The single top t-channel fiducial cross section at 8 TeV measured with the ATLAS detector
Welcome to the easter egg section of this thesis. Hi Nicole! ;)

So, first of all, why is this even here? Well, there are many things that one would want to say, but can’t. Remember, we are dealing with a tabloid. However, that doesn’t mean there aren’t things that (I think) need to be said. Hence this semi-hidden section.

My PhD-life didn’t always go smoothly. In fact, I think all those involved can agree it was rather the opposite. In my experience, I observed many times that the failures weren’t due to failings of the individuals involved, but the lack of communication. One would think that all parties involved knew what to expect, but sometimes, this isn’t the best strategy to take, and some failures of communications between them and me didn’t help either.

That being said, I don’t blame anybody. Not even when things got so heated and troubled my PhD was in serious danger of cancellation. But at the end of the day, I think this was a learning experience, for both me and my supervisors. I’d like to think we’ve all grown by this; I know I have.

That, and I’m just too stubborn to fail. :D

Also, there’s about 30% more content of this thesis that got cut due to size and for being “too technical” or “irrelevant”. Many important details... So ask me for the Extended Edition. ;)

But enough negative talk. There’s also fun things to mention.

In a way, I owe Tristan my future, for introducing me to the hacker-scene.

Nicole, you are the best! And Gwen, if you ever read this, ask your mother about “stoeptegels” and my plans with time machines. She didn’t like my plans too much, but maybe you see things differently.

With my PhD after all done,
I will be finally gone.
Leaving the academic world behind,
But never questions of the scientific kind,
As those are always in my mind.
So many friends I have made,
Oh I wish I could have stayed,
But I can not, I’m afraid.
So on with me, to larger spaces,
And moving on to happier places.

Daniël Geerts, May 9th, 2015.
The single top t-channel fiducial cross section at 8 TeV measured with the ATLAS detector
### Promotiecommissie

<table>
<thead>
<tr>
<th>Promotor:</th>
<th>Prof. dr. S.C.M. Bentvelsen</th>
<th>Universiteit van Amsterdam</th>
</tr>
</thead>
<tbody>
<tr>
<td>Co-promotor:</td>
<td>Dr. M. Vreeswijk</td>
<td>Universiteit van Amsterdam</td>
</tr>
<tr>
<td>Overige leden:</td>
<td>Dr. D.B. Ta</td>
<td>Michigan State University</td>
</tr>
<tr>
<td></td>
<td>Prof. dr. E.L.M.P. Laenen</td>
<td>Universiteit van Amsterdam</td>
</tr>
<tr>
<td></td>
<td>Prof. dr. ir. E.N. Koffeman</td>
<td>Universiteit van Amsterdam</td>
</tr>
<tr>
<td></td>
<td>Prof. dr. P.M. Kooijman</td>
<td>Universiteit van Amsterdam</td>
</tr>
<tr>
<td></td>
<td>Prof. dr. M.H.M. Merk</td>
<td>Vrije Universiteit Amsterdam</td>
</tr>
<tr>
<td></td>
<td>Prof. dr. ing. B. van Eijk</td>
<td>Universiteit Twente</td>
</tr>
</tbody>
</table>

Faculteit der Natuurwetenschappen, Wiskunde en Informatica
# Table of contents

1: Introduction ........................................................................... 1  
   1.1: Particle overview ......................................................... 1  
   1.2: Why top-quarks? ......................................................... 3  

2: Theory and event generation ............................................. 7  
   2.1: Electroweak theory ...................................................... 7  
      2.1.1: Quantum Electro Dynamics .................................. 7  
      2.1.2: Weak interactions .............................................. 8  
   2.2: Strong interaction ...................................................... 10  
   2.3: Running of the coupling strength ............................... 11  
   2.4: Proton-proton collisions in Monte Carlo generators ........ 14  
      2.4.1: Colliding partons .............................................. 16  
      2.4.2: Parton showering .............................................. 22  
      2.4.3: Hadronization .................................................... 24  
   2.5: Top-quark production .................................................. 25  
      2.5.1: Top anti-top quark pair production ...................... 25  
      2.5.2: Single top production ....................................... 28  
      2.5.3: Expected cross sections at the LHC ..................... 32  

3: Monte Carlo samples and generators ................................. 35  
   3.1: Monte Carlo generators .............................................. 35  
      3.1.1: Hard scattering .................................................. 35  
      2.4.2: Parton showering .............................................. 22  
      3.1.3: Hadronization and underlying event ..................... 37  
   3.2: Monte Carlo samples .................................................. 37  
      3.2.1: Simulation of top anti-top quark production ........... 38  
      3.2.2: Single top ........................................................ 39  
      3.2.3: $W^+\text{jets}$ .................................................... 39  
      3.2.4: $Z^+\text{jets}$ ..................................................... 40  
      3.2.5: Diboson ........................................................... 40  
   3.3: Alpgen ........................................................................ 40  
      3.3.1: Generating Alpgen samples .................................. 41  
      3.3.2: Sample definitions .......................................... 42  
      3.3.3: Jet-parton matching ......................................... 44
1.1 - Particle overview

Developed mainly in the 1960's and 1970's, the Standard Model (SM)\cite{2-5} of particle physics is a framework that describes three of the four known fundamental forces (gravity is not included). The SM describes the interactions of elementary particles and the properties of composite particles by the exchange of other particles that carry the force. The elementary particles of the SM are tabulated in figure 1.1.

Quarks
At the top left, the three generations of quark doublets are listed (pairs) horizontally. The first generation is the lightest one, with mass increasing with each generation. The quarks (anti-quarks) on the first row have a charge of +⅔ (−⅔), while the second row of quarks (anti-quarks) have charge −⅓ (+⅓). The quarks also carry color charge: red, green or blue for quarks, and anti-red, anti-green or anti-green for anti-quarks. The force particle of the strong interaction is the gluon that exchanges color charge between the quarks, and it's denoted by $g$ in the figure. This massless boson carries one color and one anti-color charge.
Hadrons
Particles composed of partons are called hadrons. The two most common groups of hadrons are the mesons (two quarks) or baryons (three quarks). The most well-known examples of hadrons are protons and neutrons, which form the building blocks of atomic nuclei.

Leptons
The leptons are listed in green, and just as the quarks, there are three generations. Each generation consists of one charged lepton and its corresponding neutrino. The first generation has the electron, present in all atoms. The other charged leptons (muons and taus) are not stable, but are produced in, for example, cosmic rays. The mass of the charged leptons increases with each generation, and the neutrinos have only very small masses.\[^6\]

Gauge bosons
The electromagnetic force is mediated by the photon (\(\gamma\) in the figure). This massless boson is coupled to the electromagnetic charge. The three other gauge bosons are associated to
Table 1.1: The percentages of each decay channel for the $W$-boson.\textsuperscript{[12]}

<table>
<thead>
<tr>
<th>Decay channel</th>
<th>Percentage [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electron</td>
<td>10.71 ± 0.16</td>
</tr>
<tr>
<td>Muon</td>
<td>10.63 ± 0.15</td>
</tr>
<tr>
<td>Tau</td>
<td>11.38 ± 0.21</td>
</tr>
<tr>
<td>Hadrons</td>
<td>67.41 ± 0.27</td>
</tr>
</tbody>
</table>

the weak interaction: the $W^\pm$ and $Z^0$. Their masses are quite significant (80.4 GeV for the $W$-bosons, and 91.2 GeV for the $Z$-boson). The weak interaction is responsible for the radioactive decay of nuclei, and allows for the flavor violation in electroweak theory.

**Higgs boson**

Without mass, there would not be three generations of quarks and leptons. For instance, a top-quark would be exactly the same particle as an up-quark. To endow particles with mass, the Higgs mechanism was proposed by Higgs et al.\textsuperscript{[7-8]} Through the interaction with the Higgs field the particles acquire their mass and the mass difference between top- and up-quarks becomes as large as the mass difference between a tennis ball and a truck. The SM does not predict the magnitude of the mass of each particle.

The Higgs boson's discovery was recently announced by CERN.\textsuperscript{[9-10]} It is the particle associated with the Higgs field. Its discovery has drawn a lot of attention, and François Englert and Peter W. Higgs were awarded the Nobel price in 2013.\textsuperscript{[11]}

### 1.2 - Why top-quarks?

The top-quark is the heaviest quark in the SM. In fact, it is the heaviest fundamental particle known, with a mass of $173.21 ± 0.51 ± 0.71$ GeV.\textsuperscript{[12]} It was experimentally verified to exist only in 1995 by the CDF and D0 collaboration at the Tevatron accelerator at Fermilab.\textsuperscript{[13-14]}

The top-quark almost exclusively decays to a $W$-boson and a $b$-quark. It has a short lifetime of $\tau_t = \frac{1}{\Gamma_t} = \sim 5 \cdot 10^{-25}$ sec. At this time scale no hadronization takes place. Hadronization is the formation of hadrons, which will be explained in more detail in section 2.4.3, and it takes place in a typical hadronization time $\tau_{\text{had}} = \sim \frac{1}{\Lambda_{\text{QCD}}} = \sim 3 \cdot 10^{-24}$ sec. The final decay mode of the top-quark is determined by the decay of the $W$-boson it produces. The percentages are given in table 1.1.
Chapter 1

The top-quark is special for two reasons. Firstly, its heavy mass is close to the electroweak scale, where it is expected that new physics will appear. Secondly, due to its lack of hadronization, measurements of top-quarks is the closest one can get to study a bare quark. In its decay, the top-quark spin is propagated to its decay products. The effect on the angular distributions of the decay products is measurable. New physics (NP) would directly influence the spin structure of the $Wtb$ vertex, resulting in deviations from the Standard Model expectations.

Measurements of the top-quark properties are a good test of the SM. Various assumptions, such as the unitarity of the CKM-matrix (which we describe in section 2.1.2) can be tested. Additionally, top-quark processes form a significant background to many new physics searches, as these often focus on events with multiple jets, high $p_T$ leptons, and missing energy. Accurate determination of the top-quark mass is essential for a reliable estimation of the top-quark backgrounds to these searches, as the production cross section of top-anti-top quarks is very sensitive to the top-quark mass.

New physics may modify the $Wtb$ vertex, and show up as deviations from the SM prediction in production rates involving this vertex. There are several different production channels for top-quarks (described in section 2.5), and each of the corresponding cross sections can be affected differently by new physics. Therefore, the production rate (both absolute and relative) of these channels are important to measure.

Many new physics models predict new large mass particles. Due to the large top-quark mass, top-quark processes may have quantum interference and show evidence of these new physics models. For example, the invariant top anti-top mass has been proved to be an almost model-independent way of searching for new physics.[15]

Additionally, many extensions to the SM predict new types of particles decaying into top-quarks pairs. These particles show up as narrow peaks in the spectrum, in the TeV range.

And finally, corrections can be expected in top-quark loops due to postulated supersymmetric (SUSY) particles. Similarly, top-quark loops can show up in Higgs propagators.

There is a strong connection between top-quarks and the Higgs-boson. In the SM the radiative corrections to the top-quark mass and $W$-boson mass depend on the Higgs mass. Figure 1.2 shows the consistency of these masses as predicted by the Standard Model.

And lastly, apart from all the interesting avenues of new physics, there is a very important reason why we are looking at top-quarks in this thesis. Because of the high collision energy of the LHC, top-quark production is relatively easily accessible. We discuss this further in section 2.5.3.

In chapter 2, 3, and 4 we discuss the theoretical framework, and the methods used for calculating theoretical predictions. In chapter 5, the ATLAS detector and the data-taking process are explained. We finish with an inclusive and a fiducial cross section measurement of single top t-channel production in chapter 6, 7, and 8.
Figure 1.2: The relation between the $W$, top-quark and Higgs masses. As the $W$ and top-quark masses are measured better and better, the constraint on the Higgs boson mass becomes tighter and tighter. Illustration from [16-17].
Chapter 1
2 - Theory and event generation

In the Standard Model the elementary particles and their interactions are mathematically represented as fields in the electroweak (EW) and the strong interaction theory. Both are relativistic quantum field theories (QFT), which enable to quantitatively explain particle phenomena. The QFT of the strong theory is called Quantum Chromo Dynamics (QCD).

In this chapter we give a phenomenological description of both EW and QCD theory. In the second part of this chapter we discuss the generation of collision events with Monte Carlo generators, describing several methods and approximations used to simulate proton-proton collisions.

2.1 - Electroweak theory

The electroweak theory is a unified description of the electromagnetic force called Quantum Electro Dynamics and the weak force. In this section, we will describe both separately.

2.1.1 - Quantum Electro Dynamics

Quantum Electro Dynamics (QED) is the quantum theory describing the electromagnetic force. The force carrier of QED is the photon which couples to the conserved electric charge of particles.

QED obeys a U(1) symmetry, and the quantum interaction is represented by a Feynman vertex that couples the fermions to photons, as is illustrated in figure 2.1. Here $e$ is the elementary charge, and the associated coupling constant is $\alpha_{EM} = \frac{1}{137}$, obtained at low energy scales where it appears in atomic physics as the fine structure constant $\alpha_{EM} = \frac{e^2}{4\pi}$. 
Chapter 2

![Figure 2.1](image_url)

Figure 2.1: The QED Feynman vertex, which couples fermions to photons. $\gamma^\mu$ are the Dirac matrices.

2.1.2 - Weak interactions

The weak interaction affects all fermions in the SM (i.e. the quarks and leptons) and it conserves the weak isospin. It is mediated through the $W$- and $Z$-bosons, which were discovered in 1983.[18-19] It obeys the SU(2) symmetry, which groups all fermions into doublets. The quark doublets are: $\begin{pmatrix} u \\ d \end{pmatrix}$, $\begin{pmatrix} c \\ s \end{pmatrix}$, and $\begin{pmatrix} t \\ b \end{pmatrix}$, and the lepton doublet are: $\begin{pmatrix} \nu_e \\ e \end{pmatrix}$, $\begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}$, and $\begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}$.

The weak interaction breaks parity symmetry (mirroring), as only left-handed particles interact weakly, and right-handed particles are unaffected. Another intriguing feature of the weak interaction is that it also breaks, at a fairly small rate, the conservation of the Charge Parity (CP) operation.[20] So far, CP violation has only been observed in weak interactions. A quark can change flavor through the emission of a $W$-boson, as shown in figure 2.2d. The neutral $Z$-boson allows for quark pair creation and annihilation (similar to figure 2.2c). The coupling factors of these processes are $g$ and $g'$ respectively. The couplings can be related to the coupling factor of the electromagnetic force:

$$ e = g \sin (\theta_W) = g' \cos (\theta_W) . $$

(2.1)

The weak mixing angle $\theta_W$[5] is the mixing angle between the SU(2) and U(1) hypercharge group, and it is dependent on the energy scale at which it is probed ($Q^2$, which we will introduce in more detail later). The symmetry breaking results in the physical gauge bosons $\gamma$, $Z^0$, $W^\pm$, and a Higgs particle. An interesting consequence on the masses of the gauge bosons is the relation given in equation 2.2:
Figure 2.2: Six fundamental Feynman vertices involving $W$- and $Z$-bosons. (a) and (b) show the Feynman vertices of a $W$-boson can change the type (flavor) of a lepton (quark). (c) shows the Feynman vertices of a $Z$-boson changing the flavor of a lepton. (d) shows a $Z$-boson interacting with a quark. (e) and (f) show the self-interacting Feynman vertices.

$$\rho = \frac{m_W^2}{M_Z^2 \cos(\theta_W)}, \quad (2.2)$$

where $\rho$ is a parameter that is very close to 1 in the Standard Model.\(^{[21]}\)

The experimentally obtained value of $\sin^2(\theta_W) = 0.23126 \pm 0.0005$ at an energy of 91.2 GeV.\(^{[12]}\) This relation is affected by higher order quantum corrections that involve the Higgs particle, allowing a theoretical prediction of the mass of the Higgs particle using
experimental inputs, long before it was observed. The magnitude of the weak factors is thus of the order \( g = \frac{e}{\sin(\theta_W)} \approx 2e \), which shows that the weak coupling is intrinsically stronger than the electromagnetic coupling. At low energies however, the propagator \( \frac{1}{m_{W,Z}^2} \) reduces the effective strength of the weak interaction by a factor \( 10^4 \) (hence its name).

The weak charged interaction describes flavor-changing currents. The mixing between the flavor eigenstates and the mass eigenstates is called the Cabibbo-Kobayashi-Maskawa (CKM) matrix, and it is shown in equation 2.3. This 3 x 3 matrix is unitary in the SM, and the elements \( V_{ij} \) are proportional to the mixing of quark \( i \) and quark \( j \). This matrix has contributions from complex phase angles that lead to the aforementioned CP violation. The magnitude of the matrix elements are obtained experimentally, using constraints of the theory.

\[
\begin{pmatrix}
d' \\
s' \\
b'
\end{pmatrix} = \begin{pmatrix}
V_{ud} & V_{cd} & V_{td} \\
V_{us} & V_{cs} & V_{ts} \\
V_{ub} & V_{cb} & V_{tb}
\end{pmatrix} \begin{pmatrix}
u \\
c \\
t
\end{pmatrix}
\] (2.3)

The value of the element that is most important in this thesis \( |V_{tb}| \) is determined to be 0.99914 ± 0.00005.

### 2.2 - Strong interaction

Quantum Chromo Dynamics (QCD) is the theoretical description of the strong interaction, and it is based on local SU(3) symmetry of 'color'. The color charge is carried by quarks and transferred by gluons that carry a combination of colors and anti-colors. Quarks are fermions, while gluons are the massless spin-1 bosons that mediate the strong interaction. Figure 2.3 shows the QCD vertices. Notable are figure 2.3b, and figure 2.3c where the gluon fields couple to themselves.

The coupling constant of the strong interaction \( \alpha_s(m_Z) = 0.1185 \pm 0.0006 \) thus of the order of ten times larger than the electroweak interaction. As will be discussed later, an overwhelming amount of interactions are indeed produced by QCD processes at the LHC.

The coupling strength is closely related to the cross section of an interaction. If new physics contributes to an interaction, the coupling is affected, and thus the cross section as well. Hence, the measurement of the cross section of several processes enables to discover or constrain new physics.
In this thesis, the probe for new physics is the single top production in the t-channel, which will be explained in section 2.5.

2.3 - Running of the coupling strength

Despite its name, the coupling constant depends on the energy scale of the interaction. We briefly discuss the dependency of its value on the energy scale heuristically, known as the running of the coupling constant.

The running of the QED coupling is a subtle effect, originating from the fact that QED is a perturbative theory describing the creation and annihilation of particle anti-particle pairs. In fact, the charge of the electron is shielded by virtual electron positron pairs and the higher the energy scale of the probe (the photon) the more of the bare charge becomes 'visible'. This principle is illustrated in figure 2.4a. As suggested by this hand-waving argument, the coupling constant $\alpha_{EM}$ increases (slowly) with the energy scale.

In the calculation the energy scale at which $\alpha_{EM}$ is expressed has to be specified. In QED this is usually the so-called renomalization scale $\mu_R$ (which will be discussed in further detail in section 2.4.1). This scale is in general different from the scale at which the interaction is probed, $Q^2$. This $Q^2$ is the energy scale$^1$ of the virtual particle connecting to the vertex, and is denoted by $Q^2 \equiv -q^2$ where $q$ is the four-momentum of the virtual particle. In QED, $Q^2$ is a measure for the resolution of the photon that is used to probe the lepton. This can be written qualitatively as:

---

1. Note that the energy, and the square of the energy are both referred to as the energy scale.
Figure 2.4: (a) shows the screening effect of QED. At large distances (small energies), the charge at the center gets screened as illustrated on the circle, and becomes seemingly smaller. (b) shows the antiscreening effect of QCD. At smaller distances, more virtual particles appear, screening the color charge at the center. Based on images from [23].

![Diagram](image)

Figure 2.5: The QED vertex, with the photon propagator modified with one virtual loop.

\[ \lambda \approx \frac{1}{\sqrt{|Q^2|}}. \]  

(2.4)

where \( \lambda \) is the wavelength of the photon.

The running of the coupling constant emerges from higher order terms to the QED vertex (figure 2.1). The photon propagator is corrected with diagrams containing virtual loops, as shown in figure 2.5. The contributions of diagrams with such loops changes with the energy scale \( Q^2 \), resulting in a running of the coupling constant.
When the value of $\alpha_{EM}$ is measured at a scale $\mu_R^2$, then its value at the energy scale $Q^2$ can be evaluated with:

$$\alpha_{EM} (Q^2) = \frac{\alpha_{EM} (\mu_R^2)}{1 - \frac{\alpha_{EM} (\mu_R^2)}{3\pi} \log \left( \frac{Q^2}{\mu_R^2} \right)}.$$

As expected, at higher energies the coupling becomes stronger. At energies relevant in this thesis around values similar to the mass of the $Z$-boson, the value of $\alpha_{EM}$ has grown to approximately $1/128$.

Similar to QED, the strength of the interaction vertex is also energy dependent in QCD. However, unlike QED, the strength decreases with energy. This can be explained, albeit a bit handwaving, that at close distances to the original parton the combined color of radiated virtual gluons, carrying both a color and an anti-color, lead effectively to a more 'blurred' surrounding and thus neutral color. This behavior is called antiscreening of color charge, illustrated in figure 2.4 and gives rise to asymptotic freedom.

Just as with the photon propagator, virtual loops cause the running of the QCD coupling constant. The gluon propagator however gets modified with two different virtual loops. Figure 2.6a shows the lepton virtual loop, similar to QED. But, due to the gluon self-coupling, another virtual loop is possible: the gluon virtual loop: figure 2.6b. This changes the way the coupling constant runs with the energy scale.

2. Only a single active flavor has been assumed in this equation.
The QCD coupling constant $\alpha_s$ can be expressed as:

$$
\alpha_s (Q^2) = \frac{\alpha_s \left( \mu_R^2 \right)}{1 + \frac{\alpha_s (\mu_R^2)}{12\pi} (33 - 2n_f) \log \left( \frac{Q^2}{\mu_R^2} \right)},
$$

(2.6)

where $n_f$ stands for the number of quark flavors involved in the interaction. Particles under the strong interaction behave as free particles at very short distances or equivalently at high energies. This is because a higher energy $\mu_R^2$ leads to a smaller value of $\alpha_s$, and thus a weaker coupling. This feature of QCD allows to consider quarks inside the proton as free particles when probed at high energies. On the other hand, when the energy becomes lower (around 1 GeV), the coupling strength grows to unity, where perturbative QCD no longer holds. The running of the strong coupling constant can then be written as:

$$
\alpha_s (Q^2) = \frac{12\pi}{(33 - 2n_f) \log \left( \frac{Q^2}{\Lambda_{QCD}^2} \right)},
$$

(2.7)

where $\Lambda_{QCD} = \sim 0.2$ GeV is the energy at which the effective coupling becomes undefined. In fact, at this scale the quarks are bound together forming hadrons. Removing a single quark from a hadron further increases the coupling and hence increases the energy density for separation to such a high energy density that new partons will be created. In other words: color-charged particles cannot be isolated in the macroscopic world. Stable particles will therefore always be color-neutral; this is called color confinement.

Effectively, high energetic partons that emerge from the proton-proton collision will emit QCD radiation in order to form color-neutral hadrons. As a result, each high energetic parton produced in the hard scattering will create a spray of other, collimated partons around it that form final state hadrons. These collimated sprays of final state particles are called jets. The direction and momentum of the initial parton corresponding to a jet can be inferred. The reconstruction and selection of jets will be described in more detail in section 6.3.

### 2.4 - Proton-proton collisions in Monte Carlo generators

A full calculation of the physics of a proton-proton collision is very complex, and the calculation of many observables directly from the Lagrangian of the QCD and EW theories is infeasible. Instead, the predictions are obtained using Monte Carlo (MC) techniques. The event generation starts with a description of the collision of two protons. The result of these Monte Carlo techniques is a list of all final state particles that are produced in the collision.
Figure 2.7: The breakdown of an MC event. The various phases of MC event generation are indicated: hard scattering (HS), initial state radiation (ISR), final state radiation (FSR), hadronization, and subsequent decay, the underlying event (UE), and the parton distribution function (PDF). These phases are further described in the text. Illustration from [25].

Figure 2.7 schematically illustrates the event generation. This figure shows the various phases of the theoretical prediction. These phases are mostly separated by different effective energy scales. Even though the complexity of this illustration may seem overwhelming, this figure illustrates the steps of event generation. In the center of the figure the colliding protons are drawn, traditionally represented by three lines that disappear in large ellipses (in green). From there, the individual partons are drawn that contribute to either the underlying event (UE) or the hard scattering (HS). The HS plays the central role in the event generation and is
usually calculated using perturbative techniques. Processes like initial state radiation (ISR) and final state radiation (FSR) also take place. As a result of all these phases, a collection of partons is created. In the next phase, these partons hadronize into colorless hadrons or leptons. Unstable particles will subsequently decay.

During the event generation these phases are treated separately. Some phases use the result of the previous phase, so the generation of these phases is not independent. In the following sections we discuss these phases of the Monte Carlo event generation further.

2.4.1 - Colliding partons

In this section we describe the hard scattering and parton density functions, which form the core of the event generation.

The subprocess of two interacting partons at high energy is illustrated in figure 2.8. This interaction is governed by either the strong or electroweak force and can lead to a number of final state partons with a high $p_T$. The energy scale $Q^2$ of the hard scattering is less than the full center of mass energy squared ($s$) of the proton-proton system, because each parton in the proton carries only a fraction of the total proton's energy. This fraction is denoted by $x_1$ ($x_2$), which is the fraction of each respective protons four-momentum $P_1$ ($P_2$) carried by the struck partons. The squared center of mass energy of the hard scattering $\hat{s}$ is expressed as:

$$\hat{s} = x_1 x_2 s .$$  \hspace{1cm} (2.8)
The hard scattering describes the phase of the proton-proton collision in which the two incoming partons interact and produce the final state $X$. The cross section associated with this hard scattering is indicated with $\hat{\sigma}$, and it can be calculated through perturbative QFT. The fractions $x_1$ and $x_2$ are obtained by the parton density functions. In order to calculate the production cross section of proton collisions leading to the final state $X$, a technique called factorization is used.

**Factorization**

In proton-proton scattering at high energy scales, the cross section $\sigma_{1,2 \rightarrow X}$ of two protons leading to the final state $X$ can be calculated by accounting for all the possibilities for the two protons to provide two partons with momentum fraction $x_1$ and $x_2$. The cross section $\hat{\sigma}$ of the process is then calculated where these two partons interact to produce a certain final state. This approach is described by the factorization theorem. Factorization enables to separate between the hard perturbative physics and non-perturbative phenomena. Deep inelastic scattering experiments at CERN, SLAC, and DESY have successfully used this description in order to obtain a measurement of the parton density functions. The parton density function (PDF), denoted as $f_i(x_1, \mu_F)$, describes the probability of obtaining a parton with flavor $i$ and fraction $x_1$ in the proton at the so-called factorization scale $\mu_F$ (which will be discussed later). The proton-proton cross section $\sigma_{1,2 \rightarrow X}$ is obtained by a convolution of the parton density functions with the hard scattering process:\[27\]

$$\sigma_{1,2 \rightarrow X} = \sum_{i,j} \int_0^1 dx_1 \int_0^1 dx_2 \ f_i(x_1, \mu_F) \ f_j(x_2, \mu_F) \ \hat{\sigma}_{i,j \rightarrow X} \left( \hat{s}, \alpha_s(\mu_R), \frac{Q^2}{\mu^2_R}, \frac{Q^2}{\mu^2_F} \right) \tag{2.9}$$

where the coupling constant $\alpha_s(\mu_R)$ is evaluated at the renormalization scale $\mu_R$, and non-perturbative terms have been neglected. The PDFs and the cross section for the hard scattering, $\hat{\sigma}$, enter the equation separately.

The square of the factorization energy scale $\mu_F$ is usually set equal to the energy scale of the hard interaction itself, $Q^2$. The perturbative part of the process is calculated with QFT, and this gives an adequate approximation at high collision energies. The low-energy, non-perturbative physics that spoil this calculation are absorbed in the PDFs. The PDFs can be determined in measurements, for example in electron-proton scattering.

There are several parameterizations of PDFs available, made by various groups, and at different orders of the perturbation expansion. These were measured in previous experiments, and can be evolved to the ATLAS energy scales using the QCD evolution, or DGLAP equations\[28-31\] (see section 2.4.2). The PDF parameterizations used in this thesis (CTEQ6\[33-34,42-43,46-61\] and Martin-Stirling-Thorne-Watt (MSTW)\[32-45\]) are determined mainly by deep inelastic scattering experiments. Each PDF parametrization is only valid at
Figure 2.9: The CTEQ6M PDF$^{[61]}$ as function of $x$ for several parton types. The energy scale at which the PDF is evaluated ($\mu_F$) in the left plot (a) is 2 GeV, and 100 GeV for the right plot (b). Illustrations adapted from [61].

The PDF used in equation 2.9 needs to match the order of $\alpha_s$ in the hard scattering calculation.

In figure 2.9, the CTEQ6M PDF$^{[61]}$ is shown. The dependency of the parton densities on the value of $x$ is clearly visible. Because the proton consists of two up- and one down valence quark, the PDFs for the $u$ and $\bar{u}$, as well as $d$ and $\bar{d}$, are not identical; the probability of extracting an up- or down-quark are significantly higher than finding their anti-counterparts in the proton, especially for high values of $x$. At low values of $x$, charm-quarks dominate, for which the charm and anti-charm distributions are identical. The dominant role of the gluon at lower values of $x$ is also clear. At high values of $\mu_F^2$, the parton densities at low $x$ also increase. For this reason, the gluons play a dominant role in many (but not all) production mechanism at LHC energies (for example, many top-quark production processes are gluon-dominated at LHC energies; see section 2.5.1).

**Choice of energy scale**

The choice of the numerical value of the energy scale $Q^2$ for each collision is not always obvious. However, extrapolations and approximations made during the calculation introduce a dependency on the energy scale. The choice of $Q^2$ is such that it reduces the uncertainty on the final cross section calculations.

Normally, $Q^2$ is set to the transverse mass $m_T^2$ of the virtual particle participating in the interaction, where the transverse mass $m_T$ is defined as: $m_T \equiv \sqrt{m^2 + p_T^2}$.3
In the case of top-quark production, usually $Q^2$ is set to $m_t^2$, the top-quark mass squared. At this value, the energy scale is well above $\Lambda_{QCD}$, hence the perturbative expansion converges as required.

Apart from the non-perturbative physics that are absorbed into the PDF at low energy scales, other divergences emerge at high scales. These high scale divergences correspond to the loop corrections to the propagator. As the momentum of the virtual particle in the loop is unbounded, the integral over this free momentum-variable usually diverges. These divergences are handled by renormalization. Effectively, renormalization cuts this integral into two parts, with the divergent part absorbed into the coupling strength of the virtual particle. The energy scale cut-off chosen for the integral is called the renormalization scale, denoted by $\mu_R$. The value of the renormalization scale is chosen such that $\alpha_s$, which is evaluated at the renormalization energy scale, satisfies $\alpha_s(\mu_R) \ll 1$. In practice, the renormalization scale and factorization scale are set equal.

In the end, all these choices are expected to have only a small effect on the observables, although the actual size of this effect depends on the order of the calculation. We will demonstrate this explicitly for the Monte Carlo samples used in this thesis in chapter 4.

Figure 2.10: (a) shows a LO diagram that is part of the $W^+\text{jets}$ MC sample, and (b) shows a Feynman diagram showing vector boson fusion: the production of two vector bosons, which then merge. This particular process can only be made by using at multiple EW interaction vertices.

3. Note that in the rest of this thesis, an alternative definition is used: $m_T \equiv \sqrt{2 p_T^1 p_T^2 (1 - \cos \phi^{ij})}$. 
Chapter 2

Figure 2.11: Two Feynman diagrams that contribute to the W+jets process. (a) shows an LO diagram, while (b) is one of the NLO diagrams, producing an additional gluon in its final state.

Perturbation theory

Higher order diagrams traditionally have extra QCD and electroweak vertices in the hard scattering process. However, we adopt a different definition. Only if a diagram has additional QCD vertices, we call it a higher order diagram. All other diagrams (that is, diagrams with additional QED and/or EW vertices) are categorized as different processes. An example where additional electroweak vertices are added, resulting in a different process, is vector boson fusion (VBF), as illustrated in figure 2.10.

Each additional QCD vertex adds an extra coupling constant \( \alpha_s \) to the cross section. Its perturbative expansion can thus be expressed in orders of \( \alpha_s \). The expansion of the parton-level cross section in order of \( \alpha_s \) can be written as:

\[
\hat{\sigma} = \hat{\sigma}_0 + \frac{\alpha_s}{2\pi} \hat{\sigma}_1 + \left( \frac{\alpha_s}{2\pi} \right)^2 \hat{\sigma}_2 + \left( \frac{\alpha_s}{2\pi} \right)^3 \hat{\sigma}_3 + \ldots \tag{2.10}
\]

The first term describes the Born approximation (this is the leading order (LO) QCD prediction), the second term the next-to-leading order (NLO) corrections, and so forth. LO diagrams do not contain virtual loops. Higher order QCD diagrams can produce additional final state partons. An example of an LO diagram is given in figure 2.11a, with figure 2.11b showing an NLO diagram of the same process.

Many MC generators use hard scattering processes only to LO. Different techniques are applied for higher order QCD effects, as which will be discussed later. However, nowadays several NLO generators for 2→2 and 2→3 hard scattering processes are available. There are also generators that, in a way, fall between LO and NLO: they are able to generate diagrams with many final state partons, but ignore virtual loop corrections (and thus it cannot be seen as an NLO generator). We will give more details about this in chapter 3.

4. The PDF set used in the factorization must match the order of expansion in \( \alpha_s \).
Ideally, the final predictions for any measurable quantity should be independent of both the factorization and renormalization energy scales. However, if the perturbative expansion is cut-off at a fixed order in $\alpha_s$, there will be a residual dependency on $\mu_F$ and $\mu_R$, originating from missing higher order terms. These dependencies are unphysical, and their effect on the final result should be verified to be small. A detailed study into these residual dependencies for $W$+jets is presented in chapter 4. The largest theoretical uncertainty on the calculated cross section and the simulated particle kinematics comes not only from neglecting higher-order corrections, but also from the uncertainty on the PDF.

The order at which the perturbative calculation in $\alpha_s$ is cut-off depends on various factors. This restriction is often present simply because a full higher order calculation for many final state partons is not yet available. Even when available, for most processes it is too computational intensive to perform a full higher order calculation, and thus only the leading order calculations for those processes are used.

### Cross section calculation

To obtain the cross section for a particular final state all Feynman diagrams that lead to this state have to be summed coherently. The resulting amplitudes, or matrix elements (ME), have to be squared and integrated over the relevant phase space:

$$\hat{\sigma}_n = \frac{1}{2s} \int d\Phi_p \sum_{h,c} \|A_n\|^2,$$

(2.11)

where $A_n$ is the ME calculated to the $n$-th order, $\Phi_p$ is the phase space of all partons, and $h$, $c$ are all the spin and color states respectively. After full integration of the phase space, this is called the full cross section. The domain of this phase space integration can purposefully be limited to only a fiducial volume, for example a region corresponding to the detector acceptance. This is called the fiducial cross section, which will be measured in chapter 8 for single top production.

In Monte Carlo event generation, this integration is often performed using phase space sampling. This technique leads to the assignment of event weights to each generated event, in order to ensure that the sum over events converges to the integral of equation 2.11.

Although the major features of the kinematic distributions are often predicted correctly by the LO generators, the normalization tends to be more problematic. This results in a wrongly predicted cross section, often by a few dozen percent. This factor is purely due to the LO nature of the generated events. To compensate for this, the MC events are "rescaled" with a so-called $k$-factor to the NLO (or NNLO) inclusive cross section calculated by dedicated programs, defined as:

5. For example, the Alpgen cross section prediction for the $W$+light jets sample, which will be introduced in section 3.3.2, is a factor 1.2 off from the FEWZ$^{[62]}$ NLO calculation.
In general, the \( k \)-factor depends on the region of the integrated phase space that is probed in the calculation. Hence, a global \( k \)-factor alone is not sufficient not produce accurate results, and other techniques are used to normalize the Monte Carlo events to data. These so-called data-driven techniques will be explored in more detail in chapter 4.

2.4.2 - Parton showering

In this section we discuss a common technique to radiate additional partons called parton showering. Parton showering is used with a wide variety of generators to further simulate the collision from the hard scattering onwards. Here, we present the parton showering based on splitting kernels and Sudakov form factors. Other approaches exist, such as the color dipole model. This method was not used in this thesis, and will not be further discussed.

The calculation of the hard scattering as previously described is only performed to a certain order in QCD. A parton showerer takes the outgoing particles of the hard interaction, and through subsequent radiation of partons evolves it to lower values of \( Q^2 \). This process is terminated at some low cut-off energy scale \( Q_{\text{cut}}^2 \), where perturbation theory is no longer valid. This energy scale is where hadronization starts, which will be described in section 2.4.3.

The showering (sometimes referred to as a cascade) is based on three possible interactions, described by the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) splitting kernels:

\[
K \equiv \frac{\sigma_{\text{NLO}}}{\sigma_{\text{LO}}}. \tag{2.12}
\]

These kernels are shown in figure 2.12: a gluon splitting into two gluons (figure 2.12a), gluon splitting into a quark-anti-quark pair (figure 2.12b), or a quark radiating a gluon (figure 2.12c). The mathematical description of these three splitting kernels is given by the DGLAP splitting functions \( P_{a \rightarrow bc} (z) \). These functions describe the probability of a parton \( a \) splitting
into a parton $b$ and $c$, where parton $b$ carries a fraction $z$ of the original parton $a$, and parton $c$ a fraction $1 - z$. The full expression for radiating partons becomes:

\[ dP_{a \to bc} (Q^2_{\text{max}}, Q^2) = \frac{dQ^2}{Q^2} \sum_{b,c} \int_{z_{\text{min}}}^{z_{\text{max}}} dz \frac{\alpha_s (Q^2)}{2\pi} P_{a \to bc} (z) \cdot \Delta (Q^2_{\text{max}}, Q^2) \]

where $\Delta (Q^2_{\text{max}}, Q^2)$ is the Sudakov form factor, which gives the probability that no splitting (or branching) occurs between the two energy scales $Q^2_{\text{max}}$ and $Q^2$. $Q^2_{\text{max}}$ gives the starting energy scale of the partons, and $Q^2$ the energy scale of the particles after their splitting.

The resulting partons, which have a lower $Q^2$, then again have a probability to split, and this iterative procedure is repeated until the $Q^2$ of the partons has evolved to $Q^2_{\text{cut}}$, the lower cut-off of the parton shower.

The effects of the parton showering shows up in equation 2.9 as behavior following equation 2.14. Most of the parton shower implementations are leading log (LL) approximations. They only include the first term ($n = 1$) of this perturbative expansion.

\[ \sigma \sim \sum_n c_n \left( \alpha_s \log \left( \frac{Q}{Q_{\text{cut}}} \right)^2 \right)^n, \]

where $n$ denotes the order and $c_n$ are the corresponding coefficients. It is clear that if $Q^2$ is not of equal magnitude as $Q^2_{\text{cut}}$, the higher order logarithmic terms grow larger. This means the convergence becomes slower, and more orders would need to be included for a good result. The parton showerers that we use in this thesis all implement LL parton showers. Often however, additional effects such as the running of $\alpha_s$ are taken into account in such a way that the PS includes various NLL effects.

Similarly to the final state partons of the ME, the initial state partons can also radiate before entering the ME. This is called initial state radiation (ISR), and it is also indicated in figure 2.7. ISR is similar to parton showering which could also be called final state radiation (FSR). ISR is calculated from the hard scattering backwards in time, describing the probability that a parton was extracted from the proton with a higher energy and radiated before the HS took place.

The beam remnants, consisting of the partons that are not involved in the HS, will continue on the proton’s original trajectory, carrying the rest of the proton energy. These partons no longer form a color-neutral object, and this leads to QCD interactions mostly in the forward
direction. The resulting particles in general will have a much lower transverse energy than the hard scattering. These interactions are usually described as $2 \rightarrow 2$ QCD interactions, being LO in $\alpha_s$. The process of modeling these beam remnants is called the underlying event.

### 2.4.3 - Hadronization

After the parton showering, the resulting partons are hadronized in order to form color-neutral particles. There are two methods of performing the hadronization that are used in this thesis; cluster fragmentation and string fragmentation. String fragmentation is employed in most Monte Carlo samples used in this thesis.

#### String fragmentation

String fragmentation (also called the Lund String Model)\textsuperscript{[64]} uses the concept of imaginary strings between the partons. Each string represents a color connection between them. This is illustrated in figure 2.13. Gluons are represented as kinks in the string. As the partons move further apart, the potential energy in the string increases until the string breaks, triggering the creation of new $q\bar{q}$ pairs. The newly created partons are chosen such to suppress the creation of new heavy quarks. Once no further fragmentation takes place, the partons are grouped together per string into color-neutral hadrons.

#### Cluster fragmentation

Cluster fragmentation works by grouping the partons into color-neutral clusters, as shown in figure 2.14. At the end of the parton showering, all the final state gluons are split into quark-anti-quark pairs. Quark-anti-quark pairs that are close together in phase space are then combined into color-neutral clusters. These clusters are then split into smaller clusters using preset decay
Figure 2.14: This figure illustrates cluster fragmentation. The horizontal axis measures spatial distance, while the vertical axis shows time. Starting from the bottom, parton showering creates final state partons. Next, quarks are grouped into clusters (drawn as grey ovals), which are decayed to color-neutral hadrons. Illustration adapted from [63].

Tables until they are can be split no further. At that point, each cluster is converted into a color-neutral hadron.

**Stable particles**

Many of the hadrons created during the hadronization are not stable. These unstable particles are decayed into stable particles, using the PDG decay branching ratios.[12] All particles with a lifetime larger than 10 mm are considered stable in this phase.6 These particles are therefore handed over to the detector simulation, which will be discussed in section 5.3.

### 2.5 - Top-quark production

At the LHC top-quarks are either produced in pairs, called $t\bar{t}$, or as single top. In this section we will discuss these two production mechanisms.

#### 2.5.1 - Top anti-top quark pair production

Two incoming partons can interact through the strong interaction, resulting in the production of a top and anti-top quark pair ($t\bar{t}$). There are two possible combinations of incoming partons at leading order: quark-anti-quark annihilation ($q\bar{q}$) and gluon fusion ($gg$). These are

---

6. Their travel paths will fall within the ATLAS detector, and their interactions with the material of the detector need to be taken into account.
Chapter 2

Figure 2.15: The different channels for $t\bar{t}$ production at leading order. (a) shows quark anti-quark annihilation leading to a top-quark pair creation; (b) shows the gluon fusion process; and (c) shows another gluon-gluon process leading to a top-quark pair.

illustrated in figure 2.15. Of these processes, gluon fusion has the largest cross section at the LHC.

The cross sections of the $t\bar{t}$ process as predicted by theory at various energies are listed in table 2.1. The cross sections increases rapidly with the collision energy.

<table>
<thead>
<tr>
<th>$t\bar{t}$ production</th>
<th>Collision energy</th>
<th>$\sigma$ (pb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tevatron ($p\bar{p}$)</td>
<td>1.96 TeV</td>
<td>$7.08^{+0.36}_{-0.24}$</td>
</tr>
<tr>
<td>LHC ($pp$)</td>
<td>7 TeV</td>
<td>$163^{+9}_{-5}$</td>
</tr>
<tr>
<td>LHC ($pp$)</td>
<td>8 TeV</td>
<td>$234^{+12}_{-7}$</td>
</tr>
<tr>
<td>LHC ($pp$)</td>
<td>14 TeV</td>
<td>$920^{+33}_{-35}$</td>
</tr>
</tbody>
</table>

Table 2.1: Theoretical predictions for the $t\bar{t}$ production cross section at various collision energies. The first uncertainty gives comes from the scale variation $\frac{m_t}{2} < \mu < 2m_t$, and the second from the 90% C.L. on the PDF. Table from [65].
More than 80% of the pair production of top-quarks at the LHC is expected to be produced via the gluon fusion process. About 20% is produced via $q\bar{q}$ production. In the $\overline{\text{MS}}$ scheme the NLO $qg$ contribution is small, in the order 1%. It is interesting to note that instead at the Tevatron in $p\bar{p}$ scattering at 1.96 TeV, more than 85% of the top pairs are produced in the $q\bar{q}$ channel.\footnote{7}

The decay of the top-quark pair leads to two oppositely charged $W$-bosons and $b$-quarks. Both $W$-bosons can either decay leptonically or hadronically. Combined, this results in three decay modes:\footnote{8}

- Dileptonic, where both $W$-bosons decay leptonically. About 9% of $t\bar{t}$ events all into this category.
- Semi-leptonic, where only one $W$-boson decays leptonically, and the other hadronically. About 45% of $t\bar{t}$ events all into this category.
- (Fully) hadronic, where both $W$-bosons decay hadronically. About 46% of $t\bar{t}$ events all into this category.

Figure 2.16 illustrates these modes per lepton- and quark flavor combination.

The current ATLAS $t\bar{t}$ cross section measurement for these channels combined at a collision energy of 7 TeV is:\footnote{[69]}

$$\sigma_{t\bar{t}} = 165 \pm 2 \, \text{(stat.)} \pm 17 \, \text{(syst.)} \pm 3 \, \text{(lumi.)} \, \text{pb}.$$  \hspace{1cm} (2.15)

And the measurements at 8 TeV give:\footnote{[70]}

$$\sigma_{t\bar{t}} = 241 \pm 2 \, \text{(stat.)} \pm 31 \, \text{(syst.)} \pm 9 \, \text{(lumi.)} \, \text{pb}.$$  \hspace{1cm} (2.16)

Both 7 TeV and 8 TeV measurements are compatible with the theoretical predictions (table 2.1). This is illustrated in figure 2.17 for the 7 TeV result. This figure shows the ATLAS results, and compares these to the world average measurement. As the results are compatible, there is no indication of new physics in this combined, inclusive cross section measurement, confirming the Standard Model prediction.

These measurements of the $t\bar{t}$ cross sections are summarized with the measurement of Tevatron in figure 2.18, together with the theoretical predictions.

The top-quark mass is an input parameter of the SM. This means it cannot be calculated from theory, and needs to be measured. The current measurement of the top-quark mass is:

\footnote{7. The quoted cross sections were calculated with MCFM\cite{66} using the $\overline{\text{MS}}$ scheme. The MSTW2008(N)LO PDF was used, and the top-quark mass was set to 172.5 GeV. (Numbers from [26])}

\footnote{8. The fractions quoted are without taking into account the tau-decay. In other words, all $W$-bosons decaying to a tau lepton are counted as a leptonic decay. Numbers from [67].
$m_t = 173.21 \pm 0.51 \pm 0.71 \text{ GeV}$,\cite{12} where this value is currently mostly driven by previous Tevatron studies.

### 2.5.2 - Single top production

The production of a single top-quark always involves a $W$-boson and a $b$-quark through the weak interaction. As the name suggests, only a single top-quark is produced in this process, in contrast to the pair production mentioned above. Due to the unambiguous coupling with a $W$-boson, this process is a good check of weak charged current interactions of the top-quark, thus a direct probe of the $Wtb$ production vertex.

Single top-quarks are produced in three different channels as illustrated in figure 2.19. Note that any top-quarks produced in the final state will decay to a $W$-boson and a $b$-quark. The s- and t-channel are named after the Mandelstam variable that refers to the square momentum of the intermediate $W$-boson.\cite{72} The s-channel requires a quark and an anti-quark, both with relatively high momentum fractions to produce an off-shell $W$-boson that decays to a $b$- and top-quark. As listed in table 2.2, the LHC cross sections for this channel are only a little higher than the Tevatron, due to the relatively large PDFs of the initial (anti-)quark in proton anti-proton collision. Experimentally the s-channel can be recognized...
by the additional $b$-quark in the final state that results into a jet, that can be 'tagged' as a $b$-jet (see section 6.4).

The t-channel has a considerably larger cross section at the LHC than at Tevatron. This channel is sometimes referred to as $W$-gluon fusion, referring to the NLO picture where a gluon provides the $b$-quark. The unique experimental signature of the t-channel is the additional light quark jet that emerges from the collisions with high rapidity. In this thesis we focus only on the leptonic decay for which the cross section is 28.3 pb. The Wt-channel, named after its decay products, has a (close to) on-shell $W$-boson in the final state. This feature allows to experimentally discriminate this channel from the other top-quark production channels. Due to the massive particles in the final state, the cross section of the channel is extremely low at Tevatron (0.088 pb), but the cross section at the LHC is relatively large. The experimental signature of these channels is further determined by the decay of the top-quark and thus the decay mode of the final state $W$-boson.

The cross sections of all three single top processes are proportional to $V_{tb}^2$. An accurate cross section measurement can thus lead to a direct measurement of this quantity.
Figure 2.18: This figure shows the high energy measurements of the $t\bar{t}$ cross sections, and the theoretical predictions. Illustration from [71].

Figure 2.19: The different single top production channels. (a) shows single top s-channel production. The final state has two outgoing $b$-quarks, and one $W$-boson; (b) shows single top t-channel production. The final state has an outgoing $b$-quark, a $W$-boson, and a light quark; (c) shows single top Wt-channel production. The final state has an outgoing $b$-quark, and two $W$-bosons.
Table 2.2: Theoretical predictions for the single top production cross section at various collision energies. These values were calculated in approximate NNLO assuming a top-quark mass of $m_t = 173$ GeV and using the MSTW2008 NNLO PDF. The first uncertainty comes from the scale variation, and the second from the 90% C.L. on the PDF. Table from [65].

**Single top t-channel**

The t-channel has the largest cross section and a measurement at the LHC with a center of mass of 8 TeV using the ATLAS detector is presented in chapter 7 and chapter 8. The LO expression for the partonic cross section is:[73]

$$\frac{d\sigma}{d\cos \theta} = \frac{\pi \alpha^2}{8 \sin^4 \theta_W} |V_{tb}|^2 |V_{12}|^2 \frac{1}{(s - M_W^2)^2} t \left(1 - \frac{m_t^2}{t}\right)^2,$$  \hspace{1cm} (2.17)

where $\cos (\theta)$ is the scattering angle in the parton-parton center of mass frame, $s$, and $t$ the Mandelstam variables, and $V_{12}$ the CKM-matrix element of the light quarks that couples to the $W$-boson.
Chapter 2

The measurements of single top quarks test the SM prediction and probe any new physics that affects the \(Wtb\) vertex.

In literature many new physics models are discussed. For example in [74] the possibility of a heavy \(W'\) is discussed that affect the s-channel, whereas it does not affect the t-channel because there its contribution drops with \(\sim \frac{1}{m_{W,Z}^2}\). Also, the t-channel is rather insensitive to a fourth family \(b'\) quark, while it would certainly show up in the Wt-channel. While the Wt-channel is insensitive to Flavor Changing Neutral Currents in the production, the t-channel receives a contribution. Models with charged Higgs boson, CP-violating anomalous couplings and contributions from supersymmetry\[^75\] are other possibilities that are seriously considered. In conclusion, single top-quark production is an important process to study.

2.5.3 - Expected cross sections at the LHC

Figure 2.20 shows the measured and predicted cross section as function of center of mass energy for \(p\bar{p}\) collisions below 4 TeV and for \(pp\) collisions for higher values, corresponding to the Tevatron and LHC domains respectively. The production of \(b\)-quarks with a cross section \(\sim 10^6\) nb (at 8 TeV), which is driven by the QCD processes, has an overwhelming cross section, as is expected from the discussion of the coupling strength in section 2.2. When requirements on the transverse energy on the jets are imposed the cross section becomes comparable to the (weak) production cross section of the \(W\)- and \(Z\)-bosons. The total top cross section including top pair production, is much smaller (250 pb) as a result of the low values of the PDF at higher \(x\), which are probed due to the high mass of the top-quark.

For the discussion of the backgrounds to top production it is important to note that the measurements described in this thesis focus purely on the leptonic decay modes of the top-quark. The requirement of an electron or muon strongly suppresses most other reducible backgrounds, such as events where the hard scattering producing only partons. This process is called QCD multijets (often abbreviated to QCD). The selection of a QCD event has such small probability that, combined with the large uncertainty associated with the theoretical prediction, an accurate prediction of the QCD background is not available. This background is therefore obtained entirely from the data itself, as discussed in section 6.6.2.

In figure 2.21, all the physics processes important to this thesis (and several others) are shown. The theoretical prediction is indicated, showing good agreement with the measurements. One process that is not shown, is the single top t-channel, a measurement of which will be given in this thesis.

Even though figure 2.20 gives a good impression what processes can be expected as backgrounds to top-quark analyses, there are other effects that modify this picture. For
Figure 2.20: The cross sections of various processes as a function of the center of mass energy. These have been calculated using the MSTW2008 PDF.\cite{32} Illustration from [76].
example, detector effects can lead to a relative enhancement of a background process with respect to top-quark signal (see chapter 5). In section 3.2 these backgrounds are further explored, with chapter 4 going into much more detail of the dominant background. In chapter 6 the top-quark analysis event selection is introduced, which will be used in chapter 7 and chapter 8 to measure the single top t-channel cross section.

Figure 2.21: The measurements of various physics processes done by ATLAS, compared to their theoretical predictions. Illustration adapted from [77].
In this chapter, we describe the Monte Carlo (MC) generators and samples used for the analyses described in this thesis. The events from $W^+$jets events are a significant background when trying to isolate events with top-quarks at the LHC. Simulation of these $W^+$jets events is not straightforward, and systematic uncertainties have to be studied carefully. In the second half of this chapter, emphasis is put on the MC generator used to produce the $W^+$jets events, Alpgen.

3.1 - Monte Carlo generators

There are various Monte Carlo generators that can be used to simulate the various phases of event generation. Some are general purpose MC generators that can perform full event generation up to the complete list of final state particles. Others are specialized generators, with improved accuracy and/or shorter generation time for specific processes only. In this section we give a brief overview of the generator packages used most frequently in top-quark analyses at the LHC.

Hard scattering

- \textbf{AcerMC}\textsuperscript{[78]}
  \textbf{AcerMC} produces LO matrix elements for various types of hard scattering processes. It can be interfaced to Pythia, Herwig or Ariadne for parton showering. \textbf{AcerMC} was used for the single top t-channel production at 7 TeV, but its LO behavior can be insufficient in describing the shapes of various distributions (as well as the normalization), given the relatively high level of statistics available at 8 TeV.
Chapter 3

• Alpgen\cite{79}

Alpgen is a LO matrix element generator, and thus it generates tree-level hard scattering diagrams only, and specializes in multi-parton final states (2→N). It needs to be interfaced to either Pythia or Herwig for parton showering. More details about this generator are given in section 3.3.

In this thesis Alpgen is used for both the $W^+\text{jets}$ and $Z^+\text{jets}$ Monte Carlo samples.

• MadGraph\cite{80-81}

MadGraph is a general purpose HS generator that can generate NLO matrix elements. MadGraph is not limited to Standard Model processes, but this makes it also more computational intensive.

• MC@NLO\cite{82-84}

MC@NLO is a fully NLO calculator for the HS. As it includes real and virtual corrections, the weights of some of the events can be negative (typically ~13% of events). It can be interfaced with Herwig for hadronization (see below).

• aMC@NLO\cite{85} is an automated version of MC@NLO, which has been specifically designed to reduce the need for user intervention. Internally, it uses MadGraph to generate any tree-level and one-loop process.

aMC@NLO is used in this thesis for comparisons of the single top t-channel MC samples in the analysis presented in this thesis.

• Powheg\cite{86}

Powheg is a NLO generator that can generate hard scatterings up to one loop (just like MC@NLO). In contrast to MC@NLO, it produces only positive event weights. The produced events can be hadronized with Herwig or Pythia (see below).

Powheg is used for the default single top t-channel Monte Carlo sample used in this thesis.

• Sherpa\cite{87}

Sherpa is an LO generator which implements several different multileg ME generators. Sherpa has a parton showering mechanism build in, including parton-matching.

Sherpa is used for comparisons of the $W$ and $Z^+\text{jets}$ samples.

Parton showering

• Ariadne\cite{88}

Ariadne is a parton showerer that can be interfaced with Pythia. It uses a dipole showering mechanism, and will not be further discussed in this thesis.
Monte Carlo samples and generators

- **Herwig**[^89]
  Herwig, sometimes called fHerwig (Fortran Herwig), is a full event generator written in Fortran. It can generate the matrix elements of several processes, but here it is mostly used for its parton showering capabilities. It uses the cluster fragmentation method, as described in section 2.4.3.

- **Herwig++[^90]**
  Herwig++ is a particle physics event generator, written in C++. It is the successor to Herwig, built on the knowledge gained during the development of Herwig. It implements similar or better capabilities, of which the B-meson decay the most important to note.
  The B-meson decay tables fHerwig uses internally haven't been updated in years. This mainly has a significant effect on the B-meson decays. Due to the way cluster fragmentation is implemented, simply adding the new decay modes however worsens the accuracy of decay rates, and this is thus not an option.
  Herwig++ resolves this issue by using different techniques to handle these decays, and thus provides better and more up-to-date results.

- **Pythia[^91]**
  Pythia is a leading order full event generator. It can perform parton showering (supporting both QED and QCD radiation through string fragmentation, as described in section 2.4.3), hadronization, and decay. In this thesis, Pythia6 is interfaced to several alternative generators and used only for its showering and fragmentation capabilities.

**Hadronization and underlying event**

- **Jimmy[^92-93]**
  Jimmy is mostly used in combination with Herwig, where Jimmy specifically simulates the Underlying Event. It does this through an eikonal model which is based on a phenomenological description of the UE.

- **Pythia** is also used for hadronization and the UE.

**3.2 - Monte Carlo samples**

The analyses presented in this thesis use Monte Carlo simulation to estimate the efficiency of the single top-quark cross section measurement with one isolated electron or muon in the final state. Other processes with a similar signature constitute the background for which

[^89]: The combination of Herwig+Jimmy is often just called Herwig for shorthand.
additional Monte Carlo samples are required. It should be noted that the QCD multi-jet background, which contributes when one of the jets is falsely identified as lepton, is not simulated. This background, which has a large uncertainty in the simulation, is determined in an entirely data driven procedure as explained in section 6.6.2. This section summarize all Monte Carlo sample used in the further analyses.

**Multiple parton interaction**

It is possible that in a single proton-proton collision, more than one pair of partons interacts. This is called multiple parton interaction (MPI), or, if only two collisions overlap, double parton interaction (DPI). In these events, more than one hard scattering takes place, and their decay products cannot easily be distinguished. This could potentially lead to significant misclassification: two $W^+\text{jets}$ events overlapped will mimic a $WW$ diboson event, or even a $t\bar{t}$ event; these processes are described in more detail below. Estimating the impact of MPI is thus important for a correct handling of data. At the LHC energy scale, the hard scatterings of a MPI-event do not significantly interfere with each other. In other words, a DPI-event could be simulated by overlaying two hard scatterings, and similar for MPI-events. Of course, the underlying event needs to be adjusted properly when applying this method. However, MPI is currently not taken into account in the event generation. In fact, none of the Monte Carlo predictions used in this thesis explicitly account for MPI. The fraction of events estimated to be affected by DPI is about 8% of $W^+2\text{jets}$ events,\(^{94}\) a fraction that is not insignificant. However, because the hard scatterings in MPI can be treated as happening independently, this is effectively indistinguishable from pile-up. This means that MPI is already largely handled by pile-up corrections (see section 5.3.1). It is further handled through the data-driven renormalization of the $W^+\text{jets}$ MC sample, which will correct additional MPI effects. Special MPI samples are therefore not required for a good physics description for the studies done in this thesis.

**3.2.1 - Simulation of top anti-top quark production**

The phenomenology theoretical foundations of the top anti-top quark pair production process was discussed in section 2.5.1. Here we focus on the Monte Carlo samples associated with that production process.

For 7 TeV and 8 TeV analyses, an MC@NLO sample with a cross section of 177 pb and a Powheg sample with a cross section of 252 pb are used, respectively. Both these cross sections (and the others mentioned below) have been normalized to their NNLO predictions, as presented in table 2.1. The 7 TeV MC@NLO sample was parton showered with Herwig, with Jimmy handling the underlying event. The 8 TeV Powheg sample was parton showered and hadronized with Pythia.
3.2.2 - Single top

The single top production process was already introduced in section 2.5.2, with the theoretical predicted NNLO cross sections listed in table 2.2.

For 7 TeV analyses on 2011 data, an LO t-channel single top sample generated with AcerMC and Pythia was used. The cross section for this sample is 96 pb. At 8 TeV, this sample has been generated at NLO with Powheg and Pythia, with a cross section of 137 pb.

For the s-channel, both samples are NLO. Their cross sections are 1.5 pb at 7 TeV, and 1.8 pb at 8 TeV. Both samples were generated with MC@NLO and Herwig with Jimmy.

And for the Wt-channel: $15.7 \times 10^3$ pb ($24.7 \times 10^3$ pb) at 7 TeV (8 TeV). Both of these samples are NLO. Again, both samples were generated with MC@NLO and Herwig with Jimmy.

3.2.3 - $W^+\text{jets}$

In the $W^+\text{jets}$ process, a single $W$-boson is created through an electroweak interaction. This $W$-boson subsequently decays, as was illustrated in figure 2.2. Events with a leptonically decaying $W$-boson result in a similar event signature as top-quark production. The cross section for this process is generally larger than that of top-quark production, and it is therefore one of the main backgrounds.

At 7 TeV, the samples associated with this process were generated with Alpgen, with the parton showering and underlying event handled by Herwig and Jimmy respectively. At 8 TeV, the parton showering and UE were performed by Pythia.

As was listed in table 1.1, the leptonic branching ratios to electrons, muons and tau-leptons of the $W$ are equal at leading order but get tiny higher order corrections that depend on their masses. Experimentally the electron and muon can be identified with high efficiency with the ATLAS detector. The detection and identification of a tau is more complicated, but when it decays to an electron or muon, the event is selected.

Even though this background produces an isolated lepton, the production of a $b$-quark, which is part of the top-quark signature as well, has a small rate. Still, as it will turn out, the $W^+$ jet background is the largest background to most top-quark measurements with the cross section predicted by the generator being $3.9 \times 10^4$ pb. Because this background is dominant, it is essential to estimate this background correctly, in order to be able to subtract the $W^+$jets contribution from the measured data.

In the second half of this chapter, we discuss the MC simulation of the $W^+$jets samples in detail.
3.2.4 - Z+jets

A background related to $W^+\text{jets}$ is the Drell-Yan process with intermediate $Z$-boson production, called $Z^+\text{jets}$. In this process an quark anti-quark pair annihilate, producing a $Z$-boson or a virtual photon. This particle then decays into a lepton-pair. The event signature thus is 2 leptons, and additional jets. The cross section is approximately a factor 10 smaller than $W^+\text{jets}$ (3576 pb), and due to the presence of multiple leptons, it is easier to distinguish from signal events than $W^+\text{jets}$. The samples at both collision energies were generated with the same configuration of Monte Carlo generators as the $W^+\text{jets}$ samples.

3.2.5 - Diboson

The three processes where two weak gauge bosons are produced ($WW$, $WZ$, and $ZZ$) are collectively known as the diboson processes. The cross sections are 29.7 pb, 2.31 pb, and 1.47 pb respectively, with the samples generated by Alpgen, combined with Herwig and Jimmy. The event signatures of diboson events can mimic the top-quark production processes; for example, $WW$ is very similar to top anti-top production, with the exception of the $b$-quarks. Due to the low cross sections, diboson processes only constitute a small background to top-quark searches.

3.3 - Alpgen

In this section we give a detailed description of Alpgen. Alpgen is a Fortran program\cite{Alpgen} that calculates the matrix elements used in the calculation of the hard scattering. It is not a general purpose MC generator; in other words, it is restricted to only specific hard scattering processes (the list of supported hard scatterings is reasonably large. It is used within ATLAS for the generation of many MC samples; specifically the $W^+\text{jets}$ MC samples that we will discuss in this thesis. Alpgen is a LO multi-leg generator. Loop diagrams are not taken into account; the calculation only involves tree-level diagrams. Currently, up to six final state partons in the matrix element are supported. All these additional partons produced by Alpgen are treated as being massless.

Internally, Alpgen uses the ALPHA algorithm\cite{ALPHA} from which the generator derives its name. This algorithm uses a Legendre transform of the effective quantum action, from which the matrix element is derived. The calculation of the Legendre transform is formulated in a way where a numerical approach is especially suited. Also used are the Berends-Giele recursion relations,\cite{BerendsGiele} which results in an exponential growth of the required computational time to evaluate hard scattering processes with more final state partons, and not a factorial growth as would be the case when evaluating Feynman diagrams directly. This allows for
the generation of samples for processes with many final state partons in a reasonable amount of time, as is required in the LHC-era.

**Alpgen** internally handles the leptonic decay of the electroweak gauge bosons and the Higgs boson that can be produced during the interaction. For example, the resonant $W$-boson and $Z$-boson leptonic decays are performed by Alpgen. **Alpgen** handles these decays internally to prevent the mismodeling of several angular distributions of these particles, as the information which ensures the correct treatment of the spin correlations between the outgoing particles cannot be stored in the output file format.\(^\text{10}\)

Overall, **Alpgen** has proven to give a reasonable description of the shapes of various distributions. However, the normalization of these distributions is often not well modeled; most LO generators suffer from this problem. All **Alpgen** samples thus have to be properly normalized in order to correct for this effect. This is done by renormalizing their overall cross section to the NLO or NNLO cross section. The relative contribution of heavy flavor quarks ($c$ and $b$), coming mostly from gluon splitting in the parton showers, is also not very accurately modeled. To correct these issues, a data-driven method is used to 'recalibrate' the normalization of the MC samples. More details about these corrections are given in chapter 4.

3.3.1 - Generating **Alpgen** samples

Generating **Alpgen** samples proceeds in several steps. First, the input parameters that are used for the event generation need to be determined. **Alpgen** has many input parameters, some of which do not apply to all processes. The parameters can be categorized into three groups:

1. Process specification parameters. Other parameters that need to be set include the collision type (proton-proton or proton anti-proton) and the beam energy, the number of outgoing partons (parton multiplicity), and the random seeds to use.
2. Phase space parameters. These include various $p_T$, $\Delta R$ and $\eta$ cuts on particles. Three examples are:
   - 'ptjmin' is the minimum $p_T$ of a light additional parton (called a light jet by **Alpgen**).
   - 'drbmin' is the minimum $\Delta R$ between two massive $b$-quarks.
   - 'etalmax' is the maximum absolute $\eta$ at which charged leptons are generated.
3. Non-physical parameters, or modeling parameters. These include parameters that are needed to avoid singularities or divergences caused by approximations and limitations within the ALPHA algorithm. The effect of these unphysical modeling parameters would ideally be negligible. Due to the LO nature of **Alpgen** (and other approximations) they

---

\(^{10}\) The particles that were decayed can easily be reconstructed, as the output order of the particles of **Alpgen** is fully deterministic. This is indeed done by a subsequent phase in the MC generation in order to reconstruct the complete truth tree.
often do have an effect on the final results, and thus their values need to be tuned to reflect the data. In section 4.2.4 a detailed study of the sensitivity to these parameters is presented.

Lastly, before running Alpgen, the number of events to generate per run has to be selected. Before generating the events, several small iterations (called warm-up iterations) are performed to obtain a phase space grid to optimize the bulk generation. Because the integral over the phase space that was previously shown in equation 2.11 is difficult to evaluate, it is often performed by using an average over a set of generated events, which converges to the integral. With a good sampling of the phase space, this approximation is more than adequate. The sampling is established by running a number of iterations over the phase space, using the outcome of the phase space density as a sampling probability in the next iteration. After a configurable number of iterations and a configurable number of events per iteration, the optimized grid is stored, and then the actual events can be generated. If not enough events were used in generating the grid (in other words, if the grid hasn't converged yet), the sampling of the phase space becomes inadequate, and the cross section calculated for the process will be biased. Additionally, the more complex the simulated process is, the more complicated the probed phase space will be. Such a phase space will require more warm-up events and iterations to probe properly. All together, it is quite difficult to determine the number of events needed for the warm-up iterations without closer study. The effect of an improperly generated grid is certainly not negligible, even though averaging over many runs reduces it. This effect has affected several official ATLAS MC samples. The samples used in this thesis are not affected by it.

After the generation of ME processes, the events have to be further processed by the parton showerer. Alpgen can be interfaced with two showering methods: Herwig or Pythia. Currently, only the Fortran version of Herwig and Pythia version 6 are supported. The showerers also need to perform jet matching in order to prevent double-counting certain events: this is described in more detail in section 3.3.3. This procedure is run in the Athena framework using joboptions (which are explained in section 5.4) that also include Photos[97] (to handle photon conversions) and Tauola[98] (to handle tau-decays).

3.3.2 - Sample definitions

In this thesis, Alpgen is used to generate various different Monte Carlo samples. Most of these samples are $W^+\text{jets}$ (see section 3.2.3).
Monte Carlo samples and generators

It should be noted that for the analyses in this thesis, the contribution of the Z+jets samples is very modest as compared to W+jets. Therefore, the Z+jets events are not discussed in detail, but they are generated and processed in a similar way as the W+jets samples.

Normally, when an event is generated that produces more hard jets after parton showering than the parton multiplicity of the sample, the event is rejected in order to prevent double-counting (it overlaps with an event in a higher parton-multiplicity sample). However, if the sample in which this event is generated is the highest parton multiplicity, there is no risk of an overlap, and these events can be kept (under the condition that the extra jets are softer than any other jets in the sample, to prevent overlap with other events in the sample itself). This means that the highest parton-multiplicity sample is inclusive: it contains events with an even higher parton-multiplicity.

**Alpgen** treats all partons as being massless, with the possible exception of only one quark-flavor. However, the mass of both b- and c-quarks need to be taken into account for a proper description of the physics. For this reason, multiple separate samples are generated and combined to form the full set of samples that represent the W+jets process. In total, there are four processes that make up the full W+jets set: W+light jets (Wq\textsuperscript{12}), W+bb (Wbb), W+cc (Wcc), and W+c (Wc). In the case of Wq, only up to six u-, d-, s- or c-quarks, all treated massless, contribute to the ME. The Wbb sample contains ME's with two massive b-quarks, and up to four additional massless u-, d-, s- or c-quarks. Similarly, the Wc (Wcc) sample's ME has one (two) massive c-quark(s), and up to five (four) additional massless u-, d-, s- or c-quarks (the Wc sample doesn't contain any additional c-quarks). Only the combination of all these samples forms a consistent set of events. When the Wq and these heavy flavor (HF) samples are combined, the overlap between the samples has to be carefully removed. As parton showers underestimate the heavy quark production with high transverse momentum, the events generated by **Alpgen** in this region of the phase space are used from the Wbb, Wcc and Wc samples (the precise phase space cuts used in the generations for the different W+jets samples are listed in appendix A). This is done by a combination of shower matching (discussed in section 3.3.3) and heavy flavor overlap removal (discussed in section 3.3.4).

Taking the mass of the heavy flavor partons into account is important, because the mass does have a significant impact on the showering and hadronization of the heavy flavor parton. An accurate description of the contribution of HF quark production is important, because later we will rely on rather heavily on a technique called b-tagging, that identifies jets originating from a b- (or c-)quark (see section 6.4). Generating the four different samples mentioned above and combining them guarantees that the effect of the mass of both b- and c-quarks is handled properly.

A separate Wb sample is not used, because events with a W and b in the final state are included indirectly via the Wbb and the single top sample.\textsuperscript{13} The Wb-like events from

\textsuperscript{12} Sometimes also called Wqq or Wll, even though there may not be two light jets present in the events.
Chapter 3

Figure 3.1: A Feynman diagram of a Wbb event as generated by Alpgen in the Wbb sample, with the two $b$-quarks being created by gluon splitting. If one of these $b$-quarks goes undetected with no impact on the event, this event could be seen as a Wb-like event.

the Wbb sample stem from events where the (anti) $b$-quark escapes detection through the beampipe, see figure 3.1. (This discussion does not hold for the Wc sample, as there is a significant contribution of single $c$-quarks being extracted directly from the proton.)

The other source of Wb events is single top production, where a $b$-quark is produced through a $t \to bW$ vertex. There are other possible production methods that do not involve top-quarks, but these have a cross section that is lower than the uncertainties of the LO approach. For these reasons, support for these kinds of processes was not build into Alpgen, and the absence of a dedicated Wb sample has no discernible impact on the full collection of $W^\pm$-jets events.

3.3.3 - Jet-parton matching

The parton shower is simulated with the ME's $Q^2$-scale lower cut-off as an upper limit, but special care needs to be taken to avoid double-counting. The parton shower produces partons that mimic those that are produced by the ME generator; they can occupy the same phase space. For example, an event that was generated by Alpgen as a Wq event with 3 partons can emit an additional high-energy parton in the parton showering, producing an event that is identical to an event produced directly by Alpgen in the Wq with 4 partons process. This leads to double-counting of the production of partons in the final state, when these samples are added to obtain an inclusive sample of $W$-bosons with jets.

In order to avoid this double-counting, a special prescription is applied at the interface between the ME generator and the PS code. This kind of process is called jet matching, as it matches the jets produced by the PS to the partons produced by the ME in order to

13. Another reason is because Alpgen does not extract $b$-quarks from the proton through the PDF.
Figure 3.2: This figure illustrates the separation in phase space between ME and PS generated radiated partons. If a parton is radiated with a \( \Delta R \) larger than \( R_{clus} \), it is forced to originate from the ME. Otherwise, the event containing the parton generated by the PS will be selected. Illustration adapted from [25].

To avoid overlap between the samples. After the matching through this procedure, and when an overlap occurs with another event, this event is removed from the sample. There are various different jet matching schemes available. We introduce only the MLM matching method,[99-101] as this is the scheme used with Alpgen.

MLM matching uses a shower veto to prevent the double-counting of events. This means that some generated events will be dropped from the final samples, but has the benefit that it can be used with any parton showerer and does not need special implementations for interoperability between the ME generator and the parton showerer.

After all the partons have been showered by the parton showerer, the events are at so-called 'parton level'. The events have to be reweighted for a proper description of the physics. For this, MLM jet matching uses CKKW \( \alpha_s \) reweighting.[102-103] A branching history is created by constructing parton-jets through a Kt-algorithm, merging final state particles one-by-one into larger and larger jets. The presumption is that the energetic wide-angle emissions must have happened earlier, and thus running this clustering algorithm backtraces the branching history.

At each branching vertex in the history, the value of \( k_{perp} \) (a measure of the transverse energy of two partons; see the 'ktfac' variation in section 4.2.2) is calculated. This value acts as a scale for recalculating the \( \alpha_s \) at that particular vertex. The event is then reweighted based on these new \( \alpha_s \) values to renormalize these events.

Normally, after the \( \alpha_s \) reweighting the full CKKW prescription dictates a compensation for the overlap by additionally reweighting events with Sudakov factors in such a way that
combining the overlapping events does not introduce double-counting. Instead, in MLM matching this additional reweighting is replaced by a procedure that rejects overlapping events using a jet cone algorithm. The cone algorithm is defined by the following parameters:

- $R_{\text{clus}}$ for the size of the cone. This parameter is illustrated in figure 3.2.
- $E_{T,\text{clus}}$ for the minimum transverse energy of the jet.
- $\eta_{\text{max}}$ determines a cut-off in absolute pseudo-rapidity for the partons that are used in the jet reconstruction.

These three parameters together are called the 'matching parameters'. They are set in Alpgen and are passed through to the parton showerer. The values used for the matching parameters in ATLAS are set as a function of Alpgen input parameters, according to the following prescription:

$$E_{T,\text{clus}} = \text{max}(\text{ptjmin} + 5, 1.2 \times \text{ptjmin}) ,$$  

(3.1)

$$R_{\text{clus}} = \text{drjmin} ,$$  

(3.2)

$$\eta_{\text{max}} = \text{etajmax} .$$  

(3.3)

The samples used in this thesis follow the MLM matching parameters as given above. Typical values are 20 GeV for $E_{T,\text{clus}}$, 0.7 for $R_{\text{clus}}$, and 6 for $\eta_{\text{max}}$. The jet algorithm is performed on a virtual calorimeter, which is an evenly spaced grid in $\eta$ and $\phi$. This virtual calorimeter consists of uniform bins: 60 in the $\phi$ direction, and 100 bins in the pseudo-rapidity range up to $\eta_{\text{max}} + R_{\text{clus}}$. The procedure works by first summing the energy deposited by all final state particles (excluding top-quarks, leptons, and prompt photons) on the grid. Then the jet algorithm finds the calorimeter cell with the highest (remaining) $E_T$, and constructing a cone around that; the cells that (partially) fall within the $R_{\text{clus}}$ are all summed together, forming a jet candidate. The selected cells are then cleared, so they cannot be re-used by another jet. The next most energetic cell is found, and the algorithm then repeats until no more pseudo-jets can be constructed. All pseudo-jets that pass the $E_T$ and $\eta$ requirements ($E_{T,\text{clus}}$ and $\eta_{\text{max}}$) are promoted to jets and are stored for further processing.

Now that jets have been reconstructed, the actual jet matching can be performed. Initially, all final state showered partons produced by light ME partons are used in the jet cone algorithm. Once all these jets have been constructed, all the ME light partons are matched to the jets of showered partons: starting with the hardest (light) ME parton, if it

---

14. A prompt particle is a particle that is produced by the hard scattering directly.

15. Where $E_T$ is defined as: $E_T = \frac{E}{\cosh(\eta)}$. 
Monte Carlo samples and generators

is closer than 1.5 times \( R_{\text{clus}} \), the ME parton is considered matched and the jet is marked as matched. This is continued until all partons are matched. After the procedure, if any unmatched ME partons or unmatched jets remain, the event is rejected. The event is also rejected if the number of matched jets is not identical to the ME parton multiplicity of the sample (this rejection is what prevents the double-counting).

Secondly, all particles produced by HF-partons from the ME are reconstructed into jets. Contrary to the light ME partons, the 'drjmin' parameter is used as the matching criterium, instead of 1.5 times \( R_{\text{clus}} \). Also, during the matching, multiple HF-partons are allowed to match to a single jet (this to allow for small angles between HF-partons). Once the HF matching has completed, the event is rejected if an unmatched jet remains, or if additional jets were produced and the sample is not inclusive. Once again the extra jets are required to have a lower transverse momentum than the matched jets, otherwise the event is rejected.

The MLM matching procedure creates a clear separation in phase space between the region where the ME produces partons and where the PS can produce its particles. Because a hard scattering generator is better in describing the hard \( p_T \) objects, the 'ptjmin' value is set such that it at least covers the entire observable phase space. This means that the 'ptjmin' cannot be set higher than 5 GeV below the cut on the \( p_T \) of jets. Similar arguments hold for the 'drjmin' setting; the parton showerer provides a better description for the collinear particles, and the 'drjmin' value is chosen such that the partons generated by Alpgen roughly correspond to the radius of the jets used in the analysis, while the parton showering code mainly generated the (softer) radiation within the jets. These boundaries in the phase space are schematically illustrated in figure 3.3.

3.3.4 - Heavy flavor overlap removal

In the previous section, we discussed how overlap, or double-counting, of events is removed from an individual sample through MLM matching. To improve the accuracy of the predictions, the \( W^+ \)jets process is subdivided in several samples, namely the aforementioned Wq, Wbb, Wcc and Wc samples. However, these samples have double-counting as well, i.e. they overlap. For example, a Wq event can produce \( b \)-quarks in the parton showering phase that overlaps with an event produced in the Wbb sample, as illustrated in figure 3.4. The subject of this section is the removal of this double-counting.

The overlap between the various \( W^+ \)jets samples results in a confusion of the classification of the event. For example, an event in the original Wq sample that contains two \( b \)-jets, should eventually be categorized as an Wbb event. Hence, all the events in the various \( W^+ \)jets samples after the full parton showering and underlying event steps will need to undergo the

---

16. The exception to this is if the sample being matched is an inclusive sample, in which case additional jets are allowed, because in this case, there is no possibility for double-counting with another parton multiplicity sample. However, these extra jets are required to have a lower transverse momentum than the matched jets, again to prevent double-counting.

17. The motivation for this choice is not well documented.
Chapter 3

Figure 3.3: Schematic representation of the phase space boundaries between the ME and PS. The smallest $\Delta R$ between a jet and any other is set on the vertical axis, and the $p_T$ of that jet is set on the horizontal axis. The grey area indicates the phase space that is best described by the PS, and the white area the phase space best described by the ME. The 'ptjmin' and 'drjmin' parameters, for which the default values (15 GeV and 0.7 respectively) define the regions of phase space for either the ME or the PS.

Figure 3.4: A Feynman diagram that can be generated in both the Wq sample (through parton showering) and Wbb sample (both by the matrix element generation, and the parton showering), resulting in this diagram potentially being double-counted.

Heavy Flavor Overlap Removal (HFOR) procedure,[106] in order to re-evaluate what sample the events belong to. All events processed by the HFOR-tool are flagged with one (and only one) of the flags listed in table 3.1. The 'kill' flag indicates the events that are removed in order to avoid double-counting.
HFOR first searches for all $b$- and $c$-quarks among the (intermediate) partons in the branching history of the event. The resulting collection at this stage contains multiple instances of $b$- and $c$-quarks that after gluon radiation proceeded with different momentum. The information of the branching history is used to select only the last unique $b$- or $c$-quark in each branch.

The remaining quarks are classified as originating from different processes: the hard scattering, gluon splitting, multiple parton interaction, top-quark decay, or directly from the PDFs. Only quarks originating from the hard scattering and gluon splitting are used in the event classification; all other quarks are ignored.

The flowchart in figure 3.5 summarizes the classification rules. In the flowchart $b_{ME}$ represents the $b$-quarks coming from the matrix element (hard scattering), $c_{GS}$ the $c$-quarks coming from gluon splitting, etc. The various 'collinear'-branchings indicate whether two particles are close to each other, i.e. $\Delta R < 0.4$, whereas the 'Any'-branchings indicate whether the named type of particle is present in the event at all. The final result is that each event in the $W^+\text{jets}$ samples is unambiguously flagged by the HFOR-tool.\(^\text{19}\)

---

\(^{18}\) The Unknown-flag is only used if the used sample cannot be identified as requiring HFOR-flags by the HFOR-tool, or if the classification scheme used in the HFOR-tool fails to be conclusive. Neither scenario should ever happen when using $W^+\text{jets}$ samples.

\(^{19}\) The double-counting from a very small part of phase space is not properly removed, resulting in a small over-estimation of the cross section in the HF flagged regions. This effect is currently believed to be negligible, but will be investigated more closely by ATLAS in the future.

---

\(\text{Table 3.1: A list of the different HFOR-flag values with which an event gets flagged by the HFOR-tool. Certain restrictions apply to which truth quarks qualify for use in the flagging; this is explained in the text in more detail.}\)
Chapter 3

Figure 3.5: A flowchart detailing the decision making process of the HFOR-tool. The HFOR-flag with which the event is tagged is given at the end of each path, where 'kill' signals that the event should be dropped to avoid double-counting.

In the end the samples form a consistent set of events without any double-counting, while simultaneously the best physical description is selected for the process in each overlap region. For example, the overlap of $b$-quarks produced in the Wq and Wbb samples ($b_{ME}$ from the ME and $b_{GS}$ from the parton shower), is removed by rejecting events in the Wbb sample where the two ME $b$-quarks satisfy the $\Delta R < 0.4$ cut. These events are instead
selected from the Wq sample's parton showered $b$-quarks, because parton showering gives a better description of quarks that are radiated collinearly.

The HFOR-tool contains an option to apply a cut on the minimum $p_T$ that partons need to have in order to be considered for the HFOR-tool decision. This value is set to 0 GeV in the current implementation, but it has been shown that this introduces discontinuities between the samples. As the impact of these discontinuities on the overall cross section is small, this is not considered a problem.

In figure 3.6 two examples of HFOR-flag distributions are shown for various jet-multiplicities. In figure 3.6a (figure 3.6b), the different HFOR-flags for the Wq (Wbb) sample where the $W$-boson decays into an electron and electron-neutrino are shown (after applying the cuts that are described in section 6.5). As can be seen, the higher the jet-multiplicity, the more overlapping events are removed, and more events in the Wq sample are flagged as 'BB'- and 'CC'-events. The overlap between the samples is substantial, which demonstrates the need for the HFOR-tool.

3.3.5 - Summary

We presented the various event generators, and discussed in detail the generation of $W$+jets events with the Alpgen generator. The generation of these events is challenging because double-counting between the production of jets by parton showers and matrix elements has to be avoided. The production of dedicated samples with heavy flavor jets was also discussed. These samples are essential for analyses that use so-called $b$-tagging, as will be the case in this thesis. The combination of these events requires special attention as they
Chapter 3

contain a class of similar events, again due to the overlap of parton showering and the matrix element technique. To remove this overlap, the HFOR-tool was used. After all this work consistent $W^+$jets sample are available, although they are not yet normalized correctly. In chapter 4 we discuss the normalization of the $W^+$jets sample and its Heavy Flavor contents.
One of the main backgrounds to the top-quark analyses described in this thesis is $W^+\text{jets}$.\[107\] This is a process described by the Standard Model where two partons from the incoming protons interact to form a $W$-boson and a number of additional partons that create jets. As we discussed previously in section 3.2.3, we focus on $W^+\text{jets}$ events in which the $W$-boson decays leptonically. The resulting $W^+\text{jets}$ event topology thus contains at least one high energetic lepton, missing energy due to the escaping neutrino, and a number of jets. In general, this event signature closely matches that of top-quark production, especially when one or more of the resulting jets is identified to contain a $b$-quark.

In this thesis, the Alpgen $W^+\text{jets}$ samples are used, which includes the generation of up to five additional partons. It turns out that Alpgen adequately describes several distributions measured in data, but the predicted total cross section is only accurate to the level of 20%. Furthermore, the relative normalization of $b$- and $c$-quarks with respect to the light parton emissions has uncertainties of at least 60%.\[108\] These different types of heavy flavor (HF) events have different probabilities to identify $b$-quarks. Therefore the normalization of the $W^+\text{jets}$ for the HF samples ($W_{bb}$, $W_{cc}$, $W_c$) and the light parton ($W_q$) sample has to be determined from data itself.

In this chapter, we discuss a data driven method that extracts the normalization separate for each $W^+\text{jets}$ sample and as function of the number of jets. This method is described in section 4.1 and has been developed for $t\bar{t}$ analyses. We discuss the Alpgen generator uncertainties which affect the method in detail in section 4.2. The results presented here are used in several publications in ATLAS containing the analysis of $t\bar{t}$ production at 7 TeV and 8 TeV.\[109-110\]

The method is also used in the single top t-channel analysis as discussed in section 6.6.
4.1 - Data-driven $W$+jets estimation method

Data driven estimates of background contributions rely on the assumption that the distributions of most quantities are reasonably well described by the MC simulation, while its normalization can not be trusted. The correction factor is usually called the $k$-factor, defined by:

$$
k \equiv \frac{N_{\text{sel, data}}}{N_{\text{sel, MC}}}, \tag{4.1}
$$

where $N_{\text{sel, data}}$ represents the number of events after a certain selection to obtain a sample enriched with events of interest. Often, small additional corrections to $N_{\text{sel, data}}$ have to be applied to remove contributions from unwanted other (background) processes. The predicted number of events for the selection, $N_{\text{sel, MC}}$ uses the original normalization of the MC generator. Hence, the $k$-factor corrects not only for the MC normalization, but also for possible mismatches in efficiencies between data and MC detector simulation or luminosity.

For the $W$+jet events, we need to extract four different factors: $k_{bb,i}$, $k_{cc,i}$, $k_{c,i}$, and $k_{q,i}$, corresponding to the Wbb, Wcc, Wc, and Wq samples respectively, separately for each jetbin $i$.\footnote{In this thesis we use the term "jetbin" to denote the events with that specific number of reconstructed jets in the event.} As we will see, we cannot extract so much information for each jetbin and therefore simplifications have to be made. For convenience we first introduce in section 4.1.1 the relative contribution of each sample, called the HF fractions. In section 4.1.2 the procedure to extract the HF fraction from data is discussed, and the final step that results in the $k$-factors is presented in section 4.1.3.

4.1.1 - Heavy flavor fractions

We define the Heavy Flavor (HF) fractions $F_{x,i}$ in $W$+jets events as:\footnote{\[111\]}

$$
F_{x,i} = \frac{N_{x,i}}{N_{q,i} + N_{bb,i} + N_{c,i} + N_{cc,i}}. \tag{4.2}
$$

In this expression $F_{x,i}$ denotes the $x$-flavor fraction, where $x$ represents light (q), bb, c, or cc, indicating the sample's flavor type, and $i$ denotes the number of jets. By construction, all the flavor fractions in a single jetbin (a single value for $i$) add up to unity:

$$
F_{q,i} = 1 - F_{bb,i} - F_{cc,i} - F_{c,i}. \tag{4.3}
$$
To measure the HF fractions in data, a region of phase space has to be chosen where the top quark signal is small. This requirement excludes the 3- and 4-jetbin (and higher jetbins) in which the number of $\bar{t}t$ events is significant. Hence, we determine the HF fractions in the 1- and 2-jetbin and use information from Monte Carlo to extrapolate these to higher jetbins. The extrapolation factor to obtain the heavy flavor fractions in other jetbins is defined as:

$$R_{x,2 \rightarrow i} = \frac{F_{x,i}^{MC}}{F_{x,2}^{MC}} \quad (4.4)$$

where the $F_{x,i}^{MC}$ in this equation is the HF fraction taken from MC simulation. In this way we circumvent the need to determine the HF fraction in data originating from $W^+jets$ production in events containing three or more jets. As is evident from this construction, the factor $R_{x,2 \rightarrow i}$ is sensitive to theoretical uncertainties in the production of $W^+$jet events, which have to be evaluated carefully.

### 4.1.2 - Determination of the HF fractions

For each jetbin $i$ a relation exists between the number of event before $b$-tagging is applied (pre-tag), and after it has been applied (tagged):

$$N_{\text{tag},i}^W = N_{\text{pretag},i}^W \left( F_{q,i} P_{q,i} + F_{bb,i} P_{bb,i} \right) + F_{c,i} P_{c,i} + R_{cc-to-bb,i} F_{bb,i} P_{cc,i} \right) \quad (4.5)$$

where $P_{x,i}$ is the $b$-tagging probability given the event signature with flavor type $x$. These $b$-tagging probabilities are taken from Monte Carlo simulation.

To extract the number of (pre-)tagged $W^+$jets events from data $N_{(\text{pre-})\text{tag},\text{data}}^W$, the contribution of the non-$W$ events is subtracted:

$$N_{(\text{pre-})\text{tag},\text{data},i}^W = N_{(\text{pre-})\text{tag},\text{data},i}^W - N_{(\text{pre-})\text{tag},i}^{\text{non-}W} \quad (4.6)$$

where $N_{(\text{pre-})\text{tag},i}^{\text{non-}W}$ represents all (pre-)tag events with $i$ reconstructed jets that do not originate from $W^+$jets production. These contributions are either estimated from data (QCD) or MC simulation (top-quark events, $Z^+$jets, and diboson events).

The usage of equation 4.5 and the fact that the sum of the HF fractions equals unity (equation 4.3) in both 1- and 2-jetbin leads to four equations with a total of eight unknown quantities: $F_{x,i}$ (with $x = bb$, $cc$, $c$, $q$ and $i = 1, 2$). Hence, the use of additional MC information is
required, which has to be selected carefully to prevent spoiling the data driven technique. Below, we describe the method used to obtain this additional information.

Due to the similarity in the production of $W_{bb}$ and $W_{cc}$, they are assumed to have similar behavior. Therefore, we do not estimate $F_{cc,i}$ and $F_{bb,i}$ separately, but use the Monte Carlo simulation to predict their relative yields for each jetbin through $F_{cc,i} = R_{cc-to-bb,i} F_{bb,i}$, where $R_{cc-to-bb,i}$ is taken from MC. This reduces the number of unknown quantities by two (one per jetbin).

By using two factors derived from Monte Carlo: $R_{bb,2\rightarrow1}$ and $R_{c,2\rightarrow1}$ for the $bb$ and $c$ jet sample respectively, the number of unknown fractions is further reduced to four: $F_{q,1}$, $F_{q,2}$, $F_{bb,2}$ and $F_{c,2}$. These are determined in a numerical $\chi^2$ minimization fit using MINUIT. In the fit model $F_{q,1}$ and $F_{q,2}$ are obtained analytically through equation 4.3 to ensure that the HF fractions add up to unity in each jetbin. The fractions $F_{bb,2}$ and $F_{c,2}$ are the fit parameters. Note that these fractions defined for the 2-jetbin also determine the HF fractions in the first jetbin by construction.

The $\chi^2$ contains the ratio of tagged and pre-tagged events in the two jetbins:

$$\chi^2 = \sum_{i=1}^{2} \left( \frac{N_{W_{tag,MC,i}} W_{tag,MC,i}}{N_{W_{pretag,MC,i}} W_{pretag,MC,i}} - \frac{N_{W_{tag,data,i}} W_{tag,data,i}}{N_{W_{pretag,data,i}} W_{pretag,data,i}} \right)^2 \sigma_{data,i}^2,$$  (4.7)

where $\sigma_{data,i}$ is the statistical uncertainty on the observed ratio. The $\chi^2$ depends implicitly on the fit parameters via $\frac{N_{W_{tag,MC,i}} W_{tag,MC,i}}{N_{W_{pretag,MC,i}} W_{pretag,MC,i}}$ as can be directly inferred from equation 4.5. The corresponding ratio is obtained for data by using equation 4.6.

Since this method depends on the Monte Carlo description, we have to take systematic uncertainties of this description into account as well. This discussion is reserved for a later section in this chapter.

4.1.3 - $k$-factors

In the previous section we presented the analysis strategy used to determine the heavy flavor fractions of $W+\text{jets}$ while our objective was to normalize the $W+\text{jet}$ samples, separate for each jetbin. This final step we make in this section.

The overall normalization for the pre-tagged $W+\text{jets}$ events is simply obtained by using the pre-tagged event ratio between data and MC simulation in each jetbin:
Data driven $W$+jets prediction

\[ \text{Norm}_i = \frac{N_{W_{\text{pre-tag, data},i}}}{N_{W_{\text{pre-tag, MC},i}}} . \]  \hspace{1cm} (4.8)

With $\text{Norm}_i$ and the previously derived HF fractions, a $k$-factor for each sample and jetbin can be calculated:

\[ k_{xx,i} = \frac{F_{xx,i}}{F_{xx,i}^{\text{MC}}} \times \text{Norm}_i . \]  \hspace{1cm} (4.9)

The $\frac{F_{xx,i}}{F_{xx,i}^{\text{MC}}}$ ratio in this expression contains the correction from MC simulation to data for a particular HF fraction. The $\text{Norm}_i$ ensures that the absolute number of $W$+jets events of the MC prediction equals data. When these $k$-factors are applied (as weights) to their corresponding $W$+jets samples, the predicted number of events is correct for the pre-tagged sample and for the sample after $b$-tagging.

4.2 - Generator uncertainties

In this section we discuss the theoretical uncertainties of the Alpgen generator. These uncertainties affect the extraction of the heavy flavor fractions via the factor $R_{x,2 \rightarrow i}$, and have to be taken into account in further physics analysis. We determine the theoretical systematic uncertainties of $R_{x,2 \rightarrow i}$ from various parameter variations of the Alpgen generator. For each parameter variation, we generate a separate Alpgen sample.

When a parameter for a Alpgen Monte Carlo sample is modified, the event kinematics of the produced events will in general be different. In other words, the event selection has to be applied, on each sample separately, before the impact of the variations can be properly evaluated. This study uses the 7 TeV $t\bar{t}$ selection applied to the truth-level of the MC objects, meaning that no detector effects have been modeled. The results obtained in this section of the thesis are used by the ATLAS collaboration as part of the systematic uncertainties in the production of top quarks for both the 7 TeV and 8 TeV measurements.

Below, we first briefly describe the event selection (section 4.2.1), followed by an overview of the relevant MC parameters to be varied in section 4.2.2. In section 4.2.3 and section 4.2.4 the results are presented.
4.2.1 - Signal region cuts

The selection presented here is applied on truth-level. The reconstruction and selection of the events thus require a different strategy than for data and detector-level MC events.

Physics objects are 'reconstructed' by using the final state truth particles after parton showering and hadronization. Jets are reconstructed by using the AntiKt 0.4 algorithm, see section 6.3. The neutrino content of the event leads to the missing transverse energy, $\not{E}_T$, which is calculated from the vector sum of all the neutrinos in the event. The transverse mass of the $W$-boson, $m_T^W$ is reconstructed from the the $\not{E}_T$, the electron (or muon), and the $b$-jet. The $b$-jet is tagged by checking whether there was a $b$-quark involved in the production of the final state particles that form the jet. Finally, a procedure called overlap removal is used in order to remove overlap between jets and leptons. Overlapping jets are removed if a cone around this jet contains any reconstructed lepton.

The complete selection requirements for the electron and muon streams are summarized in table 4.1. This selection on truth-level mimics the selection that would be used by a top-quark analysis at the detector-level as closely as possible. The differences come mainly from the different definition of physics objects, not the cuts themselves.

<table>
<thead>
<tr>
<th>Electron stream</th>
<th>Muon stream</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>\eta_{\text{jet}}</td>
</tr>
<tr>
<td>$p_{T\text{jet}} \geq 25 \text{ GeV}$</td>
<td>$p_{T\text{jet}} \geq 25 \text{ GeV}$</td>
</tr>
<tr>
<td>Electrons:</td>
<td></td>
</tr>
<tr>
<td>$</td>
<td>\eta_{\text{el}}</td>
</tr>
<tr>
<td>$p_{T\text{el}} &gt; 25 \text{ GeV}$</td>
<td>$p_{T\text{el}} &gt; 20 \text{ GeV}$</td>
</tr>
<tr>
<td>Muons:</td>
<td></td>
</tr>
<tr>
<td>$</td>
<td>\eta_{\text{mu}}</td>
</tr>
<tr>
<td>$p_{T\text{mu}} &gt; 20 \text{ GeV}$</td>
<td>$p_{T\text{mu}} &gt; 25 \text{ GeV}$</td>
</tr>
<tr>
<td>$\not{E}_T &gt; 35 \text{ GeV}$</td>
<td>$\not{E}_T &gt; 20 \text{ GeV}$</td>
</tr>
<tr>
<td>$m_T^W &gt; 25 \text{ GeV}$</td>
<td>$\not{E}_T + m_T^W &gt; 60 \text{ GeV}$</td>
</tr>
<tr>
<td>$N_{\text{el}} &gt; 0$</td>
<td>$N_{\text{el}} = 0$</td>
</tr>
<tr>
<td>$N_{\text{mu}} = 0$</td>
<td>$N_{\text{mu}} &gt; 0$</td>
</tr>
</tbody>
</table>

Table 4.1: Summary of the selection cuts used in this chapter.
4.2.2 - Parameter variations

In this section we describe the variations of the Monte Carlo parameters that we consider to estimate the theoretical uncertainties of the Alpgen generator.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Default value</th>
<th>Down</th>
<th>Up</th>
</tr>
</thead>
<tbody>
<tr>
<td>drjmin</td>
<td>0.7</td>
<td>0.4</td>
<td>1.0</td>
</tr>
<tr>
<td>ptjmin</td>
<td>15 GeV</td>
<td>10 GeV</td>
<td>20 GeV</td>
</tr>
<tr>
<td>iqopt</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>qfac</td>
<td>1.0</td>
<td>0.5</td>
<td>2.0</td>
</tr>
<tr>
<td>ktfac</td>
<td>1.0</td>
<td>0.5</td>
<td>2.0</td>
</tr>
</tbody>
</table>

Table 4.2: The parameters of Alpgen that are varied to estimate the systematic uncertainties. Listed are the default values, and their variations. Note that the 'iqopt' variations are technically not up- and down-variations.

The opening angle between partons: drjmin

The minimum opening angle between any two of the outgoing light partons, $\Delta R$, is called 'drjmin'. The perturbative expansion diverges when the angle between the partons goes to zero. This is called the collinear divergence. As the parton showerer is not affected by this issue, the ME generation is restricted to only branchings of partons with opening angles larger than $\Delta R$. Below this opening angle the parton shower takes over from the matrix element generator.

The default value is $\Delta R = 0.7$, with the up-and-down variations being 0.4 and 1.0 respectively.

The minimum of the parton transverse momentum: ptjmin

The minimum $p_T$ of any outgoing light parton is determined by the parameter 'ptjmin'. Similar to the opening angle of two outgoing light partons, at low $p_T$ the matrix element prediction diverges for soft transverse momentum (soft divergence), and the parton showering algorithm will take over, as was discussed in more detail in section 3.3.3.

The default value of 'ptjmin' is set to 15 GeV, well below the analysis requirement on the $p_T$ of the jets of (at least) 25 GeV. If the 'ptjmin' parameter is equal or larger than the jet $p_T$ cut, a fraction of the jets would be generated fully by the parton showering instead of originating from the matrix element calculations. For systematic variations the 'ptjmin' value is varied to 10 GeV and 20 GeV.

Note that varying the 'ptjmin' has no effect on other $p_T$ related parameters, such as the minimum $p_T$ of the $b$- or the $c$-quark partons, 'ptbmin' and 'ptcmin'.
The scale of the matrix element: iqopt

Alpgen supports several different ways of calculating the energy scale of the hard
scattering of the event, the $Q^2$. This value is also used as the factorization and
renormalization scales ($Q^2 = \mu_F^2 = \mu_R^2$). The parameter ’iqopt’, an abbreviation for 'integer Q
options', controls the functional form of the energy scale calculation.
For Wq ($W +$ light jets) processes, 'iqopt' can take the following values (the definition for
each value for the heavy flavor samples is similar):

<table>
<thead>
<tr>
<th>iqopt</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q_0^2$</td>
<td>1</td>
<td>$m_W^2 + \sum p_T^2$</td>
<td>$m_W^2$</td>
<td>$m_W^2 + p_T W$</td>
<td>$\sum p_T^2$</td>
</tr>
</tbody>
</table>

Table 4.3: The possible options for 'iqopt'. The sums run over all the outgoing partons.
The default value used in this thesis is 'iqopt' = 1.

The default value is 'iqopt' = 1, and the values 2 and 3 are used as variations. Scheme 2
is a somewhat unrealistic option, since it doesn't actually describe the energy of the event
properly ($m_W$ being a constant per definition). Scheme 3 uses the $p_T$ of the $W$-boson,
instead of that of the partons produced by the ME.

The scale of the matrix element: qfac

To further vary the values of the energy scale, Alpgen offers the 'qfac' parameter which is a
simple multiplication factor for the $Q_0$ calculated by the 'iqopt' parameter: $Q = qfac \times Q_0$.
The default value is 1.0, and the variations are up-and-down with a factor of 2: ’qfac’ = 0.5, 2.0.

The scale of $\alpha_s$: ktfac

The strong coupling constant $\alpha_s$ is evaluated at an energy scale determined by the option
'ktfac'. Specifically, the 'ktfac' parameter is a multiplication factor for the (internal) 'kperp'
('k perpendicular') variable, which is used in the internal clustering algorithm for the jet
matching procedure.
The default value of 'ktfac' is 1.0, with an down and up variation with 0.5, and 2.0 respectively.

The PDF used for Alpgen samples is by default CTEQ6L1.$^{33-34,42-43,46-60}$ There are other
choices available, such as MSTW2008LO.$^{32-45}$ Varying the PDF parametrization set will
change the distribution of the parton flavors that are extracted from the proton. This may in
principle have a significant effect on the number of heavy flavor quarks produced. However,
our analysis obtains the HF fractions directly from data and is thus not directly sensitive to
variations of the PDF parametrization. Hence we do not to include the PDF parametrization
variation in our definition of systematic uncertainties at this point.
As Alpgen supports both Pythia and Herwig for the parton showering, the choice of parton showerer can also be varied. A simple comparison between the Herwig-showered and Pythia-showered samples\textsuperscript{21} was done in the context of the 8 TeV sample validation. Most distributions of observables are indistinguishable between the Pythia and Herwig samples. This is not the case for the $\eta$ distribution of the jets and the opening angles between the jets, which exhibits significant differences. These are presented in figure 4.1. The origin of these differences is not understood, and they mainly affect low $p_T$ jets. After applying a $p_T^{\text{jet}} \geq 20$ GeV cut, the difference largely disappears, as shown in figure 4.2. The double-bump shape seen in the distribution of the jet $\eta$ for the Alpgen+Herwig samples disappears, and the tail of the jet $\Delta R$ distribution also matches much better. Any remaining difference after the $p_T$ cut is covered by a systematic uncertainty in the analysis.

4.2.3 - Variations of the cross section

Generating the samples with all the systematic variations took the equivalent of slightly more than 100 CPU-years on a single contemporary high-end processor, and about 10 TB of storage space. Most events were generated on the Nikhef tier 1 Grid site,\textsuperscript{[114]} using 1000 of its cores, reducing the required processing time to about a month. The vast majority of these samples were made available to the ATLAS production system (see section 5.4.2) and are thus available for all ATLAS analyses.

\textsuperscript{21} The Alpgen+Herwig 8 TeV sample used here is not affected by an issue that was found in the first set of official samples, where $B$-hadrons would decay improperly.
Chapter 4

Figure 4.2: Two comparisons of Herwig- and Pythia-showered Alpgen samples at both 7 TeV and 8 TeV. (a) shows the $\eta$ distributions of jets, with a cut $p_T^{\text{jet}} \geq 20$ GeV. (b) shows the $\Delta R$ between all jet pairs in an event. Illustrations from [113].

Figure 4.3: The visible cross section$^{22}$ of all events (excluding 'kill'-flagged events to prevent double-counting) from all the $W^+\text{jets}$ samples combined as function of the jet multiplicity. The results for systematic variations are also shown.

22. The visible cross section is the full cross section without any cuts applied.
Figure 4.3 shows the full visible cross section of all the variation samples after the selection criteria have been applied, as a function of the number of reconstructed jets. Most of the up- and down-variations lie distributed around the nominal sample. In table 4.4 the largest deviations are listed.

<table>
<thead>
<tr>
<th>Jetbin</th>
<th>Variation</th>
<th>Deviation [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 → 0 jet qfac20</td>
<td>8.6%</td>
<td></td>
</tr>
<tr>
<td>2 → 1 jet ktfac05</td>
<td>18.8%</td>
<td></td>
</tr>
<tr>
<td>2 → 2 jet ktfac20</td>
<td>21.8%</td>
<td></td>
</tr>
<tr>
<td>2 → 3 jet ktfac05</td>
<td>49.2%</td>
<td></td>
</tr>
<tr>
<td>2 → 4 jet ktfac05</td>
<td>57.6%</td>
<td></td>
</tr>
<tr>
<td>2 → ≥5 jet ktfac05</td>
<td>64.4%</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.4: The largest cross section deviations per jetbin in percent.

Note that as the jet-multiplicity increases, the cross section decreases exponentially, expected from theoretical predictions called Berends-Giele scaling.\[115\]

The largest deviation from the nominal sample is caused by the pair of 'ktfac' variations, with effects of 50% and larger. The effect on the cross section from the variation of the parameter 'iqopt' is asymmetric. This can be expected as the variation itself is not symmetric.

The fluctuations, or better, uncertainties, shown in figure 4.3 confirm that a data driven approach to normalize the cross section is required. As explained above the approach we adopted is still sensitive to the factor \( R_{x,2\rightarrow i} \). The variations on this quantity are studied in the remainder of this chapter.

### 4.2.4 - Extrapolation factor uncertainties

In this section, we construct the factor \( R_{x,2\rightarrow i} \) from the cross sections from the various HFOR-samples. First, in figure 4.4 we present the variations of the cross section as function of the reconstructed jets, for events that contain light jets only, and events that contains \( c \) - and \( b \)-quark jets. These results are obtained using the HFOR-flag (see section 3.3.4).

The two largest effects come from the 'qfac' variation in the lower jetbins, and from the 'iqopt' variation in the higher jetbins.

The HF fraction is obtained with equation 4.2 and shown in figure 4.5. As can be seen, the fraction of events containing a \( b \)-quark jet ('BB'-flagged) grows with the jet-multiplicity of the event. More (high energetic) partons from the hard scattering causes more energy to be available in the parton showering, and this increased activity results in more creation of \( b \)-quarks. As expected, the behavior of the \( c \)-quark fraction is similar.
Figure 4.4: The visible cross section of the four types of HFOR-flagged events in the electron stream from all the \( W^+\text{jets} \) samples. (a) shows the 'BB' cross section; (b) shows the 'CC' cross section; (c) shows the 'C' cross section; and (d) shows the 'light' cross section.
Figure 4.5: The ratio of the cross section of the HFOR-flagged events over the total cross section of the relevant sample from all the samples in the electron stream combined. (a) shows the 'BB' ratio; (b) shows the 'CC' ratio; (c) shows the 'C' ratio; and (d) shows the 'light' ratio.
Figure 4.6: The HF-ratio of HFOR-flagged events from all the samples in the electron stream combined, as given by equation 4.4. Note that all points in the 2-jet bin are unity by definition. (a) shows the 'BB' HF-ratio; (b) shows the 'CC' HF-ratio; (c) shows the 'C' HF-ratio; and (d) shows the 'light' HF-ratio.
As discussed above, the factor $R_{x,2\rightarrow i}$ is used in the data-driven methods to normalize the $W^+\text{jets}$ background (through equation 4.4). Figure 4.6 shows the effect of the variations on the ratio $R_{x,2\rightarrow i}$.

<table>
<thead>
<tr>
<th>Extrapolation</th>
<th>BB</th>
<th>CC</th>
<th>C</th>
<th>light</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2 \rightarrow 0$ jet</td>
<td>$13.6^{+1.4}_{-1.6}$ (drjmin04)</td>
<td>$16.0^{+1.3}_{-1.2}$ (ptjmin10)</td>
<td>$9.5^{+0.46}_{-0.44}$ (ktfac05)</td>
<td>$3.95^{+0.019}_{-0.019}$ (qfac05)</td>
</tr>
<tr>
<td>$2 \rightarrow 1$ jet</td>
<td>$6.3^{+1.5}_{-1.1}$ (drjmin04)</td>
<td>$8.3^{+1.1}_{-0.9}$ (drjmin04)</td>
<td>$4.34^{+0.47}_{-0.46}$ (iqopt3)</td>
<td>$1.87^{+0.20}_{-0.21}$ (qfac05)</td>
</tr>
<tr>
<td>$2 \rightarrow 2$ jet</td>
<td>$1.0$ (-)</td>
<td>$1.0$ (-)</td>
<td>$1.0$ (-)</td>
<td>$1.0$ (-)</td>
</tr>
<tr>
<td>$2 \rightarrow 3$ jet</td>
<td>$4.3^{+2.0}_{-1.5}$ (ktfac05)</td>
<td>$3.9^{+1.5}_{-0.9}$ (ktfac05)</td>
<td>$6.1^{+0.76}_{-0.67}$ (iqopt3)</td>
<td>$1.5^{+0.36}_{-0.35}$ (drjmin10)</td>
</tr>
<tr>
<td>$2 \rightarrow 4$ jet</td>
<td>$7.4^{+3.9}_{-2.3}$ (qfac20)</td>
<td>$7.3^{+2.3}_{-1.5}$ ( ktfac05)</td>
<td>$10.4^{+1.1}_{-1.0}$ (iqopt3)</td>
<td>$2.4^{+1.0}_{-1.0}$ (drjmin10)</td>
</tr>
<tr>
<td>$2 \rightarrow \geq 5$ jet</td>
<td>$13.5^{+5.2}_{-4.8}$ (iqopt2)</td>
<td>$16.8^{+4.7}_{-4.8}$ (drjmin04)</td>
<td>$28.4^{+2.1}_{-2.1}$ (iqopt3)</td>
<td>$4.4^{+1.9}_{-1.8}$ (iqopt2)</td>
</tr>
</tbody>
</table>

Table 4.5: The largest deviations per jetbin for electron stream events in percent, with the statistical uncertainty on that percentage.

<table>
<thead>
<tr>
<th>Extrapolation</th>
<th>BB</th>
<th>CC</th>
<th>C</th>
<th>light</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2 \rightarrow 0$ jet</td>
<td>$14.2^{+1.6}_{-1.3}$ (ptjmin10)</td>
<td>$15.6^{+1.0}_{-0.9}$ (drjmin04)</td>
<td>$9.60^{+0.45}_{-0.45}$ (ktfac05)</td>
<td>$4.150^{+0.019}_{-0.019}$ (qfac05)</td>
</tr>
<tr>
<td>$2 \rightarrow 1$ jet</td>
<td>$6.0^{+1.1}_{-0.9}$ (drjmin04)</td>
<td>$8.2^{+0.83}_{-0.91}$ (drjmin04)</td>
<td>$5.12^{+0.45}_{-0.45}$ (iqopt3)</td>
<td>$1.8^{+0.21}_{-0.21}$ (qfac05)</td>
</tr>
<tr>
<td>$2 \rightarrow 2$ jet</td>
<td>$1.0$ (-)</td>
<td>$1.0$ (-)</td>
<td>$1.0$ (-)</td>
<td>$1.0$ (-)</td>
</tr>
<tr>
<td>$2 \rightarrow 3$ jet</td>
<td>$3.8^{+1.6}_{-1.7}$ (ktfac05)</td>
<td>$3.5^{+1.3}_{-1.3}$ (ktfac05)</td>
<td>$5.8^{+0.65}_{-0.67}$ (iqopt3)</td>
<td>$1.8^{+0.47}_{-0.47}$ (drjmin10)</td>
</tr>
<tr>
<td>$2 \rightarrow 4$ jet</td>
<td>$7.1^{+2.5}_{-2.6}$ (iqopt2)</td>
<td>$9.3^{+2.0}_{-1.8}$ (ktfac05)</td>
<td>$10.6^{+1.0}_{-1.0}$ (iqopt3)</td>
<td>$2.5^{+1.0}_{-1.0}$ (drjmin10)</td>
</tr>
<tr>
<td>$2 \rightarrow \geq 5$ jet</td>
<td>$10.4^{+4.6}_{-3.9}$ (ktfac05)</td>
<td>$16.3^{+3.9}_{-3.9}$ (drjmin04)</td>
<td>$23.4^{+1.7}_{-1.7}$ (iqopt3)</td>
<td>$5.5^{+1.6}_{-1.6}$ (iqopt2)</td>
</tr>
</tbody>
</table>

Table 4.6: The largest deviations per jetbin for muon stream events in percent, with the uncertainty on that percentage.

Not all variations are uncorrelated, hence summing all variations quadratically overestimates the total uncertainty somewhat. Therefore, we select the largest of the variations as the total systematic uncertainty for each jetbin separately. These largest variations in each jetbin are summarized in table 4.5 and table 4.6 for the electron and muon stream respectively. For most jetbins the uncertainty on this factor is about 10%.

We decided to obtain these results for electrons and muons separately because the event selection is slightly different, which could lead to different systematic behavior. However, the results for electrons and muons appear comparable within their statistical uncertainties.
These systematic effects have to be taken into account by top quark analyses where the $W^{+}$jets is a significant background.

4.3 - Conclusion

In this chapter we have described a data-driven method to determine the normalization and Heavy Flavor fractions of $W^{+}$jets background, using events form the 1- and 2-jetbin. We defined the 2-jetbin as the basis and extrapolate the results to other jetbins using the factor $R_{x,2\rightarrow i}$ based on MC information. This introduces systematic uncertainties, which are evaluated by the variation of the key parameters of the Alpgen generator, which involved the simulation of almost 2.5 billion events.

The results for the systematic uncertainties on $R_{x,2\rightarrow i}$ from the Alpgen MC generator are summarized in table 4.5 (electrons) and table 4.6 (muons). At this moment we can not simply predict the effect of these $W^{+}$jets variations in top quark measurements. That depends on various details, and therefore the effects of these variations have to be evaluated by each analysis individually by using the uncertainties presented in this chapter.

For most jetbins the uncertainty on these extrapolation ratio's is about 10%, which is a significant improvement with respect to the previous study\textsuperscript{[116]} that quoted 25%. A significant contribution to this improvement is the large statistical size of the samples used in the work presented here, which were obtained using the Nikhef-ATLAS tier 1 facility in Amsterdam.

For the top cross section measurement presented in this theses we use the results obtained for electrons (table 4.5) as only these were available at the time of the measurement. Since the results for electron and muons are similar, this does not introduce any additional uncertainty.
In this chapter, we describe the data taking and data processing of the ATLAS detector. We start this chapter with a description of the LHC collider and the ATLAS detector with its various subdetectors. Next, we discuss the trigger and data quality systems, followed by a description of the simulation of the ATLAS detector. In the final section of this chapter, we describe how the recorded events are processed by making use of a distributed computing network called the Grid.

5.1 - The LHC

At the end of the year 2000, the Large Electron-Positron collider (LEP) was shut down in order to make place for the Large Hadron Collider (LHC). LEP was build after the discovery of the W- and Z-bosons to allow precise measurements of the electroweak section of the SM. While LEP produced head-on collisions between an electron and a positron beam, the LHC collides two proton beams head-on. The former LEP and currently LHC tunnel is between 45 and 170 meters deep underground, crossing the France-Swiss border, and located to the north-west of Geneva, at the foot of the Jura mountains. The nearby Lac Léman (commonly and incorrectly referred to as the Lake of Geneva) provides a geological stabilizing effect, making the region a very good location for an underground accelerator. The tunnel itself was dug between 1983 and 1987, and has a circumference of 27 km. It consists out of several sections of curved and straight segments; see figure 5.1. At various locations along the tunnel, access shafts lead from the surface to caverns where the experiments are located. The accelerator itself runs through the caverns and tunnel, with the proton beams circulate in two separate beampipes containing an ultra-high vacuum. One beam goes in the clock-wise direction and the other beam anti-clockwise. A large part of the LHC accelerator structure is taken up by the superconducting magnets in order to produce the 8.33 Tesla magnetic fields that are used to curve the proton beams. The magnets become superconductive by cooling with liquid helium. Dipole superconductive
magnets keep the beams orbiting, while quadrupole and higher moment magnets are used to keep the beams focused, and can be used to correct and adjust the orbit. The protons are obtained from ionizing hydrogen, and are sent through several separate accelerators (including the PS and SPS, the predecessors to LEP), eventually injecting them into the LHC with an energy of 450 GeV. Once the LHC is filled, the protons are accelerated to a beam energy of 4 TeV, as used in this thesis. In the future the energy will be increased to near the maximum design beam energy of 7 TeV. After squeezing the beams (focusing them to the smallest possible size at the interaction points), the beams are allowed to cross each other at dedicated interaction points, providing the high-energy proton-proton collisions for experiments.
There are various experiments that observe the collisions produced at LHC:

- **ALICE**: An experiment geared towards the observation and recording of heavy ion collisions runs. The scientific challenge is to investigate the quark–gluon plasma (QGP) and other strongly interacting matter at extreme energy densities.

- **ATLAS**: The detector that we use in this thesis.

- **CMS**: A general-purpose detector, much like ATLAS.

- **LHCb**: An experiment focusing on $B$-hadron interactions, and the associated CP violation.

In this thesis we analyze collisions as recorded by the ATLAS detector. In this thesis, data from the 2012 data-taking period (the last year of Run I) is used. The LHC produced proton-proton collisions with centre-of-mass energy of 8 TeV during that year. The integrated luminosity of both the 2011 and 2012 data-taking periods are illustrated in figure 5.2. The efficiency of recording the produced collisions was high during this period, and the run conditions overall were such that 84% of all delivered collisions in 2011 and 89% in 2012 can be used for analyses.
Chapter 5

5.2 - The ATLAS detector

ATLAS (an acronym for A Toroidal LHC ApparatuS) is a general purpose particle detector. It is 46 meter long and has a diameter of 25 meter.\[125\] Situated at Point 1, it is located in the UX-15 cavern about 100 meter below ground. As shown in figure 5.3, it consists of various sub-detectors, each with their own design for recording a part of the collisions. The collisions take place in the beampipe in the center of the detector. Going radially outwards, the inner detector with a total radius of 1.15 m and 3.5 m long, measures the trajectories of charged particles in a 2 Tesla solenoidal magnetic field. This system is surrounded by the calorimeters; first the electromagnetic calorimeter, then the hadronic calorimeter with a total radius of 2.1 m. The calorimeters also cover the forward directions and extend to $z \approx 6.4$ m. Around that are three toroid magnets (one barrel, two end-caps) provide the magnetic field for the muon spectrometer, with an integrated field of about 2 Tm. In the next sections, these components are discussed in more detail.
### 5.2.1 - Coordinate system

In ATLAS, a right-handed coordinate system is used. Its origin is the center of the detector, the point where the proton-proton collisions take place. The conventional Cartesian coordinate system has its z-axis parallel to the proton beam, pointing in the counter clockwise direction when viewing the LHC from above. The x-axis points inwards to the centerpoint of the LHC ring, and the y-axis upwards.

Due to the geometry of the detector, a more natural choice is to use a cylindrical coordinate system. The z-axis lies in the same direction, with the azimuthal angle φ-coordinate laying in the x-y plane, with a range of (−π, π], with φ = 0 pointing in the x direction. The inclination θ is the angle with positive z-axis, running from [0, π).

Instead of θ, an alternative quantity, the rapidity y is often used. The rapidity of a particle with an energy \( E \) and a momentum of \( p_z \) along the z-axis is given by:

\[
y = \frac{1}{2} \ln \left( \frac{E + p_z}{E - p_z} \right).
\]  

(5.1)

For high energetic particles, when the particle’s mass can be ignore, the so-called pseudo-rapidity can be derived:

\[
\eta = -\ln \left( \tan \frac{\theta}{2} \right).
\]  

(5.2)

The pseudo-rapidity is convenient to use, as the expected particle flux as a function of pseudo-rapidity is approximately constant. Additionally, with this definition, a difference between particles (\( \Delta \eta \)) is invariant under Lorentz-boosts along the z-axis (for massless particles). This means that a particular distance measure \( \Delta R \) (which is used for various selections and reconstruction cuts) is insensitive to Lorentz boosts, where \( \Delta R \) is defined as:

\[
\Delta R \equiv \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2}.
\]  

(5.3)

### 5.2.2 - The inner detector

The inner detector (ID)\(^{126-127}\) consists of three main sub-detectors. From the beampipe outwards, these are the pixel detector, the semi-conductor tracker, and the transition radiation tracker. The purpose of the inner detector is to measure the spatial layout and the momentum of tracks of charged particles with great precision. The particle momenta are obtained from the track’s curvature caused by the solenoid magnetic field. The ID can resolve the tracks of over a thousand charged particles in a single collision.

The second task of the inner detector is the proper identification of the primary and
secondary collision vertices. For this reason, the first sensitive material of the ID is at only 5 cm distance from the center of the beam-line.

In figure 5.4 the inner detector is shown, with its sensitive area covering $|\eta| < 2.5$. Below, we describe the various subcomponents of the ID separately.

**The pixel detector**

The pixel detector consists of semiconducting solid silicon cells. Charged particles coming from the collision ionize the silicon atoms, creating electron and hole pairs. The freed electrons create, under the influence of an applied 150 V potential, a current through the cell. These cells are positioned in three cylindrical layers around the beam-line, at 5.05 cm, 8.85 cm and 12.25 cm distance. The innermost layer covers the full $|\eta| < 2.5$, with the other two layers covering only up to $|\eta| < 1.7$. At each end of the pixel detector are also three disks of silicon detectors, called the end-caps, to measure the charged particles too forward for the second and third layers. The layers are constructed from $50 \times 400 \, \mu m^2$ pixels while the disks have $50 \times 600 \, \mu m^2$ pixels, with the pixels slightly overlapping to prevent particles from passing through undetected. This results in a position resolution in the $r-\phi$ direction of about 12 $\mu m$. Together, these layers and disks with their high pixel granularity contribute
significantly to ATLAS’s ability to distinguish secondary from primary vertices (section 6.5.3).

**The semi-conductor tracker**
The semi-conductor tracker (SCT) surrounds the pixel detector. It uses strips of silicon of 50 μm wide, divided over 4 cylindric layers around the pixel detector at a distance of about 30 cm) to 51 cm from the beam-line, and 9 disks that function as end-caps. The goal of the SCT is to add more spacepoints to a track measured in the inner detector.
The SCT is designed so that each particle traversing the SCT within its $|\eta| < 2.5$ coverage will generate at least 8 hits. The strips in the layers (two each) are oriented with a stereo angle of 40 mrad with respect to each other. The hits on both strips are combined to resolve the coordinate along the strip, resulting in a position resolution of 16 μm in the $r$-$\phi$ direction.

**Transition radiation tracker**
The transition radiation tracker (TRT) consists of multiple layers of drift tubes with radii of 4 mm, called straws. They are positioned radially from about 55 cm to 108 cm around the barrel-region, running parallel to the beam-line. They cover $|\eta| < 0.7$, with additional disks as end-caps covering up to $|\eta| < 2.0$, where the straws are instead radially mounted. Each straw is a thin 4 mm hollow cylinder, with an anode wire running along its center, and filled with a gas mixture (Xe, CO$_2$, O$_2$). Charged particles ionize the gas and the free electrons drift to the wire, which is for this purpose set at 1530 V. On average, each charged particle traverses about 30 to 36 straws. The read-out of the resulting current is only done on one side of the straw. Therefore, no $z$-coordinate information is available, and thus no $\eta$-information is measured by the TRT.
The TRT adds more spacepoints to the measured tracks, that improves the efficiency of the pattern recognition used to reconstruct the tracks. Additionally, the difference between pions and electrons can be measured because polypropylene fibers are mounted with the straws, causing transition radiation when the charged particles traverse them. The produced photons get absorbed by the Xenon gas in the tubes. Because the energy of the transition photons relates to the mass of the charged particle, with the differences in energy of the radiation it is easy to distinguish both types of particles from each other. Also, the transition radiation is much more energetic than the ionization energy, so the two cannot be confused.
The TRT, SCT, and pixel detector combined result in a design momentum resolution on charged particles of: $\frac{\sigma(p_T)}{p_T} = 0.05\% \times p_T \text{ (in GeV)} + 0.1\%$.\(^{124}\)

### 5.2.3 - Calorimeters

The next set of sub-detectors are the calorimeters (CALs), which form a shell around the ID and its solenoid magnet. The CALs are illustrated in figure 5.5. The purpose of

---

23. A hit is the energy deposition of a particle in a sensitive region of the detector.
Chapter 5

Figure 5.5: A cutaway view of the calorimeter systems. Illustration from [124].

the calorimeters is to measure the energy of particles, including neutral particles, and to distinguish between hadrons and electromagnetic activity. The absorber in the calorimeter causes particles to shower, while the active material measures the produced charge or light depending on the active material.

In order to be effective, the calorimeters need full $\eta$ coverage; this is particularly essential for the calculation of the $E_T$ originating from neutrinos that escape detection. The barrel and end-cap calorimeters reach up to $|\eta| < 3.2$, and specially placed forward calorimeters cover the area up to $|\eta| < 4.9$.

Almost all particles traversing the calorimeter are stopped, and their energies measured; ranging from a few GeV up to multiple TeV. Muons with reasonable energy, but below 300 GeV, loose about 6 GeV due to ionization and then enter the muon spectrometer, discussed below. Practically all other particles that enter the calorimeter are stopped.

**Electromagnetic calorimeter**

The electromagnetic calorimeter (EM CAL) is positioned in front of the hadronic calorimeters, because electrons and photons rapidly lose their energy due to bremsstrahlung and pair production respectively, leading to an electromagnetic shower. The EM CALs have
a depth of more than 22 radiation lengths $X_0$ and therefore can contain most electromagnetic showers. Hadrons on the other hand, cause hadronic showers (via strong interactions), but they still show up in the EM CALs because through decay pairs of photons are created. The EM CALs measure the energy deposited by electrons and photons in liquid argon (LAr) layers that are interleaved with passive material layers made out of lead. These layers are shaped in a zig-zag pattern (an accordion) to increase the radiation lengths particles have to traverse, and to create a structure with full $\phi$-coverage, yet without azimuthal cracks. The resolution in the barrel is $\Delta \eta \times \Delta \phi = 0.003 \times 0.1$ for the first layer, $0.025 \times 0.025$ in the barrel for second layer, and $0.050 \times 0.025$ in third layer.

Due to support-structure considerations, there is a 4 mm gap at $z = 0$, and the barrel and end-caps overlap in the $1.375 < |\eta| < 1.475$ region (this is called the crack region). The EM CAL covers the region up to $|\eta| < 3.2$.

The design energy resolution for electrons is $\sigma(E) / E = 10\% \times \sqrt{E} + 0.7\%$ ($E$ in GeV).[124]

**Hadronic calorimeter**

The hadronic calorimeter (HCAL) consists of three different components. In the barrel region at 2.28 m to 4.45 m distance from the beam-line, the tile calorimeter (Tile CAL) extends up to $|\eta| < 1.0$. The Tile CAL is built up out of steel absorber material with tiles made out of scintillating material. Its large size is due to the fact that hadronic showers tend to be much wider and longer than EM showers. The other two components are the extended barrels ($0.8 < |\eta| < 1.7$), and end-caps ($1.5 < |\eta| < 3.2$), where the end-caps are also called the liquid argon (LAr) hadronic end-cap calorimeters (HEC).

The energy resolution for a single hadron is $\sigma(E) / E = 50\% \times \sqrt{E} + 3\%$ ($E$ in GeV). For jets, it is difficult to quote the measurement accuracy. The resolution depends on the constituents of the jets and its $\eta$ range. A rough indication for the average jet considered in this thesis would be a design resolution of $\sigma(E) / E = 100\% \times \sqrt{E}$.[124]

**Forward calorimeter**

The forward calorimeter (FCal) extends from $3.1 < |\eta| < 4.9$. It has three layers: the first is tuned for EM showering measurements, and the last two layers are for hadronic shower measurements.

The design energy resolution is $\sigma(E) / E = 100\% \times \sqrt{E} + 10\%$ ($E$ in GeV).[124]

**5.2.4 - Muon spectrometer**

The muon spectrometer (MS) is the outermost sub-detector of ATLAS.[128] As all other types of (non-neutrino) particles have already been stopped by the other components of the detector, all charged particles reaching the MS are muons.
In figure 5.6, the different subcomponents of the MS are illustrated. In the barrel region $|\eta| < 1.0$, the system consists of three layers of muon chambers. A muon chamber consists typically of two multilayers, each equipped with three layers of drift tubes. The largest chamber-type are positioned in the outer layer, which has a radius 10 m. These chambers are of the so-called Barrel-Outer-Large type, which are 2 m width and 5 m long and were constructed at Nikhef in Amsterdam. For trigger purposes, fast trigger chambers called resistive plate chambers (RPCs) are used which are mounted on both sides of the middle layer.

The design resolution for the MS's energy measurement is 3% over a wide energy range, gradually changing with the muon $p_T$ to the design aim of 10% at 1 TeV.\[124\]

### Monitored drift tubes

The muon chambers are equipped with alignment-, temperature- and B-field sensors to monitor the drift tubes (MDT). These are tubes filled with Ar+CO$_2$ gas, with a diameter of 3.0 cm, and they are between 1 m and 6.5 m long. A wire at 3080 V runs through the center
of each tube. When a muon ionizes the gas, electrons drift to the wire. The drift time provides an accurate measure of the drift radius. This results in a 80 micrometer positional resolution transverse to the tube. The coverage of the MDTs is up to $|\eta| < 2.7$.

5.2.5 - Trigger system

During the 2012 data-taking period, LHC produced about 20 million bunch crossings per second in the ATLAS detector, which implies about 20 proton-proton collisions. This rate is too high to fully process, as the maximal rate at which events can be read out from the detector and recorded is below 1 kHz. Additionally, many of the events contain only non-interesting physics. Hence, a trigger system is required to reduce the rate a factor of a million by rejecting the non-interesting events. In many of these non-interesting collisions there is no large exchange of transverse momentum, resulting in so-called soft collisions. Interesting events feature, for instance, large (missing) transverse energy, isolated electromagnetic energy deposits, or muonic activity. The trigger system decides in real time whether to save or reject an event. In case of a positive decision, the detected signals are further processed by the data acquisition (DAQ) system.\(^{129-130}\)

In figure 5.7, a schematic overview of the trigger system is shown. The data from the detector (about 1.5 MB per event) flows through the diagram from the top to the bottom, passing through (or being rejected by) various trigger levels. Each level reduces the event rate by rejecting events that don't pass that level's criteria. This multi-level design is necessary, because the processing time per event is very limited, especially for the first level trigger. The next few paragraphs explain these levels in more detail. The numbers that are given in this section are for the 2012 run.\(^{131}\)

**Level 1 trigger**

The Level 1 (LVL1, or L1) hardware trigger uses specially designed electronics in order to come to a decision of whether to keep or drop an event with a decision latency of a mere 2.5 microseconds. This short time is necessary due to the high rate of incoming events; the L1 trigger is able to reduce the event rate from 20 MHz down to 70 kHz. It mainly triggers on high-$p_T$ leptons, hadronically decaying tau's, photon and jets. Examples of such selections are cuts on the $p_T$ of these objects, isolation of EM clusters, etc. The exact combination and used thresholds are set by the selected trigger menu. L1 uses the muon trigger chambers clustered together for muon-identification, and reduced granularity calorimeter information for hadrons. It also performs rudimentary reconstructions of the total transverse energy in the event, and the missing energy. Regions of Interest (RoIs) that have found to contains particles, are selected as the center of the interesting area in $\eta$-$\phi$ space and passed to the L2 trigger.
Chapter 5

Figure 5.7: The ATLAS trigger system. At the top the information from the detectors enters the trigger system. On the right are several system processing the data, feeding the processed information to the triggers on the left side. Illustration adapted from [132].

**Level 2 trigger**
The Level 2 (LVL2, or L2) trigger is implemented mostly in software. It takes the events that pass the L1 trigger, and uses object reconstruction software to re-examine L1’s ROIs based on more detailed information. Contrary to the L1 trigger, the L2 trigger uses information from all sub-detectors, and thereby reduces the fake rate. This entire process takes about 40 ms per event, and reduces the event rate down to 5 kHz.

**Event filter trigger**
Finally, the Event Filter (EF) trigger is fully implemented in software, and uses all the detector information and full object reconstruction to reconstruct all the objects in the events. This takes about 4 seconds per event. The cuts employed here reduce the rate to approximately 700 Hz, of which about 200 Hz is written directly to tape for delayed processing, because not all these events can be processed immediately. The remaining 500 Hz of events are stored on the Grid for further processing, which is explained in more detail in section 5.4.

The L2 and EF triggers together are also called the High-Level Triggers (HLT), and they run on a large computing farm located in the ATLAS cavern.
**Trigger chains**

Many different settings can be used for the different levels of triggers described above. However, as only a limited amount of bandwidth and computing power is available, only a subset of triggers can be active at the same time. The collection of active triggers is called the trigger menu, and this can be varied from run to run.

Some triggers are the prescaled triggers. They are mainly used in physics analysis, but they have been prescaled down at high luminosity to keep bandwidth usage acceptable. As in this thesis only unprescaled triggers are used, we will not discuss these types of triggers.

The active triggers are chosen to optimally select events containing interesting physics objects. Often, multiple triggers using different reconstruction or selection algorithms are used sequentially to do this; this is called a trigger chain. Most trigger chains are between 2 to 10 algorithms long. The outcome of all the trigger decisions are stored so that they are available for further analyses further down the pipeline. In this section, we describe the triggers used for the top-quark analyses in this thesis.

The analysis of top-quark events in this thesis are based on the following triggers:

- For electrons, events needs to pass at least one of the following two triggers:
  - EF_e24vhi_medium1
  - EF_e60_medium1
- For muons, events needs to pass at least one of the following two triggers:
  - EF_mu24i_tight
  - EF_mu36_tight

The numbers in the trigger name signify the GeV-threshold placed on the object (the turn-on energy of the trigger). The "medium1" and "tight" refer to the lepton selection, which is explained in more detail in section 6.1 and section 6.2. "i" signifies that an isolation criteria is applied for the lepton, "v" that L1 uses thresholds varying as function of $\eta$, and "h" that the L1 trigger uses a hadronic core veto. For more details about the trigger naming, see [133].

In figure 5.8 the electron trigger efficiencies are illustrated as a function of $E_T$ and $\eta$ respectively. The turn-on of the trigger is clearly visible around 24 GeV in figure 5.8a, as is the efficiency plateau at higher energies. Figure 5.8b features the crack region around $|\eta| \sim 1.5$, and the reduced efficiency in the end-caps.

Figure 5.9 shows the muon trigger efficiencies as a function of the muon's $p_T$, in the barrel and end-cap regions. Just as with the electron triggers, both the turn-on and the efficiency plateaus are clearly visible.

**Event streams**

All recorded events are categorized into collections, called streams, based on the type of reconstructed objects. A single event can belong to multiple streams. These streams makes it
Figure 5.8: The electron trigger efficiencies at 8 TeV as a function of $E_T$ and $\eta$. Illustrations from [134].

Figure 5.9: The muon trigger efficiencies at 8 TeV as a function of $p_T$ in the barrel (a) and end-cap (b) region. Illustrations from [135].

It is easier to select certain types of events for analyses, as less events have to be processed. The most important streams in the context of this thesis are:

- **Physics**: The Physics stream contains the events that are used for most physics analyses. This stream is subdivided into three other streams (events can be shared between these streams):
  - **Egamma**: this stream contains events passing the electron or photon trigger chain;
  - **Muons**: this stream contains events passing the muon trigger chain;
  - **JetTauEtmiss**: this stream contains events passing the jet, tau, $E_T$, or $\sum E_T$ trigger chain.
Figure 5.10: The event stream recording rates from the 2012 data-taking period. Illustration from [136].

- **Minbias**: the minbias stream contains events passing the so-called minimum bias triggers. These triggers select a small (random) fraction of events based on minimal hadronic activity originating from one vertex. This provides a set of events that are free of non-collision background, whilst not selecting for any particular type of hard scattering;

- **Delayed**: these events have been recorded, but cannot be processed immediately. They are stored on tape for later retrieval and processing. The events in this stream are not used in this thesis.

In figure 5.10, the event rates for the different streams during the 2012 data-taking period are shown. At the start of longer data-taking periods, the minbias stream is dominating, as this stream is used to calibrate various subdetectors. The recording of the 'delayed' direct-to-tape events can be seen to start around May. The cumulative rate in the graph can exceed the maximum recording rate due to events being part of multiple streams.

In this thesis, events from both the Egamma and Muons streams will be used.
5.3 - Detector simulation

The previous sections have dealt exclusively with the recording and handling of data events taken with the ATLAS detector. As was discussed in chapter 3, simulated events are used as well. After having been generated, these Monte Carlo events are processed by a procedure called detector simulation in order to simulate the event signatures of the events in the ATLAS detector.

For this purpose, a virtual ATLAS detector was simulated using the GEANT4[137-138] software package. This model of the detector consists of more than 316,000 virtual volumes,[139] and uses a 'conditions database' to obtain information about the calibration, misalignment and possible dead-channels of the real ATLAS detector. GEANT4 then simulates the propagation of the particles through the detector, taking into account the effects of the magnetic field. It calculates where energy deposits are made, as well as the energy loss of the particles as they travel through the detector. Additionally, all the possible particle decays are also handled. This procedure is known as full simulation (FullSim), and is relatively computing intensive. Processing a single event can take up to tens of minutes.[142] However, since the ATLAS detector doesn't cover the entire 4\pi solid angle, it is not necessary to simulate the path of all particles. A truth filter of |\eta| < 6 is applied to ignore all particles that are too forward to hit the detector.

A full simulation is often not needed: many of the finer details are irrelevant for analyses, and need not be modeled very accurately. Therefore, a simplified process called fast simulation is often used.

Fast simulation has been mostly done with Atlfast I.[140] Instead of simulating the detector, Atlfast I smears the Monte Carlo truth information directly with resolutions obtained from full simulation. This is much faster than full simulation, with some loss in accuracy.

A newer approach is Atlfast II.[141] This is a combination of full simulation and fast simulation, and is now the default simulation for MC samples in ATLAS. With Atlfast II, a selection can be made which subdetector to do fast, and which to simulate fully. The Atlfast II simulation uses pre-stored templates of detector deposits per particle, and is therefore much closer to a full simulation than Atlfast I with only a minimal impact on performance. However, most samples in this thesis have been simulated with full simulation.

After detector simulation, the response of the detector is digitized. The energy deposits modeled by the detector simulation (with the exception of Atlfast I, which includes the digitization step) are converted into virtual electrical signals, which represents the signals in the detector. The resulting data format is identical to that of the actual ATLAS detector readout, and the output of the detector simulation is used as input by the same processing flow as the actual recorded data.
5.3.1 - Pile-up

Another experimental effect that has to be simulated is pile-up, where two or more distinct collision events are so close together time-wise, that they overlap in the detector. This results in real objects being reconstructed that do not belong to the collision of interest. These additional collisions can come from the same bunch-crossing (in-time pile-up) or from neighboring bunch crossings (out-of-time pile-up). The effects of pile-up have become more significant with the increasing peak luminosity. The number of collisions that occur per bunch-crossing (i.e. in-time pile-up) for 2012 data is shown in figure 5.11.

The MC simulation accounts for pile-up by overlaying simulated events with (parts of) recorded minimum bias (minbias) events. The exact amount of pile-up can differ from run to run, depending on the conditions of the detector and the beam, which are usually not exactly known at the time of the simulation. Therefore the MC events are eventually reweighted to match number of overlapping events in the ATLAS data.

The effects of pile-up are shown in figure 5.12, where a $Z\rightarrow\mu\mu$ candidate event is shown. It is clear that pile-up has a significant impact on the data recorded by the ATLAS detector.
Figure 5.12: An event recorded in 2012 that passes the $Z\rightarrow\mu\mu$ selection criteria. The two tracks associated with the two muons are drawn in thicker, yellow lines. In total, there are 25 reconstructed vertices in this single event, of which only one belongs to the $Z$-boson decay. All other vertices are caused by in-time pile-up. Illustration from [143].

5.4 - Data processing

All the events that pass one or more trigger chains are immediately stored on tape, in the RAW format. This is done primarily to always have a backup of the data. Additionally,
if improvements to the code are made or a better understanding of the run conditions is achieved, the storage of the RAW files allows for the reprocessing of the data. From this point onward, the processing continues offline, which means that it runs independently from the detector. Online processing runs while data taking takes place, for example the trigger chain algorithms. Everything mentioned in this section however, are examples of offline processing.

The recorded RAW files are not a convenient format for physics analyses, as this contains little more than the raw output of all the detector components. This information needs to be reconstructed into (for example) a list of detected particles. RAW files are typically also very large, taking up a lot of storage space, and taking a long time to process. The next step is thus to process this raw information, and to step-by-step reconstruct the particles that were produced in the collision. The output of this step is much smaller in size, as only the particle information is now stored, instead of the raw detector output.

The RAW data is stored in POOL files.\textsuperscript{145} POOL is a file format that can store C++ objects, and can store metadata about these objects as well. Objects are stored through the Gaudi Framework,\textsuperscript{146-148} which is an experiment independent set of services for data processing specialized for high energy physics (HEP) experiments. This framework has been adopted by ATLAS. In addition, a number of general utility libraries have been implemented, such as the CLHEP common libraries\textsuperscript{149} which contain common functions, like Lorentz boosts. Also used is the HepMC library,\textsuperscript{150} which provides a framework for storing and handling Monte Carlo particles. Combined together, this ATLAS framework is called Athena.\textsuperscript{151} Athena performs all the data processing and object reconstruction (chapter 6), and is able to process data as well as Monte Carlo files. Certain of the tasks associated with this can take up a lot of computing time. For example, detector simulation can take as much as tens of minutes per event. Demanding this amount of computing power from a single computer is not feasible, so a cluster of computers called the Grid is used.

5.4.1 - The Grid

The data produced by the ATLAS detector over the course of a data-taking period ranges into the petabytes (PB; 1000 terabytes). About 10 petabyte was recorded during all the Run I periods combined.\textsuperscript{152} At various points during the processing, the events are split into various streams, with each stream focusing on a particular signature of the collision event. Each additional stream increases the amount of storage needed. Further more, because the processing is generally done in several steps, intermediate results have to be stored as well. As for some studies these intermediate results are reused, they cannot be deleted. This means that many dozens of petabytes of storage are needed. No single computer, or even a single cluster of computers would practically be able to process and store all the data. Instead, a distributed computing solution is used, called the
World-wide LHC Computing Grid (WLCG), or Grid for short. The Grid is a heterogeneous collection of computer clusters in which many international institutes have connected their local computing clusters. Each of these clusters is called a site. (The Grid is also open to projects outside of CERN, but here only the ATLAS-specific component is discussed.)

The sites are roughly categorized into four groups: tier 0, 1, 2, and 3 sites:

- **Tier 0**: The Tier 0 site is located at CERN, and is the primary processing location of data. It is used for the immediate production of the RAW files. Additionally, the first steps of data processing take place here, as access to various databases located at CERN is needed. From here, processed files are copied throughout the Grid to different sites for further processing.

- **Tier 1**: Tier 1 sites are National Centers with large computing and storage facilities. Each tier 1 site also has tape backup facilities for collision data. ATLAS currently has 11 tier 1 sites of which Nikhef/SARA is one. Each tier 1 site further acts as the central point of a so-called "cloud" of other computing sites.

- **Tier 2**: Each cloud can have an arbitrary number of tier 2 sites. These are typically smaller sites managed by institutes that have local groups working on LHC analyses.

- **Tier 3**: Tier 3 sites are for the benefit and purpose of the individual institute, and are often not large enough to allow users from other institutes to extensively use the site, but they are assigned to a cloud in order to have transparent usage. They are officially not part of WLCG, and are mainly used for Monte Carlo sample production purposes.

- **Tier 4**: An informal term used to indicate workstations of end users.

The layout of the sites in tiers is illustrated in figure 5.13. As the layout of clusters get more integrated with each other, the distinction between tier 1, 2 and 3 sites gets vague and diffuse. It is likely that in the future the distinction between tier 2 and tier 3 sites will vanish. Note that the tier 0 site is not available for users, as this site is dedicated to the processing of the detector output. However, all other sites can, in principle, be used by any ATLAS user worldwide.

All Grid sites are used for processing and storage of the ATLAS data, and many of them are available for users to run their jobs on. To this end, all sites have both a data storage and a job processing role.

**Data storage**

Files on the Grid are grouped together in so-called datasets. Every file is part of a dataset, and every dataset has a unique name. To this end, there is a dataset naming scheme that enforces this.[154]

When a new dataset is created, its status is set to "open". This means the owner (the dataset's creator) can freely add new files to the dataset. Afterwards, the dataset must be frozen ("closed"). This prevents new files to be added to the dataset. Since a closed dataset cannot

---

24. The Dutch cloud is actually a special case, as it is located at both SARA and Nikhef, both in Amsterdam, the Netherlands.
be changed, the information of the dataset can be cached, resulting in a much better performance.

Datasets can be requested to be copied to other sites (replication). When such a Data Transfer Request (DaTRi) is submitted, the permissions of the user are verified to make sure the destination is write-able for that user. If that is the case, the Grid will automatically start copying files over, notifying the user when the transfer is completed.

There are restrictions to datasets. Most importantly, because there is a transfer overhead per file (and not all implementations scale), the number of files per dataset is limited. Most dataset tools employ a hard limit, usually set to 10,000 files. Due to this overhead, the performance with a few large files is better than many small files. This can be achieved through merging files. For example, the official production jobs (the jobs that process the ATLAS data and Monte Carlo, section 5.4.2) have several such steps in between the different conversions, in order to keep performance high.

To overcome these limitations, dataset containers can be used. A dataset container is simply a collection of datasets, and the content of the container can be changed any time (there is no freezing procedure for dataset containers). This also allows data samples stored in dataset containers to be extended by adding more datasets to them.

25. This is not entirely true, but this kind of advanced usage will not be considered in this thesis.
Chapter 5

All the files, datasets, and dataset containers are available through the Grid to run over, and can be downloaded by all users. This is done by using a toolset called Don Quixote 2 (DQ2).\textsuperscript{155} The DQ2 tools allow for the creation and deletion of datasets and dataset containers, as well as the downloading and replication.\textsuperscript{26} This, and all things Grid file related, all fall under the umbrella of ATLAS Distributed Data Management (DDM).

Internally, DQ2 uses the LCG utility tools for file manipulation. These tools communicate with the Storage Resource Manager (SRM)\textsuperscript{157} server that is hosted by each Grid site.\textsuperscript{27} There are different sets of SRM software: dCache,\textsuperscript{158} DPM (Disk Pool Manager),\textsuperscript{159} StoRM,\textsuperscript{160} and others. Each SRM server implements the same interface, but handles the underlying disks differently.

The LCG tools include the LCG File Catalog (LFC), a database that connects the DQ2 name of a file with the logical file name (LFN) that the SRM servers recognize. These relations are illustrated in figure 5.14. Each file stored in this system can have multiple replicas on different sites at the same time, and even multiple copies on different space-tokens (which are discussed below) at a single site.

File transfers (both between two sites, and between a user and a site) are done using a high-performance version of FTP\textsuperscript{163} called GridFTP\textsuperscript{164-165}. Bulk transfers are handled through the file transfer service (FTS) channels, which have been set-up between all tier 1 sites, and between tier 1 and tier 2, 3 sites to bundle GridFTP transfers, and optimize bandwidth usage. The physical transfer of data is done using dedicated glass fibers (although, if needed, these transfers can be re-routed over the internet). This fiber network is called LHCOPN (LHC Optical Private Network),\textsuperscript{162} and it is shown in figure 5.15. It connects many of the Grid sites worldwide together with dedicated fiber cables.

Storage on a site is divided into several groups, called storage elements (SE). An SE is a separate server storing files, and each SE can contain several space-tokens. Each space-token has its own file hierarchy; they can be seen as a root-directory. Each of these space-tokens has a specific purpose, and separate access rights. For example, all output files of jobs are put on the SCRATCHDISK space-token, from which the files are automatically deleted after about 30 days (files have to be replicated to other space-tokens, or downloaded before then). Another space-token, the LOCALGROUPDISK, is only writable for users from the institute that hosts the storage, and can be used to replicate datasets to the local Grid site for processing. Having the job run locally makes it easier to use the output in a separate local batch system, because it removes the need to replicate or retrieve the output from the Grid site when the jobs have finished.

\textsuperscript{26} A new set of tools, called Rucio\textsuperscript{156} has been deployed over the course of 2014, superseding DQ2 while remaining mostly backwards compatible with it. Rucio is more efficient with file storage, as it decouples the file location on disk from the dataset name. It also simplifies dataset handling, as it does not support dataset versioning. We will discuss file distribution only in terms of DQ2 in this thesis.

\textsuperscript{27} SRM technology is not unique to the Grid: it originates elsewhere and is used by many other projects.
Files are stored using a Globally Unique Identifier (GUID), which is unique throughout the Grid. This GUID is stored by the DQ2 system, and a one-to-one mapping between the LFN and GUID must exist. The LFN can also be an alias to another LFN; this is how replica's of files are handled. Together with the GUID of a file, DQ2 also stores the Adler32 checksum\cite{166} of each file to be able to verify the file integrity after a transfer. As files are immutable, this checksum can never change.\footnote{If an update to a file needs to be made, a new dataset or a new version of the dataset needs to be made.}

Another way to address files is through their Storage URL (SURL). This is a URL format that the SRM servers use, and it is build up of the LFN and the URL-address of the SRM server where the file is located. For direct file access, a Transfer URL (TURL) is created, with specifies the access protocol to use to access the file. This can be, for example, GridFTP, RFIO,\cite{167} or XROOTD,\cite{168} with GridFTP and XROOTD providing direct access to Grid.
storage elements from user workstations. This multitude of file access systems allows users to select the most optimal way to access any dataset for their specific circumstances.

**Job processing**

Submitting code to a site to process ("run over") files on the Grid is done by packaging the code that needs to be run, and any other needed software libraries that are not provided at the site, into a single payload (usually a tarball). The payload is then submitted to a scheduling system, together with the details of which files to run over. Such a request is called a job.

In order to reduce the need to send common libraries with jobs, several software repositories are provided by CERN and ATLAS. For example, through the Andrew File System (AFS)\(^{[169]}\) many pre-compiled software packages can be obtained.\(^{29}\) A more recent system called the CERN Virtual Machine File System (CVMFS)\(^{[170]}\) is HTTP-based\(^{[171]}\) and this is available on all Grid sites. The biggest advantage of this is that HTTP uses web- and proxy-servers, which are good in caching, providing quick access to often used software packages.

---

29. Not all sites provide AFS access.
Figure 5.16: Overview of the various ways of job submission and output retrieval on the Grid that are available to users. Illustration from [172].

The most-used scheduling and accounting system for jobs in ATLAS is PandaTools (or Panda).[173] A central Panda server performs the job scheduling and management, while user tools such as Pathena (a Panda-wrapper around Athena) allow for easy job submission.

Another popular software package to manage jobs is Ganga.[174] Ganga can be scripted through Python, giving more control over the jobs and their management than Panda. It also supports submitting to other job frameworks than Panda, as illustrated in figure 5.16. As the Panda job scheduler is the de facto default job management framework for ATLAS, we will describe that workflow below.

Each site can have multiple queues. For example, the user analysis jobs and production jobs are separated by being submitted to different queues. Some sites also have dedicated queues for multi-core jobs, or jobs that require a large amount of memory. Users can specify which
queue to submit to, but by default \texttt{Panda} handles the queue assignment (job brokerage). Multiple factors play a role in the determination which queue to allocate the job to, based on which sites have the requested files available, how busy the available queues are, if the queue has been blacklisted, etc.

Sometimes, there is a problem at a site which causes jobs submitted there to fail. Examples of problems are broken or misconfigured hardware or operating systems, network issues, disk failures, and power outages. To prevent jobs for getting assigned to such a problematic queue, two automatic blacklisting systems called GangaRobot and HammerCloud are used. Pre-defined test-jobs are frequently sent to each queue, and if such jobs fail with a large enough rate, the site is taken offline while the system administrators investigate the cause of the failures. This vastly reduces the amount of jobs that the user needs to be resubmit.

The computing element (CE) of a Grid site runs the job scheduler, and publishes the site's information (the configuration of the queues, current status, etc.) to a central database. The CE receives all (pilot) job submissions from the different collaborations, and handles the job scheduling, for which several software packages are available, such as \texttt{TORQUE}\textsuperscript{175} and HTCondor (or Condor).\textsuperscript{176} The job scheduling is configured such that each collaboration gets the necessary amount of computing resources, but also ensuring that a single collaboration doesn't monopolize the site.

Before the code contained in a job starts to execute, first a small framework that handles the input/output is started. This code is called the pilot framework.\textsuperscript{177} The job scheduler at the Grid site assigned the job to a machine in its cluster, the worker node (WN), and starts the pilot on that machine. The pilot is responsible for contacting the job server to retrieve the code tarball, and to make the input files available. It then sets up the software environment for the job to run in, and starts the job. While the job is running, it routinely sends a status update back to the \texttt{Panda} server, which allows users to keep track of the job's progress (job status). When the job is finished, the pilot reports the final status (success or failure), and in the case of success transfers the local output files to the output Grid dataset.

Often code needs to be compiled before being able to execute on the WN. This can be done by the user before submitting the code, although this might introduce binary incompatibilities if the WN has a different hardware architecture, or different versions of libraries installed. Therefore, usually the code is submitted first to the site in a single job, and compiled. This job is called the build job. When that job has finished, it creates a dataset containing the executable files to run, and the run jobs are started.\textsuperscript{30} Alternatively, each job can compile their own executable, but this is inefficient.

In the above description of the Grid, many advanced features have been left out. For example, a recently implemented protocol called Federated ATLAS \texttt{Xrootd (FAX)}\textsuperscript{178} allows users to access files from arbitrary Grid sites. This makes it possible to run on one Grid site,

\footnote{30. These jobs are submitted at the same time, but the run jobs do not start as the dataset containing the executable is still open. This is another reason for the open/close feature of datasets.}
while the files are stored on another Grid site. Such advanced features are out of the scope of the overview presented in this thesis.

Cloud computing outside particle physics
There are many parallels between the need for computing power in ATLAS and in the IT-industry, where distributed computing is known as cloud computing. The Grid provides a unique testbed for experimental techniques that can be applied in cloud computing. The push for more computational power results in multi-faceted improvements, ranging in the form of optimization of existing hard- and software, to rethinking and restructuring the fundamental infrastructure. All are needed due to the increasing workload exceeding mere hardware-based growth. Hopefully the obtained solutions will find their way into commercial applications, because many of the lessons learned with the Grid are directly applicable to the cloud computing industry. More specifically, the Grid is a good case study of providing distributed computing capabilities in a relatively decentralized, heterogeneous environment. Studies done towards the scalability and maintainability of this environment certainly provide valuable information and insights of do's and don'ts for cloud computing.

5.4.2 - Production system
The official ATLAS production system handles both the ATLAS data that was recorded, and the Monte Carlo sample generation. Both of these take place almost exclusively on the Grid. The production system automatically resubmits failed jobs, and can re-broker jobs if a site goes down, passing the jobs on to the next site like the proverbial hot potato. This makes the production system robust, even on less reliable sites, allowing a good usage of all available resources.

We will not discuss in detail the Monte Carlo sample generation here, as this is out of the scope of this thesis, although some information is given in section 3.3. Most generation jobs run within Athena, as this simplifies the handling of the output. In some cases, dedicated types of jobs are used, which are then fed into Athena for further processing.

After detector simulation, the production system also takes care of producing the data files that can be used in user analyses. This follows the pipeline shown in figure 5.17. The conversion from RAW to POOL files was already mentioned previously. This conversion is done with Athena. The output POOL files are then collected into an event summary data (ESD) dataset. During this conversion, several reconstruction steps are also performed. The goal of ESD files is to remove the raw signal output of the detector, but to keep the information of which channel of a subdetector was hit. This allows for efficiency studies of the various detector components.

Next, the ESD files are further processed into analysis object data (AOD) files. The AOD files contain no specific detector information anymore; only the outcome of the triggers is stored.
All detector signals have been reconstructed into physics objects (electrons, muons, etc.). As the ESD files are large, and not used by many studies, they are stored only on tier 1 sites. AOD files stored on both tier 1 and tier 2 sites for usage in physics analyses. However, AOD files are still relatively large, so they are further processed into Derived Physics Data (DPD) files. At this point, assumptions about the specific physics analysis are made. In other words, one set of AOD files is processed into different sets of DPD files, where each DPD set optimizes for a specific analysis by demanding a minimum number of a certain object, or the firing of a specific trigger (skimming). Additionally, unneeded variables are dropped from the file (slimming). This processes of slimming and skimming can be repeated multiple times, resulting in D2PD's, D3PD's, etc.

5.4.3 - User analysis code

Even though physics analyses can be done with Athena, it is a rather heavy software package. Most users choose to perform their analysis in a more lightweight framework called ROOT.\textsuperscript{[179-181]} ROOT has its own file format (ROOT files), and provides facilities for data analysis, statistical analysis, plotting, and more. However, the main advantage of using Athena instead is that many commonly used physics operations are already available in Athena, and the software is generally well-tested. Which package to use is left to the user's preference. A tool that is used to open POOL in ROOT is AthenaROOTAccess (ARA).\textsuperscript{[182]} This further reduces the need for Athena, but a well-establish convention for the file format is needed for this to work.

However, some of the needed physics analysis tools are not available by default in ROOT, so a lot of code has to be written in order to perform the analysis. For this reason, an ATLAS-wide framework called RootCore\textsuperscript{[183]} has been developed, which also contain scripts to support job submission to run of the code on the Grid. In the top-group, a specialized version
called TopRootCore \(^{[184]}\) is used to, for example, produce D3PD's and to perform most user analysis.

**Software used in this thesis**

For the analysis in chapter 4 Athena 16.6.4.3 on Scientific Linux 5 (SL5) \(^{[185-186]}\) was used to parton shower the samples, and save the needed information to ROOT files. ROOT 5.34/14 on SL5 was subsequently used for the processing of the events, with a custom 'hadd'\(^{31}\) that enabled the creation of large (many GBs) ROOT files.

For jet reconstruction, a software package called FastJet version 3.0.6 was used; this is described in more details in section 6.3.1.

For PDF set generation a software package called LHAPDF version 6.1.4 was used (see section 7.5).

The samples used in chapter 7 and 8 have been processed with Athena on the Grid, and then skimmed and slimmed on the Grid with TopRootCore into D3PDs. Afterwards, a self-written ROOT-based program called DannyLoop was used on SL6 to perform the analysis. More information about this program can be found in appendix C. ROOT 5.34/19 was used for appendix C, with a custom fix to resolve an SVG\(^{[187]}\) image output issue (ROOT-6994).\(^{[188]}\)

---

31. A ROOT-based program used to merge ROOT files.
Chapter 5
6 - Event reconstruction and pre-selection

In this chapter, we describe the reconstruction of physics objects from the events of the (simulated) ATLAS detector. The physics objects, such as the leptons, jets, and missing energy, are required to be reconstructed from the low-level detector information of ATLAS with high efficiency, high purity, and high resolution in order to minimize the contribution of falsely identified objects.

6.1 - Electron reconstruction

During the first part of the reconstruction information of the calorimeters is investigated. At this stage, electrons and photons are indistinguishable and considered equivalent. The photons are separated from the electrons by track requirements that are made at a later stage.\[190\]

Electron candidates are obtained from clusters of cells in the EM calorimeter.\[189-190\] The clusters are formed by using a sliding window algorithm.\[191-193\] The calorimeter is divided into areas with a size of $\Delta \eta \times \Delta \phi = 0.025 \times 0.025$ (in the middle layer cells). A window of $3 \times 5$ cells slides over the grid, and sums the transverse energy in the window to form a cluster seed. Cluster seeds with an energy below 2.5 GeV are rejected. Due to the detector acceptance, all electron candidates are required to lie within $|\eta| < 2.47$. ($|\eta| < 2.37$ for photons.) In both cases, the crack region $1.37 < |\eta| < 1.52$ is excluded due to limited instrumentation (see section 5.2.3).

Next, the electron candidates formed from the calorimeter clusters are associated with a track in the ID, and the clusters are re-investigated to optimize the cluster size: $3 \times 7$ ($5 \times 5$) in the barrel (end-cap). The quality requirements on the track are determined by a
chosen particle identification (PID) method. There are several PID methods that are used in top-quark analyses: loose, medium, and tight. Going from loose to tight, each level has stricter requirements on quantities like shower shape (for example, one of the used shower shape variable is $R_{np}$, which is the ratio between cell energies in $3 \times 7$ versus $7 \times 7$ cells), requirements on the ID tracks (for example, the transverse impact parameter $d_0$ is required to be small), and number of hits in various subdetectors in order to reduce the false selection of non-electron energy deposits, the so-called fake electrons.

Three companion PID methods are also defined: loose++, medium++, and tight++. These methods use more quantities (such as a cut on $E/p^{[194]}$) in order to have a higher performance in a high pile-up environment. The tight++ PID method is the one that is used for the electron reconstruction in the single top analysis later this thesis.

The tight++ PID method uses isolation requirements in order to select electrons. The isolation requirements on the electron candidates are specifically chosen to reduce the selection of electrons, either fake or genuine, from jets. Only minimum calorimeter activity (calorimeter isolation) and a limited number of tracks (track isolation) in an $\eta-\phi$ cone around the electron candidate's direction are allowed. The isolation requirements are optimized to have a uniform isolation efficiency across $\eta$; and the transverse energy $E_T$.

To enforce the electron track isolation, all the calorimeter energy associated with tracks within a cone of $\Delta R < 0.2$ are combined. If the electron track's $E_T$ contribution to this sum of calorimeter energy is more than 90%, the electron candidate is considered isolated. In addition, in the cone of $\Delta R < 0.3$, the electron must contribute 90% to the $p_T$ of all summed calorimeter energies.

The efficiency of the electron reconstruction and identification is measured by using a tag-and-probe method on $Z\rightarrow ee$ events. Figure 6.1 shows the electron identification efficiencies of the PID methods with respect to the triggered electron candidates. Going from the loose to tight selection, the efficiency clearly decreases. This is the price to pay to reduce the percentage of fake electrons (in other words, the purity of the selected events increases, as shown in figure 6.2). The modeling of these efficiencies in MC is reasonable, which is an important result as they are used during the detector simulation. To further calibrate the energy measurements from the EM calorimeter, energy scales are applied on data. These energy scales are obtained by studies of events from resonances such as $Z\rightarrow ee$, $J/\psi\rightarrow ee$, and $E/p$ studies using isolated electrons from $W\rightarrow ev$. (They have not yet been applied in the plots shown here.) Additionally, energy smearing and other correction factors derived from the tag-and-probe are applied to the Monte Carlo during the detector simulation phase in order to improve the agreement between data and MC.
Figure 6.1: The efficiencies of the electron PID methods as a function of the electron $E_T$ and $\eta$, measured using the tag-and-probe method on 8 TeV. The results are shown for various types of selection criteria as indicated (these are the non ++ variants). The multilepton represents a low energy electron selection, and is not used in this thesis. The ++ variants have similar efficiencies but smaller fake rate. Other criteria are explained in the text. Illustrations from [195].

6.2 - Muon reconstruction

Muon candidates are found by reconstructing tracks from hits in both the muon spectrometer and inner detector. For the MS, a track fitter is applied to the hits in the MDTs, using the space coordinate of the hit.\cite{196} The hits in the CSCs are reconstructed by applying the appropriate clusterization methods to the detector data. For the inner detector hits, the procedure is similar to what was described earlier for electrons.

There are different types of reconstructed muons, where each type uses information from the MS and ID in a different way. These four types are:\cite{197-198}

- **Combined (CB) muons** are reconstructed from full tracks in the muon spectrometer;
- **Segment tagged (ST) muons** are used to recover efficiency in poorly covered regions and at low transverse momenta;
- Muon spectrometer **stand-alone (SA) muons** are used to extend the muon acceptance from the ID to $|\eta| < 2.7$;
- **Calorimeter tagged (CaloTag) muons** are used to recover efficiency at $|\eta| \sim 0$. 
Segments are short parts of tracks (tracklets) in a single layer of the muon spectrometer. The CaloTag muons depend fully on the calorimeter, allowing muon reconstruction in the $|\eta| \sim 0$ area where there is a gap in the MS for support structures and services.

The different types of muons are often combined to provide better reconstruction efficiencies. Two reconstruction chains exist to do this:

- **Staco**, statistically combines the measurements of a muon in both MS and ID;
- **Muid**, performs a track fit using all muon hits in the MS and ID.

Recently, a third chain has been developed for Run II. This new chain uses the best performant algorithms from both Staco and Muid, but it is not used in this thesis; we use the Muid muons.

Similar to the electrons, the Muid muons can be divided into four quality levels: tight, medium, loose, and very loose. These levels are roughly defined as follows: (more details on these quality levels can be found in [198].)

- Very loose: muons constructed from track information from the TRT only;
- Loose: very loose muons that pass muon tagging algorithms, and have a track in the ID;
- Medium: all loose muons that are also standalone (which are reconstructed from MS tracks only);
- Tight: all medium muons that have a fit successfully combining the MS and ID tracks.

For the analyses in this thesis, the tight selection for muons is used.

To reduce the contribution from muons arising in the decay of heavy particles, isolation
Event reconstruction and pre-selection

Figure 6.3: The muon reconstruction efficiency for the 2012 data-taking run. (a) shows the muon reconstruction efficiency as a function of the muon $p_T$, for the chain 2 tight muons. (b) shows the muon reconstruction efficiency as a function of $\eta$. The drop at $|\eta| \sim 0$ is due to the central gap in the muon spectrometer. Illustrations from [198].

Requirements are imposed. A cone with a radius that is a function of the muon $p_T$ is placed around the muon track, and the scalar sum of the transverse momenta (with a $p_T$ above 1 GeV) of all the tracks in the cone (except the muon track) is calculated. The radius of the cone is given by:\footnote{199}

$$R_{iso} = \frac{10 \text{ GeV}}{p_T}. \quad (6.1)$$

Only muon candidates for which the scalar sum divided by the muon $p_T$ is less than 0.05 are accepted. Later, additional overlap removal (section 6.5.2) will further suppress the muons that are created during during particle decaying and interacting in the detector.

In figure 6.3, the efficiency of the muon reconstruction is shown. This efficiency is measured using a data sample enriched with $Z \rightarrow \mu\mu$ events. The high efficiency (compared to electrons, figure 6.1) is mostly due to the dedicated muon subdetector, the muon spectrometer.\footnote{198} As almost no charged particles except muons reach it, the MS has a very significant positive impact on the overall muon identification and reconstruction.

The purity of the muon reconstruction, as measured with the $Z \rightarrow \mu\mu$ events, is largest for muons with a $p_T \sim 40$ GeV. It decreases to 98.5% (97%) for $p_T = 10$ GeV (100 GeV).\footnote{198}
Similar to the electrons, energy and resolution smearing factors are derived from the Z-boson data, and they are applied during the detector simulation phase for Monte Carlo events, for a better match with data.

6.3 - Jet reconstruction

The number of outgoing hard partons in the HS of an event is an important quantity and used in many analyses to discriminate signal from background. The information needed to estimate this quantity is obtained from the energy depositions in the detector, using a technique called jet reconstruction. The energy depositions that are used for this are calibrated using the local cluster weighting (LCW) method, an energy- and $\eta$-dependent simulation-based calibration scheme, with in-situ correction based on data.\textsuperscript{201} They are then used as inputs to the jet finding algorithms.\textsuperscript{200}

In a very small fraction of events with pathological noise bursts in the calorimeter, jets can be incorrectly reconstructed from a few noisy cells. The nature of these signals have been studied in depth,\textsuperscript{202} and event cleaning cuts are applied to remove events with jets flagged as "bad" during the reconstruction, if the $p_T$ of the bad jet is > 20 GeV.

For some studies, the jet reconstruction algorithm needs to also applied directly to MC generated events, without detector simulation. In these cases the jet reconstruction uses the 'truth' information of the final state MC particles (instead of energy depositions). In this section, for brevity, we further only refer to particles.

6.3.1 - Reconstruction algorithms

Most jet reconstruction algorithms follow a prescription similar to the following iterative cone procedure:

1. Select the particle with the highest $p_T$ or $E_T$ in the event. This particle is called the seed for the proto-jet.
2. Find all particles that lie within a certain distance of that particle. The way this is calculated differs between algorithms, but usually it involves a cone/algorithm parameter. This is often parametrized by $R$, representing the radius of the cone, which is in the simplest case defined in the $\eta$-$\phi$ plane.
3. Remove all the found particles (including the originally selected particle) from the event. Their four-vector sum is the proto-jet.
4. Repeat until no particles are left.

From the proto-jets, the reconstructed jets are obtained by requiring a transverse momentum of at least 15 GeV, typically.
The simplest algorithm that can be applied, is the cone algorithm (as used by MLM jet matching, in section 3.3.3). This algorithm starts with the highest $p_T$ particle, and groups all other particles within a certain distance, the cone size, to form a proto-jet. The size of the cone is given by the cone parameter $R$. Any particle belongs to a proto-jet if it satisfies the following condition:

$$\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 < R^2,$$  \hspace{1cm} (6.2)

where $i$ denotes any of the final state particles, and $j$ is the center of the proto-jet. Note that this cone algorithm uses rapidity, not pseudo-rapidity; all other jet algorithms mentioned use pseudo-rapidity $\eta$ instead.

Instead of an iterative procedure that constantly recalculates the proto-jet, in this thesis we use a sequential algorithm for jet reconstruction that work by adding particles one by one to proto-jets. Of these algorithms, the Kt-family algorithms are most used.

The Kt-family algorithms are a generalized version of the Durham algorithm,\cite{203} and are IRC-safe by construction. The algorithms are named after $k_T$, the transverse momentum of a radiated particle with respect to its mother particle. The Kt-family algorithms work by finding the particle pair ($i,j$) that are closest together according to a distance measure $d_{ij}$, and merges these particles. For any pair of particles $i,j$, the distance measure $d_{ij}$ is defined as:

$$d_{ij} = \frac{\min\left(k_{t,i}^{2p}, k_{t,j}^{2p}\right)}{R^2} \Delta R_{ij}^2,$$  \hspace{1cm} (6.3)

where $R$ is the algorithm's size parameter, and the parameter $p$ determines the order in which the algorithm merges particles.

The beam is also treated as a particle during the merging, denoted by $B$, with $d_{i,B} = k_{t,i}^2$, and $d_{j,B} = k_{t,j}^2$. When a particle merges with it, it promotes the particle to a proto-jet (and it is removed from the list of particles).

For the parameter $p$, three different values are most commonly used:

<table>
<thead>
<tr>
<th>Value of $p$</th>
<th>Name of the algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p = 1$</td>
<td>Kt\cite{204}</td>
</tr>
<tr>
<td>$p = 0$</td>
<td>Cambridge-Aachen</td>
</tr>
<tr>
<td>$p = -1$</td>
<td>AntiKt\cite{205}</td>
</tr>
</tbody>
</table>

Table 6.1: The most-used values of the $p$ parameter for the Kt-family of jet reconstruction algorithms.
Chapter 6

The Kt-algorithm reconstructs proto-jets by trying to reverse the process of QCD branching process. It merges soft emission together first, producing larger and larger particles, until they have all merged with the beam and formed proto-jets. This works well, but one of the problems with this approach is that the shape of the proto-jet can be very irregular, as can be seen in figure 6.4a. This gives each proto-jet a different area in the $\eta-\phi$ plane, making background noise subtraction difficult.

The AntiKt-algorithm clusters emissions around the hardest particles first, and this results in much more circular shapes, as can be seen in figure 6.4d. This algorithm is therefore the algorithm of choice in this thesis.\(^{32}\)

6.4 - Identification of b-jets

Single top t-channel events contain one b-jet (a jet coming from a b-quark), while most background events do not have such a jet. This provides a powerful way of distinguishing signal from background events for top-quark analyses. Additionally, b-quarks coming from the ME (and thus their associated b-jets) are expected to have high $p_T$. For most aspects, b-jets leave very similar signatures in the detector compared to light jets, so identifying them is not trivial. For this purpose, techniques called b-tagging have been developed.

B-hadrons travel a significant distance in the detector before decaying. Consequently, the decay products (and thus the jet) will not originate from the primary vertex (section 6.5.3; the hard scattering position), but a secondary vertex (SV) is created. By reconstructing the secondary vertices in an event, the positions of decaying B-hadrons can be found, and b-jets can be identified. The tagging probability depends on the decay length parameter of the secondary vertex, $L_{xy}$. By cutting on its significance $\frac{L_{xy}}{\sigma(L_{xy})}$ secondary vertices can be distinguished from the primary vertex efficiently.

However, secondary interactions and finite tracking resolution will produce false secondary vertices, thus tagging on the secondary vertex alone is usually not enough. A second quantity that is often used, is the transverse impact parameter (IP) of the tracks in the jet, $d_0$. Again, the actual cut is done on the combined significance of all the tracks. Sometimes, in addition the longitudinal impact parameter $z_0$ is used.

Both $L_{xy}$ and $d_0$ are illustrated in figure 6.5. $L_{xy}$ is distance between the primary vertex and the secondary vertex in the $x$-$y$ plane, while $d_0$ gives the closest approach of the track to the primary vertex in the $r$-$\phi$ projection. $z_0$ is the $z$-coordinate of this point.

There is also a 3D variant of $L_{xy}$, called $L_{3D}$. As the name suggests, this variable measures the decay length in three dimensions, instead of only in the $x$-$y$ plane.

---

\(^{32}\) These algorithms have been implemented in a software package called FastJet,\(^{206-207}\) which has been used to perform the jet reconstruction for this thesis.
6.4.1 - Algorithms for $b$-tagging

There are many different $b$-tagging algorithms that can be used.\cite{210} In this thesis, only the MV1c algorithm is used. MV1c uses a neural network that combines the outputs of the several other $b$-tagging algorithms, just as MV1, but is specifically trained to reject charm jets as well.
There are two aspects to take into account when calibrating a $b$-tagging algorithm. Firstly there is the efficiency $\epsilon_b$, which is defined as:

$$
\epsilon_b \equiv \frac{N_{b\text{-jet}}^{\text{tagged}}}{N_{b\text{-jet}}^{\text{total}}},
$$

(6.4)

where $N_{b\text{-jet}}^{\text{tagged}}$ is the number of $b$-jets that are tagged, and $N_{b\text{-jet}}^{\text{total}}$ is the total number of $b$-jets.

The second aspect is the mistag rate. The more lenient the algorithm is calibrated, the larger the probability of falsely tagging a jet as a $b$-jet. This is often expressed as a light-jet rejection rate $R_l$ (which is the inverse of the mistag rate):

$$
R_l \equiv \left( \frac{N_{\text{non-b/c-jet}}^{\text{tagged}}}{N_{\text{non-b/c-jet}}^{\text{total}}} \right)^{-1},
$$

(6.5)
where \( N_{\text{non-b/c-jet}}^{\text{tagged}} \) is the number of non-\( b/c \)-jets that were tagged, and \( N_{\text{non-b/c-jet}}^{\text{total}} \) is the number of non-\( b/c \)-jets. The \( c \)-jets are removed from this definition because long-lived \( C \)-hadrons cannot be rejected. Because of the overwhelming amount of light jets, the rejection needs to be much higher than the efficiency.

When using a \( b \)-tagging algorithm, a working point is chosen, which defines the balance between the efficiency and mistag rate. The working point is expressed as the (approximate) efficiency of the \( b \)-tagging algorithm. In figure 6.6 the efficiency of the MV1 algorithm at the 70% working point is shown as a function of jet \( p_T \).\(^{33}\)

The values of the working points used for the \( b \)-tagging in this thesis have been calibrated on a sample of \( t \bar{t} \) events with \( p_T^{\text{jet}} \geq 20 \text{ GeV} \) and \( |\eta_{\text{jet}}| < 2.5 \),\(^{212}\) which have been updated for the 2012 data-taking period.\(^{211,213-215}\) For our analysis we have chosen the MV1c 50% working point. At this 50% working point, the purity is 98%, with the rejection factor (the inverse of the mistag rate) for \( c \)-jets, tau-jets, and light jets being 25.6, 119, and 1472 respectively.\(^{216}\)

\(^{33}\) Plots for MV1c were not available at the time of writing.
Chapter 6

6.5 - Top event pre-selection

In this section we present the selections applied to the reconstructed objects. Most of these selections are common for all 8 TeV top-quark analysis.[217] Several cuts are modified compared to the 7 TeV analysis: there are harder cuts on $p_T$ of the jets, because of the need to suppress pileup, and because the higher statistics allow to apply more stringent cuts, which leads to a better background suppression.

Events that are poorly reconstructed, or affected by suboptimal detector conditions need to be rejected. During a data-taking period, the conditions of the subdetectors can change for various reasons. This means that each run can be recorded under different circumstances, leading to different reconstruction efficiencies. In some cases, the information recorded is not sufficient or not precise enough to perform the reconstruction. Only runs with good beam and detector conditions are selected for physics analysis and added to the so-called good run list (GRL). All events used in top-quark analysis are required to be recorded in a run on the GRL.

6.5.1 - Object selection

Jets

As a first step, jets are reconstructed with the AntiKt-algorithm (as explained in section 6.3) with a radius parameter of 0.4. Not all of the reconstructed jets fall within the acceptance of the detector, or within regions where jets can be properly reconstructed. Only jets passing the following cut are accepted:

$$|\eta_{\text{jet}}| < 4.5.$$  \hspace{1cm} (6.6)

Figure 4.1a illustrated the $\eta$ distribution of jets in various Alpgen Monte Carlo samples, showing that only a small number of jets is cut away by this restriction.

Additionally, jets are required to have a $p_T$ sufficiently high to be on the reconstruction efficiency plateau:

$$p_{T_{\text{jet}}} \geq 30 \text{ GeV}.$$ \hspace{1cm} (6.7)

However, additional cuts to improve the jet reconstruction efficiencies in the forward region have to be applied. For jets falling in the transition region between forward and end-cap calorimeters ($2.5 < |\eta_{\text{jet}}| < 3.75$), a harsher $p_T$ requirement is used: $p_{T_{\text{jet}}} \geq 35$ GeV.

A cut is placed on the jets in order to reduce the effects of pileup. The jet vertex fraction (JVF) is a measure of how much a track is associated to a vertex.[218-221] If a jet originates
from pileup, it will not be closely associated to any vertex in the event, and it can be rejected with a cut on the JVF.

The JVF is defined as:

$$\text{JVF}(\text{jet, vtx}) = \frac{\sum (p_T(\text{trk}) \text{ if vtx } \in \text{ track.vtxlist})}{\sum p_T(\text{trk})},$$

(6.8)

where 'track.vtxlist' is the list of tracks associated to a vertex.

A cut of $|\text{JVF}| > 0.5$ is used to cut away pileup jets. The distribution of the JVF is shown in figure 6.7; values closer to 1 indicate a higher contribution from the hard scattering instead of pileup. The cut on JVF is only applied to jets $p_T < 50$ GeV and $|\eta| < 2.4$; the JVF is not well constructed for jets outside this region.

**Electrons**

For electrons we require that:

$$|\eta_{\text{el}}| < 2.47$$

$$E_{T_{\text{el}}} > 25 \text{ GeV}.$$  

(6.9)

Additionally, electrons in the crack region $1.37 < |\eta_{\text{el}}| < 1.52$ are excluded.

Electrons are required to have a small impact parameter with respect to the primary vertex:
This ensures that the electron comes from the HS, and not from a secondary decay.

Over the course of the data-taking, several read-out channels in the ID have failed, probably due to Electro Static Discharge (ESD) or Electrical Over Stress (EOS).\footnote{223} Hence, reconstructed electrons that pass through the ID in a dead region are rejected.

The spread in the \( z \)-coordinate of the collisions was observed to be different in MC than in data, as shown in figure 6.8. This difference has an up to 2\% effect on electron identification efficiency, and this needs to be corrected for. This is done by reweighting the events with the \( z \)-vertex reweighting tool.\footnote{224}


Muons
The muon cuts are almost similar to those of the electrons:

\[
|\eta_{\text{mu}}| < 2.5 \\
\text{\,\,\,\,\,\,} p_{T_{\text{mu}}} > 25 \text{ GeV} .
\] (6.11)

In contrast to electrons, the muon reconstruction is not affected by the crack region.

Similar to electrons, muons are also required to have a small impact parameter with respect to the primary vertex:

\[
|z_0| < 2 \text{ mm} .
\] (6.12)

Even though it is quite well shielded by the ground above it, some cosmic radiation still reaches the ATLAS detector. This radiation consists almost entirely out of muons. These muons can leave traces in the detector, and they need to be filtered out. They are identified by finding to back-to-back muon tracks that do not originate from a particle vertex.\[225\]

Photons
Even though photons are not explicitly used in this analysis, they are still needed for overlap removal later. The following cuts are performed:

\[
|\eta_{\text{ph}}| < 2.37 \\
\text{\,\,\,\,\,\,} E_{T_{\text{ph}}} > 15 \text{ GeV} .
\] (6.13)

Just as with the electrons, the crack region $1.37 < |\eta_{\text{ph}}| < 1.52$ is excluded here as well. In some recent studies, the cut on the crack region for photons has been extended to $|\eta| = 1.56$ in order to avoid degraded calibration performance.\[226\] This has not been done in this study.

6.5.2 - Overlap removal

After object selection, there will be several objects identified that originate from the same physical particle. This happens because a single particle can leave traces in the detector that will match the reconstruction criteria for different objects. These reconstructed objects will all point in the same direction; they overlap. Hence overlap removal is necessary in order to remove these objects, that would otherwise lead to ambiguities.

For jets, there is a small probability that (part of) a jet leads to a falsely identified electron. Because of the large cross section of the QCD multijets, the number of events with fake electrons that mimic the signal events can become significant. These "electrons" are called fake electrons. Fake (isolated) muons are only rarely produced by jets, but particle decays
can lead to genuine muons. These effects are also reduced by the overlap removal procedure, which is as follows:
1. Consider all reconstructed muons in the event.
   Remove the muon from the event if there is a jet with $p_T > 25$ GeV within $\Delta R < 0.4$;
2. Consider all reconstructed jets in the event.
   If there is an electron within $\Delta R < 0.2$, discard the jet;
   If there is a photon within $\Delta R < 0.1$, discard the jet;
3. Consider all reconstructed electrons in the event.
   If both the $\Delta \varphi$ and $\Delta \eta$ are $< 0.005$ with a muon (in other words, they share a track), remove the electron;
   If there is a jet within $\Delta R < 0.4$, remove the electron.

6.5.3 - Primary vertex

The primary vertex is the vertex of the hard scattering. In Monte Carlo, this is clearly defined, but in data, the primary vertex is defined as the reconstructed vertex that has the largest $\sum p_T^2$ (summing over its associated tracks). Vertices arising from decaying particles in the event are categorized as secondary vertices. Other vertices coming from pileup collisions are categorized as pile-up vertices.

We require that the primary vertex is reconstructed with at least four tracks. This requirement reduces soft collisions. All measurable tracks above the minimum $p_T$ of 400 MeV (with no other requirements) are used for this cut. An event without a reconstructed primary vertex, is rejected.

6.5.4 - Missing transverse energy

Although two incoming protons have balanced transverse momentum, the detectable final state can have missing transverse momentum due to escaping neutrinos. The magnitude of this momentum is called the missing transverse energy, $\not{E}_T$.

Events with an on-shell decaying $W$-boson that decays leptonically have in general a large $\not{E}_T$. Contrary, QCD events produce no or relatively small $\not{E}_T$, when final state hadrons in jets decay leptonically. Also, finite momentum resolution and misreconstructed muons can lead to (fake) $\not{E}_T$. Despite these effects, the $\not{E}_T$ is a powerful quantity to separate our signal from QCD background events.

The $\not{E}_T$ is calculated from the energy deposited in all calorimeter cells and from muons in the muon spectrometer. The $\not{E}_T$ is calibrated using the reconstructed particles (after overlap removal) in order to model the energy loss correctly. Additionally, a correction is applied for the energy lost in the cryostat, and for 8 TeV several additional pile-up suppression techniques are used as well.
In this thesis, the following cut is used:

\[ \mathbf{E_T} > 30 \text{ GeV} \, . \] (6.14)

In figure 6.9, the different Monte Carlo samples are shown without the \( \mathbf{E_T} \) cut applied. The QCD multijets and \( W^+ \text{jets} \) samples as used in the figures presented in this thesis are normalized with data-driven techniques as explained below. This cut rejects a background-dominated region of phase space, improving the \( S/B \) (signal over background) of the remaining events. The gray band represents the systematic uncertainties, which are determined in chapter 7. The QCD contribution has an uncertainty of 60% after all the cuts have been applied.

### 6.5.5 - Transverse \( W \)-boson mass

Another quantity that allows to reduce QCD events is the transverse \( W \) mass. When the \( W \)-boson decays leptonically, it is reconstructed using the charged lepton and the \( \mathbf{E_T} \), and its transverse mass can be written as:

\[ m_T^W \equiv \sqrt{p_T (l) \mathbf{E_T} - \mathbf{p_T}^2 (l) \cdot \mathbf{E_T}} \, . \] (6.15)
Events that do not contain a leptonically decaying $W$-boson will usually be constructed with a lower $m_T^W$. A cut on the $m_T^W$ removes these events. The cut used is:

$$m_T^W > 50 \text{ GeV}.$$  \hfill (6.16)

The effect of this cut can be seen in figure 6.10. Similar to the $E_T$ cut, a background-rich set of events is rejected by this cut, improving the signal selection. At low values of $m_T^W$, the prediction overshoots the data. The uncertainty on the QCD of 60% easily the difference in this region.

6.5.6 - Lepton-jet cut

QCD events are further reduced by a cut on the $p_T$ of the lepton, where the value of the cut depends on the azimuthal opening angle with the leading jet:

$$p_T (l) > 40 \text{ GeV} \left( 1 - \frac{\pi - |\Delta \phi (j_1, l)|}{\pi - 1} \right),$$  \hfill (6.17)
where \( j_1 \) stands for the leading jet in \( p_T \). This cut mainly reduces QCD events where the lepton has low \( p_T \) and is opposite to one of the jets in the transverse plane, as is illustrated in figure 6.11.

**6.6 - Data-driven backgrounds**

**6.6.1 - \( W^+ \)jets \( k \)-factors**

For the normalization of the \( W^+ \)jets background, \( k \)-factors are used with a data-driven background method as described in section 4.1, which uses the 2-jetbin. The 2-jetbin is also populated with the single top t-channel signal events, but for the used pre-selection, its contribution is small in comparison with the background. In addition, in order to avoid a circular dependency between the signal measurement and background determination, we first apply a veto on signal events. The signal veto uses the events passing the final selection introduced in the next chapter, with an additional requirement on the reconstructed top mass. This requirement vetoes t-channel events from the signal enriched region.\textsuperscript{34}
Table 6.2: The 8 TeV k-factors for the nominal sample. (a) gives the electron stream, (b) the muon stream.

<table>
<thead>
<tr>
<th>$N_{\text{jet}}$</th>
<th>$k_{bb,i}$</th>
<th>$k_{c,i}$</th>
<th>$k_{l,i}$</th>
<th>$N_{\text{jet}}$</th>
<th>$k_{bb,i}$</th>
<th>$k_{c,i}$</th>
<th>$k_{l,i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.979</td>
<td>1.03</td>
<td>0.856</td>
<td>1</td>
<td>1.53</td>
<td>1.03</td>
<td>0.942</td>
</tr>
<tr>
<td>2</td>
<td>0.979</td>
<td>1.03</td>
<td>0.856</td>
<td>2</td>
<td>1.54</td>
<td>1.04</td>
<td>0.948</td>
</tr>
<tr>
<td>3</td>
<td>0.936</td>
<td>0.981</td>
<td>0.818</td>
<td>3</td>
<td>1.52</td>
<td>1.03</td>
<td>0.938</td>
</tr>
<tr>
<td>4</td>
<td>0.914</td>
<td>0.959</td>
<td>0.799</td>
<td>4</td>
<td>1.53</td>
<td>1.03</td>
<td>0.941</td>
</tr>
<tr>
<td>$\geq 5$</td>
<td>0.637</td>
<td>0.668</td>
<td>0.557</td>
<td>$\geq 5$</td>
<td>1.24</td>
<td>0.839</td>
<td>0.766</td>
</tr>
</tbody>
</table>

Table 6.2 gives the values of the $k$-factors at 8 TeV, for the nominal sample. As $k_{cc,i}$ is identical to $k_{bb,i}$, it is not listed in the tables.

One of the systematic effects to be considered is the extrapolation uncertainty as discussed in chapter 4. Various other systematic effects are considered as well to determine the uncertainty on the t-channel cross section measurement; these will be discussed in the next chapters. For each systematic effect, the $k$-factors are re-derived.

It turns out that the systematic variation on $k_{bb,i}$ and $k_{c,i}$ is about 50%, strongly anti-correlated with each other. The uncertainty on the normalization of the tagged $W^+\text{jets}$ background is smaller, below 10%.

### 6.6.2 - QCD multijets

The QCD multijets process cannot be accurately modeled in Monte Carlo simulations. Therefore, the QCD contribution is obtained through a data-driven method as well. In this section, we estimate the QCD background using the matrix method.\[^{232-233}\]\[^{35}\]

The single top t-channel analysis requires exactly one hard lepton in the event. In the QCD process, no hard leptons are produced in the HS, but the parton shower occasionally leads to isolated energy deposits, which can be falsely reconstructed as prompt leptons. The matrix method is developed to measure the number of QCD events in data by determining the number of events with fake leptons passing signal cuts.

---

34. This selection also removes background from data and MC predictions. At first, the MC prediction is not yet normalized with $k$-factors. To check this effect, another iteration is made and the $k$-factors are re-derived. The impact turns out to be negligible.

35. The matrix method derives its name from the matrix that can be constructed that represents the transformation from the loose-tight categories to the real-fake categories.

118
The number of events with a lepton that passes the loose (tight) lepton selection is denoted \( N^{\text{loose}} (N^{\text{tight}}) \). The yet unknown number of events with real (fake) leptons is denoted as \( N_{\text{real}} (N_{\text{fake}}) \). The QCD background contribution to the selected signal region is then given by the fake leptons that also pass the tight lepton selection, \( N_{\text{fake}}^{\text{tight}} \).

The two efficiencies \( \epsilon_{\text{real}} \) and \( \epsilon_{\text{fake}} \) are defined as the fraction of real and fake lepton events respectively that pass the tight lepton requirements after already passing the loose selection:

\[
\epsilon_{\text{real}} = \frac{N_{\text{real}}^{\text{tight}}}{N_{\text{real}}^{\text{loose}}},
\]

\[
\epsilon_{\text{fake}} = \frac{N_{\text{fake}}^{\text{tight}}}{N_{\text{fake}}^{\text{loose}}}. \tag{6.19}
\]

The total number of events passing the tight lepton selection \( (N^{\text{tight}}) \) are then expressed as the sum of the events passing the tight lepton selection in both the fake and real categories:

\[
N^{\text{tight}} = \epsilon_{\text{real}} N_{\text{real}}^{\text{loose}} + \epsilon_{\text{fake}} N_{\text{fake}}^{\text{loose}}. \tag{6.20}
\]

This equation can be rewritten to give the events that comprise the data-derived QCD contribution:

\[
N_{\text{fake}}^{\text{tight}} = \frac{\epsilon_{\text{fake}}}{\epsilon_{\text{real}} - \epsilon_{\text{fake}}} (N_{\text{loose}}^{\text{real}} \epsilon_{\text{real}} - N^{\text{tight}}). \tag{6.21}
\]

Both \( N^{\text{loose}} \) and \( N^{\text{tight}} \) are measured directly in data. The \( \epsilon_{\text{real}} \) and \( \epsilon_{\text{fake}} \) are estimated in regions of phase space that are enriched with real leptons (for example, using \( Z \rightarrow ll \) events) or fake leptons (for example, events with very little \( E_T \)).

Using the derived QCD contribution, the QCD sample is generated from data events passing the loose lepton selection, by applying the derived QCD weight factor. This new sample is then used as the background sample. The accuracy of this method is estimated to be 60%, based on variation of requirements and comparison to alternative methods.

### 6.7 - Event yields and control plots

For the pre-selection, we only accept events that have two jets that pass the jet requirements, of which one must be tagged as a \( b \)-jet. The events are divided into two streams; an electron stream, and a muon stream. This is done by requiring exactly one electron and zero muons, and exactly one muon and zero electrons respectively.
Chapter 6

<table>
<thead>
<tr>
<th>Jets:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
</tr>
<tr>
<td>$p_{T_{\text{jet}}} \geq 30 \text{ GeV}$</td>
</tr>
<tr>
<td>$p_{T_{\text{jet}}} \geq 35 \text{ GeV}$ (if $2.5 &lt;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Electrons:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
</tr>
<tr>
<td>$E_{T_{\text{el}}} &gt; 25 \text{ GeV}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Muons:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
</tr>
<tr>
<td>$p_{T_{\text{mu}}} &gt; 25 \text{ GeV}$</td>
</tr>
<tr>
<td>$E_{T} &gt; 30 \text{ GeV}$</td>
</tr>
<tr>
<td>$m_{T}^{W} &gt; 50 \text{ GeV}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Electron stream:</th>
<th>Muon stream:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_{\text{el}} = 1$</td>
<td>$N_{\text{el}} = 0$</td>
</tr>
<tr>
<td>$N_{\text{mu}} = 0$</td>
<td>$N_{\text{mu}} = 1$</td>
</tr>
<tr>
<td>$N_{\text{jet}} = 2$</td>
<td>$N_{\text{jet}} = 2$</td>
</tr>
<tr>
<td>$N_{b_{\text{-jet}}} = 1$</td>
<td>$N_{b_{\text{-jet}}} = 1$</td>
</tr>
</tbody>
</table>

Table 6.3: Summary of the pre-selection cuts. (Not all cuts are indicated.)

After all the cuts mentioned above have been applied (summarized in table 6.3), the events as given in table 6.4 remain. The event yields show a good agreement between the expected number of events (Monte Carlo) and the observed number of events (data).

Several distributions of the pre-selected events are shown in figure 6.12, 6.13, 6.14, and 6.15. These show the lepton $p_{T}$, lepton $\eta$, $E_{T}$, and $m_{T}^{W}$ respectively. These figures exhibit a good agreement between data and MC, within the error bars. (The gaps in the electron $\eta$ distribution around the absolute value of 1.4 are caused by the crack region.)
Event reconstruction and pre-selection

<table>
<thead>
<tr>
<th>Sample</th>
<th>Electron</th>
<th>Muon</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-tag</td>
<td>Tagged</td>
</tr>
<tr>
<td>Data</td>
<td>1437867</td>
<td>60757</td>
</tr>
<tr>
<td>Single top s-channel</td>
<td>1100</td>
<td>520</td>
</tr>
<tr>
<td>Single top t-channel</td>
<td>19076</td>
<td>8880</td>
</tr>
<tr>
<td>Single top Wt-channel</td>
<td>7007</td>
<td>2908</td>
</tr>
<tr>
<td>Wq</td>
<td>694335</td>
<td>1648</td>
</tr>
<tr>
<td>Wbb</td>
<td>30170</td>
<td>7790</td>
</tr>
<tr>
<td>Wc</td>
<td>201940</td>
<td>7702</td>
</tr>
<tr>
<td>Wcc</td>
<td>93790</td>
<td>2312</td>
</tr>
<tr>
<td>Zq</td>
<td>100260</td>
<td>161</td>
</tr>
<tr>
<td>Zbb</td>
<td>114211</td>
<td>1687</td>
</tr>
<tr>
<td>Zcc</td>
<td>9130</td>
<td>201</td>
</tr>
<tr>
<td>Diboson</td>
<td>13151</td>
<td>233</td>
</tr>
<tr>
<td>t\bar{t}</td>
<td>46598</td>
<td>20705</td>
</tr>
<tr>
<td>QCD</td>
<td>102131</td>
<td>6221</td>
</tr>
<tr>
<td>Expected</td>
<td>1432895</td>
<td>60973</td>
</tr>
<tr>
<td>Observed</td>
<td>1437867</td>
<td>60757</td>
</tr>
</tbody>
</table>

Table 6.4: The event yields for the (nominal) pre-selection events, both pre-tagged and tagged.

![Figure 6.12](image)

(a) (b)

Figure 6.12: This figure shows the $p_T$ of the selected lepton in the pre-selection events. (a) shows the electron stream; (b) shows the muon stream.
Figure 6.13: This figure shows the η of the selected lepton in the pre-selection events. (a) shows the electron stream; (b) shows the muon stream.

Figure 6.14: This figure shows the $E_T$ in the pre-selection events. The highest bin is used as an overflow bin. (a) shows the electron stream; (b) shows the muon stream.

6.8 - Conclusion

We have discussed the reconstruction of measured objects with the ATLAS detector. In preparation for the single top t-channel analysis a pre-selection has been introduced to enrich the sample with top-quark events. The QCD background is extracted from data and the
Figure 6.15: This figure shows the $m_W^T$ in the pre-selection events. The highest bin is used as an overflow bin. (a) shows the electron stream; (b) shows the muon stream.

background for $W$+jets uses MC events is normalized by a data driven procedure. Finally, the distributions of several quantities for the pre-selection are presented. Also in these plots, the gray band represents the systematic uncertainties. The agreements between data and the predictions is excellent.

In the next chapter, we will use these pre-selected events as a starting point for a single top t-channel analysis.
Chapter 6
The measurement of Standard Model processes at new energy scales is interesting by itself and provides an important check of the theory. Deviations of the Standard Model predictions would indicate contributions of unknown physics processes. To unravel the properties of potential unknown processes requires a combination of several measurements. Our measurement presented here corresponds to the first step towards a model independent search and flags any type of new physics as it emerges through the deviation of the predicted single top t-channel cross section. We use 8 TeV data from ATLAS and a template fit to the top-quark mass distribution to perform a cross section measurement, and to compare the result to the theoretical prediction.

We will measure the total cross section $\sigma_{pp \rightarrow X}$ of single top t-channel production through:

$$\sigma = \frac{N_{\text{sel, data}}}{\epsilon L},$$  \hspace{1cm} (7.1)

where $L$ represents the data luminosity, and the number of signal events that is extracted from data after background subtraction is represented by $N_{\text{sel, data}}$. The reconstruction efficiency is given by:

$$\epsilon \equiv \frac{N_{\text{det}}}{N_{\text{total}}},$$  \hspace{1cm} (7.2)

where $N_{\text{det}}$ is the number of single top t-channel Monte Carlo events passing the selection cuts, and $N_{\text{total}}$ is the total number of events in the MC signal sample with leptonic decays ($5.8 \times 10^5$ events). These efficiencies include the branching ratio of all leptonic decays to electrons and muons respectively. The selected electron or muon events contain a small fraction of signal events with $W \rightarrow \tau \nu$ events in which the tau decays leptonically.
Chapter 7

7.1 - Single top t-channel selection

Throughout the previous chapter we discussed the event pre-selection (including 1 $b$-tagged jet) for single top events with a $W$-boson that decays leptonically. The large backgrounds of mainly QCD multijets and $W^+jets$ events have both been normalized using data driven techniques. That is the starting point of this section and we introduce additional requirements to isolate single top events. In the plots we display the backgrounds individually, using the legend as shown in figure 7.1. The shaded "Sys. uncertainty" band represents the systematic uncertainties without the contribution of alternative signal generators; see section 7.5.

The first quantity used in the selection is a measure of the total transverse energy in the event, called $H_T$. The $H_T$ is the scalar sum of the $p_T$ of all the objects in the event, including the $\vec{E}_T$. We require:

$$H_T > 210 \text{ GeV}.$$  \hspace{1cm} (7.3)

The effect of this selection cut is illustrated in figure 7.2. The cut removes most of the QCD and a substantial part of the $W^+jets$ background. The measured distribution of $H_T$ is...
reasonably well described by the MC simulation. In table 7.1 the number of events for the electron and muon stream are shown before and after this cut.

Next, a cut on the $\eta$ of the light jet is applied, as shown in figure 7.3. The spectator (light) quark that is produced in the hard scattering (the Feynman diagram in figure 2.19b) is expected to emerge in the forward direction:
Chapter 7

The cut on the $\eta$ of the light jet significantly improves the signal to background ratio from 17.6% (16.8%) in the electron (muon) stream to 55.4% (50.4%).

After that, a cut on the $\Delta \eta$ between the $b$-jet and the light jet is applied, which are expected to be well separated in t-channel events when compared to background events.

$$|\eta_{\text{-jet}}| > 2.0.$$ (7.4)

$$|\Delta \eta_{b\text{-jet, l\text{-jet}}}| > 1.0.$$ (7.5)

Figure 7.4 shows the corresponding distribution. The region that gets removed by this requirement is mainly occupied by background events. After this cut the expected signal to background ratio is 0.39 (0.37) for the electron (muon) sample.

7.2 - Top mass distribution

To reconstruct the four-momentum of the top-quark, all its decay products have to be identified. The charged lepton and the $b$-jet are measured, but the neutrino escapes undetected. The transverse momentum of the neutrino can be inferred from the measured $p_T$, assuming the momentum imbalance originates only from this decay product and thus ignoring mismeasurements in the calorimeter and additional neutrinos from soft decays in
Table 7.1: The event yields for the nominal samples. The first row is the starting point after pre-selection. Then the final cuts are turned on sequentially from top to bottom, with the results are listed in the corresponding rows.

The longitudinal momentum of the neutrino, $p_{z, \nu}$, cannot be determined from momentum imbalance, as most of the remnants escape through the beampipe, taking along the majority of the collision's longitudinal energy. However, the $p_{z, \nu}$ can be reconstructed using other variables and kinematic constraints after which a twofold ambiguity remains.\[234\]

The four-momentum of the $W$-boson in the event is constructed by summing the four-momenta of the lepton and neutrino. Since the $W$ is on-shell, its mass equals the invariant mass of the charged lepton and neutrino. This leads to a quadratic equation with $p_z$ as the unknown:

$$p_{z, \nu}^2 - 2 \frac{\mu}{E_l^2 - p_{z,l}^2} p_{z, \nu} + \frac{E_l^2}{E_l^2 - p_{z,l}^2} p_{T, \nu}^2 - \mu^2 = 0 , \quad (7.6)$$

with:

$$\mu \equiv \frac{m_{W^\pm}}{2} + \cos (\Delta \phi) \ p_{T,l} \ p_{T, \nu} , \quad (7.7)$$

where $\Delta \phi$ denotes the azimuthal angle between the lepton and the $E_T$.

The quadratic equation 7.6 leads to two solutions for the longitudinal momentum of the neutrino, which are given by:

$$p_{z, \nu} = \frac{\mu}{p_{T,l}^2} \ p_{z,l} \pm \sqrt{\frac{\mu^2 p_{z,l}^2}{p_{T,l}^4} - \frac{E_l^2}{p_{T,l}^2} p_{T, \nu}^2 - \mu^2} \quad (7.8)$$
Figure 7.5: The distribution of the neutrino $p_z$ in the pre-selection events. The highest bin is used as an overflow bin. (a) shows the electron stream; (b) the muon stream.

In case the discriminant of equation 7.8 is positive, we use the smallest value of $p_{z,\nu}$. However, the discriminant becomes negative if the $m_T^W$ is larger than $m_W$, which can occur due to the resolution of the $E_T$ measurement. In this case, we force the discriminant to zero by reducing the $E_T$ until the $m_T^W$ is equal to $m_W$. With this modified $p_{z,\nu}$ and $p_{y,\nu}$ the calculation of $p_{z,\nu}$ results in a real solution.

Figure 7.5 shows the distribution of the reconstructed neutrino's $p_z$ in MC and data, demonstrating good agreement.

Figure 7.6 displays the top mass distribution, after the final selection requirements. The contribution of single top t-channel is clearly visible and in good agreement with the predictions.

The predicted number of signal events $N_{det} = 2910 \text{ (3837)}$ for the electron (muon) stream. This corresponds to a total selection efficiency of 0.51% and 0.67% for the electron and muon stream respectively.

7.3 - Template fit procedure

As a next step, we estimate the top-quark events in the remaining distributions. In previous studies this was done by further increasing the top-quark purity by placing a cut on the mass.$^{[235-236]}$
Single top t-channel cross section

Figure 7.6: The reconstructed top mass distribution after the $\Delta m$ cut, i.e. the final event selection. (a) shows the electron stream; (b) shows the muon stream.

This cut removes a large portion of the top mass distribution, as can be seen in figure 7.6. Instead, in this analysis we fit the complete top mass distribution with a binned likelihood fitting procedure, separately for the electron and muon streams. The background contribution is enriched in the sidebands, which allow to determine its normalization during the fitting procedure.

The dominant backgrounds are $W$+jets and $t\bar{t}$ events. The $t\bar{t}$ background is generated to NLO accuracy where previous measurements have confirmed the predictions. For $t\bar{t}$ normalization we use the NNLO theoretical prediction, constrained in our fitting procedure with the corresponding uncertainty. The $W$+jets background is already normalized using the data driven procedure that involves separate $k$-factors for different heavy flavor contributions. The final selection is kinematically different, and therefore the fitting procedure is also used to determine the overall normalization of the $W$+jets background again. The other background samples have only small contributions and are fixed to their predictions. The modest contribution of QCD multijets has been determined from data as described in chapter 6.

The number of events in a particular bin of the mass distribution of figure 7.6, $N_i$, is parametrized by:

$$N_i = \alpha_S N_{S,i} + \alpha_W N_{W,i} + \alpha_{t\bar{t}} N_{t\bar{t},i} + N_{B,i},$$  \hspace{1cm} (7.10)
where $N_{S,i}$ represents the MC t-channel contribution in bin $i$. The total number of selected MC t-channel events is $N_{\text{det}} = \sum_i N_{S,i}$. The predicted $W$+jets background is represented by $N_{W,i}$, the predicted $t\bar{t}$ background by $N_{t\bar{t},i}$, and all other backgrounds by $N_{B,i}$. These quantities are taken from the predicted distributions, also called templates. The three parameters $\alpha_S$, $\alpha_W$, and $\alpha_{t\bar{t}}$ are determined in an extended binned likelihood fit with a Gaussian constraint for $\alpha_{t\bar{t}}$ around unity with a width of 0.059, corresponding to the theoretical uncertainty. The other two fit parameters are allowed to vary arbitrarily.

To evaluate the effects of the statistical uncertainties of the templates, 100,000 pseudo experiments are used. For each pseudo experiment all template-bins are fluctuated randomly according to their statistical uncertainty. This results in an uncertainty of $\pm 2.4\%$ ($\pm 2.2\%$) on $\alpha_S$ for electrons (muons) respectively.

**Linearity check**

To check that the fit is able to distinguish between the single top t-channel signal and the $W$+jets background, a linearity test is performed. In this test the fit procedure is repeated after the signal Monte Carlo is varied with an artificial scale factor. As shown in figure 7.7, the results demonstrate that the procedure is robust.

Figure 7.7: A linearity test of the fitting procedure. The signal Monte Carlo is varied with a scale factor, and the fit result for $\alpha_S$ and $\alpha_W$ is shown. The error bars represent the statistical error as reported by MINUIT only and are fully correlated between the scale factors.
7.4 - Fitting to data

Figure 7.8 shows the result of the fit: the top mass distribution after applying the scale factors obtained in the fitting procedure, on both the electron and muon stream. The data is well described by the Monte Carlo prediction, which is also confirmed by the ratio plots.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Number of events</th>
<th>Fitfactor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Electron</td>
<td>Muon</td>
</tr>
<tr>
<td>$N_{\text{sel,data}}$</td>
<td>2639±108</td>
<td>3827±126</td>
</tr>
<tr>
<td>Single top t-channel</td>
<td>2626±107</td>
<td>3784±127</td>
</tr>
<tr>
<td>$W^+\text{jets}$</td>
<td>1590±114</td>
<td>2725±136</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>1920±105</td>
<td>2625±134</td>
</tr>
</tbody>
</table>

Table 7.2: The number of events after applying the scale factors, and the scale factors themselves for the nominal sample. The error indicated is the uncertainty associated with the fit.

Table 7.2 shows the number of events and the scale factor for the nominal sample. Only statistical uncertainties are given. For the cross section calculation we use $N_{\text{sel,data}}$ which is also given in the table. This value is obtained by subtraction of all (fitted) backgrounds from the data, through:
Chapter 7

\[ N_{\text{sel,data}} = N_{\text{data}} - \sum_{\text{bins}} (\alpha_W N_{W,i} + \alpha_{t\bar{t}} N_{t\bar{t},i} + N_{B,i}), \]  

(7.11)

with \( N_{\text{data}} \) the number of all selected events. The resulting \( N_{\text{sel,data}} \) is compatible with the fitted number of t-channel events in the template as expected.

<table>
<thead>
<tr>
<th></th>
<th>Electron</th>
<th>Muon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single top t-channel</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( W^{+}\text{jets} )</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>( t\bar{t} )</td>
<td>-0.27</td>
<td>-0.18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Electron</th>
<th>Muon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single top t-channel</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( W^{+}\text{jets} )</td>
<td>1.00</td>
<td>-0.52</td>
</tr>
<tr>
<td>( t\bar{t} )</td>
<td>-0.43</td>
<td>-0.56</td>
</tr>
</tbody>
</table>

Table 7.3: Correlation matrix for the fitting procedure.

Table 7.3 lists the correlations between the parameters in the electron and muon stream. For both, the correlations between the parameters are modest, again signaling that the fit is able to distinguish background from signal.

The results for \( \alpha_W \) in the electron and muon stream are both below unity and are in reasonable agreement with each other. The results for \( t\bar{t} \) are for both electron and muon streams are above unity and in agreement. It should be noted that there is a negative correlation between \( \alpha_W \) and \( \alpha_{t\bar{t}} \). The result for the signal in the electron and muon stream are both below unity, but not in perfect agreement.

It will be interesting to see how much the systematic uncertainties will bridge the (small) gap, which is the subject of the next section.

7.5 - Sources of systematic uncertainties

The analysis presented this chapter is sensitive to several sources of systematic uncertainties. These uncertainties affect the shape of MC templates used in the fit, and/or the efficiency \( \epsilon \) (equation 7.2) to calculate the full cross section. Also, the uncertainty on the luminosity and the MC statistics have to be taken into account.

The sources of systematic uncertainties that need to be addressed can be divided in two groups. The first group is related to the experimental and detector effects, which includes uncertainties on energy calibration and the detection efficiency of objects and events. When
more data becomes available these effects get better understood and may be better controlled, but intrinsically these effects do not follow the $\frac{1}{\sqrt{N}}$ law. The systematic effects described here are the result of a collaborative effort spanning many years to understand the ATLAS detector.

The second group of systematic effects are the results of the use of the theoretical predictions and models used in the MC simulation. These simulations are based on experience gathered in many decades of theoretical development. In this section a short overview of all sources of systematic effects is given.\[237\]

In the description below the names of the uncertainties, listed in brackets, are given as these appear in tables and in further discussion.

Jet uncertainties
The systematic uncertainties that affect the jets are the jet reconstruction efficiency (jeff), the jet energy resolution (jer), the jet energy scale (jes), and the jet vertex fraction (jvf) efficiency and fake rate. It turns out that the jes is the most important variation. The uncertainty of the jet energy scale is determined by re-evaluating the event with the energies of simulated jets shifted up and down by one standard deviation. The size of the shift depends on the $p_T$ and $\eta$ of the jet, and is mostly between 1% and 6%.\[221\] The jet energy resolution uncertainty is evaluated by smearing the resolution in simulation. The jet vertex fraction uncertainty is estimated by varying the jvf-calibration up and down within the difference between MC and data.\[238-239\]

$b$-tagging uncertainties
There are three systematic uncertainties associated with $b$-tagging: btag_btag, btag_ctautag, and btag_mistag. These denote the uncertainties in the tagging efficiencies of $b$-jets, $c$/tau-jets, and light jets respectively. These uncertainties are evaluated by varying the MC scale factors used to calculate the $b$-tagging efficiencies. The uncertainties depend mostly on the energy of the tagged jet, resulting in an uncertainty 2% to 8% for the $b$-tag rate, 8% to 15% for the $c$-tag rate, and 15% to 40% for the mistag rate.\[208,240\]

Electron uncertainties
Specifically for electrons, there are two systematic uncertainties: the electron energy resolution (eer), and the electron energy scale (ees). These combine to an about 1.5% effect.\[241\]

Muon uncertainties
For muons, there are three systematic uncertainties: the muon identification (muid), the muon momentum scale (mums), and the muon momentum resolution (musc). The muon momentum scale uncertainty is ~0.05% and the momentum resolution uncertainty ~3%.\[242\]
Lepton uncertainties
For both the electrons and muons, there are also the lepton reconstruction efficiency (sf_Id), and the uncertainty associated with the lepton trigger efficiency (trigger_sf). All the uncertainties associated to lepton are evaluated in a similar way as the jet uncertainties. The uncertainties turn out to be smaller than that of the jets.[231]

$E_T$ uncertainties
There are two systematic uncertainties that arise during the $E_T$ reconstruction: the uncertainty on the soft energy resolution used (res_soft), and the uncertainty on the soft energy scale (sc_soft). The effects of these uncertainties on the MET are of the order of 2% to 3%.[243]

Monte Carlo uncertainties - background estimation
The following systematic effects are all related to the use of MC prediction for the shape and normalization of background distributions:

- The uncertainty on the choice of the $W^\pm$jets generator is estimated by replacing the Alpgen events with Sherpa samples (SherpaW). The deviation from the nominal is symmetrized, and used as an uncertainty. (Several additional uncertainties associated in the $W^\pm$jets estimation are discussed below.)
- The uncertainty on the choice of the MC generator used for the $t\bar{t}$ MC sample is estimated by replacing the Powheg sample by an MC@NLO sample (ttbarMcAtNlo). This is also symmetrized before use;
- The effects from uncertainty on the $t\bar{t}$ cross section measurement are also estimated. Here, the uncertainty of $\pm 5.9\%$ mentioned earlier is used (xsecvar_tbarPoPy).
Monte Carlo uncertainties - signal extraction

The MC signal predictions enter the analysis when the shape of the signal template used in the fitting procedure is extracted for the determination of $N_{\text{sel, data}}$. To estimate the uncertainty on $N_{\text{sel, data}}$ we use another high statistics signal sample produced with the LO generator AcerMC. Figure 7.9 shows the reconstructed top mass distribution for this sample. We find a modest variation of 4.7% (4.6%) for electrons and muons respectively. We will use the symmetrized values.

Monte Carlo uncertainties - phase space extrapolation

Another uncertainty originating in the MC signal samples is due to the extrapolation of the cross section from the selected phase space to the full phase space. For this extrapolation (which is ratio between $N_{\text{det}}$ and $N_{\text{total}}$) we use the variation of AcerMC and also that of an aMC@NLO sample that has a factor 300 lower statistics. For AcerMC we find 15.6% for the electron and 14.6% for the muon streams, and for aMC@NLO 0.0% and 6.7% respectively. We call this systematic check the 'phase space extrapolation'. We use the maximal variation and symmetrize the effect around the combined central value, and we obtain $\pm 7.8\%$ for the electron and $\pm 7.3\%$ for the muon streams respectively. The large differences between these MC generator predictions is the main motivation for the measurement of the so-called fiducial cross section, presented in the next chapter.

$W^+\text{jets background}$

There are several uncertainties associated with the $k$-factor calculation. First, the uncertainties associated with the heavy flavor fractions determination as discussed in chapter 4. We are affected by the theoretical uncertainty on the ratio of heavy flavors fractions between events with one jet and that of two jets. We use the values listed in table 4.5 to calculate the effects which we list here as "kfacvar_WbbWccjet1" and "kfacvar_Wcjet1". Also, the statistical uncertainty of the $W^+\text{jets}$ MC samples is taken into account: "kfacvar_Kbb", "kfacvar_Kc","and "kfacvar_Kll".

In addition, the signal prediction is used in the data driven procedure to estimate the $W^+\text{jets}$ and QCD background. To assess the possible circular dependence, we apply a 30% variation on the single top t-channel cross section of the Monte Carlo sample, and call this uncertainty "tchan". This conservative variation is based on an initial guess of the total uncertainty of the cross section measurement and should be readdressed when its impact is significant.

QCD uncertainties

There are three systematic uncertainties associated with the QCD data-driven method:[244] "fake_alternate": the uncertainty caused by using an alternative way to get fake efficiency, using a different control region (CR).
"fake_mc": a 10% variation on the normalization of the Monte Carlo samples used for the subtraction from the fake CR.
"real_alternate": the uncertainty associated with an alternative way to get the real efficiency.
The total variation of the QCD background amounts to roughly 60%.
PDF uncertainties

The "pdfvar" systematic variation is calculated by re-running the entire analyses with a signal sample with a different PDF set\cite{245-246} Instead of generating new samples, the existing sample is modified by reweighting the event, using the LHAPDF library\cite{247-248} For this, three different NLO sets are used: MSTW2008nlo, CT10\cite{33-34,39,41-43,46-60,249-256} and NNPDF2.3\cite{33-34,41-43,53,56,59-60,250,254-276}

The combined relative uncertainty is then determined by the envelope of all uncertainties of the three sets individually.

Other uncertainties

And finally, there are several other systematic uncertainties that are also taken into account. "fit" is the uncertainty on the fit procedure due to MC statistics, by varying the fit scale factor for the MC background with their uncertainty as determined by the pseudo experiments in section 7.3. "lumi" is the uncertainty on the luminosity measurement, which is ±2.8%\cite{277}

7.5.1 - Systematic effects

The systematic uncertainties on the measured cross section are evaluated by applying each systematic variation and re-measure the t-channel cross section. The effect on the cross section is listed in table 7.4. Most experimental effects are modest of order 1%. The largest effect is 5%; this is from the uncertainty in the b-tagging. (For completeness, the results on the event yields and fitted parameters for the \(W^+\)jets and \(t\bar{t}\) background and for the t-channel signal are given in appendix B). The choice of the MC generators leads to relative large uncertainties as well. Below, we will especially focus on the t-channel signal generator.

The results, separate for upward and downward variations, can be regarded as practically uncorrelated and are therefore combined in quadrature for the total uncertainty. The total systematic uncertainty for the electron channel is 12% (12%) up (down), and for the muon channel it is 12% (12%) up (down). Without the large AcerMC and aMC@NLO systematic uncertainties (which are the most dominating systematical uncertainties), the result is ±9.3% for electrons and +9.0% −8.9% for muons.

The systematic uncertainties of the electron and muon streams are mostly correlated, but some systematic sources appear to have a slightly different impact. Obviously, the effects of the MC statistics ("fit") and the lepton reconstruction are uncorrelated between the channels. If we add only the differences and these uncorrelated effects, we obtain a total uncorrelated uncertainty on the cross section of +11.1% (−7.2%) and +6.3% (−10.3%) for the electron and muon stream up (down) respectively.

In the previous section, we saw the small difference in signal scale factors of 0.902 ± 0.037 and 0.986 ± 0.033. With both statistical and uncorrelated systematical uncertainties taken
### Single top t-channel cross section

<table>
<thead>
<tr>
<th>Variation</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Electron</td>
</tr>
<tr>
<td><strong>Jet uncertainties</strong></td>
<td></td>
</tr>
<tr>
<td>jeff</td>
<td>+0.3%</td>
</tr>
<tr>
<td>jer</td>
<td>+0.9%</td>
</tr>
<tr>
<td>jes</td>
<td>-2.8% -2.0%</td>
</tr>
<tr>
<td>jvf</td>
<td>-0.8% +1.6%</td>
</tr>
<tr>
<td><strong>b-tagging uncertainties</strong></td>
<td></td>
</tr>
<tr>
<td>btag_btag</td>
<td>-4.9% +5.4%</td>
</tr>
<tr>
<td>btag_ctautag</td>
<td>-0.0% +0.1%</td>
</tr>
<tr>
<td>btag_mistag</td>
<td>-0.0% +0.3%</td>
</tr>
<tr>
<td><strong>Electron uncertainties</strong></td>
<td></td>
</tr>
<tr>
<td>eer</td>
<td>-0.8% -0.3%</td>
</tr>
<tr>
<td>ees</td>
<td>-3.3% +3.8%</td>
</tr>
<tr>
<td><strong>Muon uncertainties</strong></td>
<td></td>
</tr>
<tr>
<td>muid</td>
<td>-</td>
</tr>
<tr>
<td>mums</td>
<td>-</td>
</tr>
<tr>
<td>musc</td>
<td>-</td>
</tr>
<tr>
<td><strong>Lepton uncertainties</strong></td>
<td></td>
</tr>
<tr>
<td>sf_Id</td>
<td>-3.0% +3.2%</td>
</tr>
<tr>
<td>sf_reco</td>
<td>±0.3%</td>
</tr>
<tr>
<td>trigger_sf</td>
<td>±0.8%</td>
</tr>
<tr>
<td><strong>E_T uncertainties</strong></td>
<td></td>
</tr>
<tr>
<td>res_soft</td>
<td>-1.9% +0.0%</td>
</tr>
<tr>
<td>sc_soft</td>
<td>-0.5% +0.4%</td>
</tr>
<tr>
<td><strong>Monte Carlo uncertainties - background estimation</strong></td>
<td></td>
</tr>
<tr>
<td>SherpaW</td>
<td>±3.1%</td>
</tr>
<tr>
<td>ttbarMcAtNlo</td>
<td>-0.4%</td>
</tr>
<tr>
<td>xsecvar_ttbarPoPy</td>
<td>-1.9% +2.2%</td>
</tr>
<tr>
<td><strong>Monte Carlo uncertainties - signal extraction</strong></td>
<td></td>
</tr>
<tr>
<td>AcerMC</td>
<td>±2.3%</td>
</tr>
<tr>
<td><strong>Monte Carlo uncertainties - phase space extrapolation</strong></td>
<td></td>
</tr>
<tr>
<td>Phase space extrapolation</td>
<td>±7.8%</td>
</tr>
<tr>
<td><strong>W+jets background</strong></td>
<td></td>
</tr>
<tr>
<td>tchan</td>
<td>-0.1% +0.3%</td>
</tr>
<tr>
<td>kfacvar_Kbb</td>
<td>±0.2%</td>
</tr>
<tr>
<td>kfacvar_Kc</td>
<td>±0.1%</td>
</tr>
<tr>
<td>kfacvar_Kll</td>
<td>±0.0%</td>
</tr>
</tbody>
</table>
Table 7.4: Breakdown of the systematic uncertainties after the fitting procedure.

into account, these differences are not significant, and we conclude that the results are in agreement.

7.6 - Measured cross section

By using the number of signal events obtained from the fitting procedure, the measured cross sections for the electron and muon streams can be calculated through equation 7.1, and correcting by a factor 9/3 for the branching ratio to leptonic decays:

\[
\sigma_{t-channel,el} (8 \text{ TeV, data}) = 78.9 \pm 3.2 \text{ (stat.)}^{+9.8}_{-9.8} \text{ (syst.)} \text{ pb} \cdot 
\]

\[
\sigma_{t-channel,mu} (8 \text{ TeV, data}) = 86.8 \pm 2.9 \text{ (stat.)}^{+10.3}_{-10.2} \text{ (syst.)} \text{ pb} \cdot 
\]

Note that \(\sigma_{t-channel,el} (\sigma_{t-channel,mu})\) represents the cross section for signal events with the \(W\)-boson directly decaying to electrons (muons). Thus, the indirect contributions from the \(W\) to tau-decays have been removed.

As we discussed in the previous section, part of the total systematic uncertainty is an uncorrelated effect which is +8.8 (−5.7) pb and +5.5 (−8.9) pb for electrons and muons up
(down) respectively. Hence, the results for electrons and muons are in good agreement. When the results of both streams are combined, taking into account the correlation between systematic effects, we obtain for a single top t-channel cross section at 8 TeV:

\[ \sigma_{t\text{-channel}} (8 \text{ TeV, data}) = 83.4 \pm 2.1 \, \text{(stat.)}^{+9.8}_{-9.6} \, \text{(syst.)} \, \text{pb} . \]  

(7.14)

The theoretical prediction of the same cross section was already given in table 2.1:

\[ \sigma_{t\text{-channel}} (8 \text{ TeV, pred.}) = 87.2^{+2.8}_{-1.0} \, \text{(scale var.)}^{+2.0}_{-2.2} \, \text{(PDF)} \, \text{pb} . \]  

(7.15)

The measurement presented in this chapter is compatible with this prediction, within the uncertainties. A significant contribution to the systematic uncertainty comes from the extrapolation of the selected phase space to the full phase space. In other words, while the experimental effects are reasonably under control, the signal MC generator uncertainty dominates the overall systematic uncertainty. The next chapter will focus on a fiducial cross section measurement to reduce this signal generator uncertainty.
Chapter 7
8 - Fiducial cross section measurement

In the previous chapter, the total cross section for the single top t-channel production process has been extracted. To be more precise, it is not the total cross section which is actually measured. In fact, only events in the detector acceptance, inside the fiducial volume, are measured from which a total cross section is obtained by extrapolation to the full phase space using the theoretical prediction. Especially in the forward region, outside the fiducial area, the non-perturbative QCD effects are large. MC generators have different approaches to model these effects, which leads to different results. For instance, in the previous chapter it was shown that the difference between AcerMC and Powheg amounts to about ±2.3% in the fiducial phase space, whereas the extrapolation to the full phase space has an uncertainty of about ±7.5%.

In this chapter we extract the cross section in a limited kinematic region corresponding to a fiducial volume. The 'fiducial volume' as nomenclature can be somewhat misleading, because it does not simply refer to a geometrical volume. Instead, it is defined by the selection requirements introduced in the previous chapter that comprise kinematic cuts as well. Therefore, the fiducial volume is more like a kinematic domain than a volume in the geometrical sense.

To measure the so-called fiducial cross section, the fiducial volume (or kinematic domain) has to be defined at the truth-level of the MC events. Hence, in this chapter we introduce requirements that act on the truth information of the MC generators. To avoid confusion, we will indicate the previous requirements (chapter 7) for events selected from ATLAS data (or full MC simulation) as detector-level cuts. Ideally, the requirements at truth-level and detector-level select the same events, but due to finite detector resolution event migration occurs. Hence, (relatively small) MC corrections remain necessarily, but these are expected to be small as we show in this chapter.
A small but notable difference with the measurement in chapter 7 is the treatment of tau events that pass the selection. The tau events are always included in the MC simulation, but in the previous chapter the signal cross sections for prompt $W \rightarrow e$ and $W \rightarrow \mu$ decays are extracted. The MC corrections that remove the tau decays, of about 2% are absorbed in the efficiency (equation 7.2). In this chapter we extract the cross section for events that lead to either a selected electron or muon, which by construction includes the $W \rightarrow \tau \rightarrow e, \mu$ events.

The advantage of the fiducial cross section is not only the smaller phase space corrections as compared to the total cross section. Also, the requirements at truth-level are specifically designed to be MC generator independent, and can thus be applied to (new) MC generators. This allows to compare the predictions of any MC generator in the specific fiducial volume, where the measurement is performed. This improves and facilitates model builders (of new physics) to study the actual impact on the measurement.

In section 8.1 we introduce the fiducial cross section, and in section 8.2 the fiducial cuts. The results of these cuts are shown in section 8.3, with the event yields and associated efficiencies being the subject of section 8.4. Finally, we conclude the chapter in section 8.5 with the results of the fiducial single top cross section measurement, and the discussion of the results and an outlook in section 8.6.

### 8.1 - Fiducial cross sections

The fiducial cross section $\sigma_{\text{fid}}$ is related to the full cross section $\sigma$ that was measured in the previous chapter through the following equation:

$$\sigma_{\text{fid}} = \epsilon_{\text{fid}} \cdot \sigma, \quad (8.1)$$

where $\epsilon_{\text{fid}}$ is the acceptance of the measured fiducial cross section. It is defined as:

$$\epsilon_{\text{fid}} \equiv \frac{N_{\text{tru}}}{N_{\text{total}}}, \quad (8.2)$$

where $N_{\text{tru}}$ represents the number of Monte Carlo signal events that pass the fiducial selection cuts defined using stable particles before detector simulation at truth-level, which we introduce later in more detail. $N_{\text{total}}$ is the total size of the generated MC signal sample.

In the following we continue with the subscript 'tru' as we refer the truth-level selection. We use this notation only for MC events. For events measured by the ATLAS detector, by construction at detector-level, we use the subscript 'data'.
Preferably there should be a large overlap between the events passing the detector-level cuts as defined in the previous chapter and the truth-level cuts. We denote the previous chapter's selection with 'det' to indicate that these cuts apply to ATLAS data or MC event on detector-level. To quantify the overlap between the detector-level and truth-level selection, \(N_{\text{det} & \text{tru}}\) is defined, which is the number of events that pass both the detector-level cuts and the truth-level selection. The fraction of events selected at detector-level that also pass the truth-level cuts, the purity of the selected sample, is now given by the ratio:

\[
P_{\text{det}} \equiv \frac{N_{\text{det} & \text{tru}}}{N_{\text{det}}} ,
\]

where \(N_{\text{det}}\) denotes the total number of MC events selected at detector-level, irrespective whether these events pass the truth-level cuts.

The efficiency to select an event at detector-level that also passes the truth-level selection is given by:

\[
P_{\text{tru}} \equiv \frac{N_{\text{det} & \text{tru}}}{N_{\text{tru}}} .
\]

It is convenient to define a ratio called \(R_f\), as follows:

\[
R_f \equiv \frac{P_{\text{det}}}{P_{\text{tru}}} = \frac{N_{\text{tru}}}{N_{\text{det}}} = \frac{\epsilon_{\text{fid}}}{\epsilon} .
\]

Using \(R_f\), the fiducial cross section can be rewritten as:

\[
\sigma_{\text{fid}} = R_f \frac{N_{\text{sel, data}}}{L} .
\]

where \(N_{\text{sel, data}}\) are the selected data events after background subtraction. For completeness we remark that the MC prediction of the fiducial cross section, \(\sigma_{\text{fid}, \text{MC}}\), is obtained by replacing \(N_{\text{sel, data}}\) with \(N_{\text{sel}}\) in equation 8.6.

This equation gives the fiducial cross section that we will measure in this chapter. The \(\sigma_{\text{fid}}\) depends on events that are either selected on detector-level or generated inside the fiducial volume. The total cross section depends in addition on events that are generated outside the fiducial volume, but not selected at detector-level. This is the key difference between the two approaches, and will lead to a rather different systematic uncertainties.
8.2 - Truth-level cuts

To define truth-level cuts for the fiducial cross section measurement, the event selection is applied on truth particle level\textsuperscript{[279]} using all final state particles before detector simulation in the Monte Carlo event.\textsuperscript{[280]} It is important that a similar selection can be applied to reconstructed (measured) objects. Ideally, there would be a one-to-one correspondence, but migration of events due to detector effects will take place. This is quantified by $\epsilon_{\text{corr,sel}}$, defined in equation 8.3 above. In this section the particle level selection is discussed.

**Lepton dressing**

The bare electron (or muon) from the decaying $W$-boson is 'dressed' to add the soft photon radiation back to the original lepton, which is modeled with large systematic uncertainty by MC generators.\textsuperscript{[280]} The amount of energy radiates this way can be several GeV, which is quite significant. In order to get a good measure on the energy of the lepton as it was produced, these photons need to be added back to the lepton, to get to the truth lepton energy and direction.

To merge the photons back into the leptons, a technique called lepton dressing is used. A cone with a $\Delta R$ of 0.1 is constructed around each lepton. All the final state photons that lie within that cone are added to the four-momentum of the lepton. This dressing procedure consumes the matched photons.
In figure 8.1 the effect of lepton dressing on the $\Delta R$ distribution between the lepton and photons is illustrated. It is clear that the peak at small $\Delta R$ originates from radiation which is added back to the lepton by the dressing procedure.

**Lepton selection**

After the leptons have been reconstructed, their selection cuts are applied. For electrons it is demanded that $E_T > 25$ GeV, and $|\eta| < 2.5$. Similar cuts are applied to the muons: $p_T > 25$ GeV and $|\eta| < 2.5$.

Additionally, the same cut as on detector-level on the lepton's $p_T$ (equation 6.17) is made with respect to the leading jet:

$$p_T(l) > 40 \text{ GeV} \left(1 - \frac{\pi - |\Delta \phi(j_1, l)|}{\pi - 1}\right).$$ (8.7)

**Jets at particle level**

The jet reconstruction considers all the remaining final state particles; this thus excludes the selected leptons (before the cut on the lepton $p_T$) and photons used in lepton dressing. The algorithm for AntiKt 0.4 jets is used, see section 6.3. Selection criteria are then applied to the reconstructed jets, so that only jets with $p_T > 25$ GeV and $|\eta| < 4.5$ are selected.

**b-tagging at particle level**

$b$-tagging at particle level is done differently from $b$-tagging on detector-level events from data or simulation. A list of all the $B$-hadrons among all the truth particles with $p_T > 5$ GeV are added (with their energy set to zero) to the input particles of the jet reconstruction algorithm. This does not affect the final reconstructed jets, but it will add the $B$-hadrons to the jet when appropriate. If a $B$-hadron is added to a jet in this way, the jet is tagged as being a $b$-jet.

**Overlap removal**

Overlap removal at particle level works differently than that at detector-level. Instead of rejecting reconstructed objects, the entire event is rejected.\cite{280}

The procedure removes events when electrons (muons) and jets are close together, defined by $\Delta R < 0.4$ (0.1). There is no need here to perform a more complicated overlap removal because several types of detector misidentifications cannot occur at particle level. This overlap removal results in about 2% of the events being rejected, which impacts the fiducial volume of the cross section.

**Reconstruction of $E_T$, $m_T^W$, and $m_\ell$**

The $E_T$ is calculated directly from the neutrinos' momenta. This will include the neutrino from the $W$ decay, and additional neutrinos that are produced in the parton showering and tau-decays.
The neutrino's momentum along the $z$-axis is reconstructed by the same procedure as on detector-level (section 7.2), and the $m_W$ is calculated by reconstructing the $W$-boson in the same way as on detector-level (section 6.5.5). Note that the truth $m_W$ and $p_{z,\nu}$ are not directly used to ensure that differences arising due to their reconstruction efficiencies are not made part of the fiducial cross section.

Finally, the top-quark mass is also reconstructed in the same manner as on detector-level.

| Jets: |
|---|---|
| $|\eta_{\text{jet}}| < 4.5$ |
| $p_{T,\text{jet}} \geq 30 \text{ GeV}$ |
| $p_{T,\text{jet}} \geq 35 \text{ GeV}$ (if $2.5 < |\eta_{\text{jet}}| < 3.75$) |

| Electrons: |
|---|---|
| $|\eta_{\text{el}}| < 2.5$ |
| $E_{T,\text{el}} > 25 \text{ GeV}$ (no crack region) |
| $\Delta R > 0.4$ with jets |

| Muons: |
|---|---|
| $|\eta_{\text{mu}}| < 2.5$ |
| $p_{T,\text{mu}} > 25 \text{ GeV}$ |
| $\Delta R > 0.1$ with jets |

| Others: |
|---|---|
| $E_T > 30 \text{ GeV}$ |
| $m_W > 50 \text{ GeV}$ |

Table 8.1: Summary of the object selection cuts on particle level.

The selection cuts on the leptons and jets are similar to those used on detector-level, in order to match their kinematics. These cuts are summarized in table 8.1. This table does not include the final event selection cuts (such as the cut on $H_T$, etc.), which are defined at particle level identical to their detector-level counterparts.

### 8.3 - Sample comparison

In this section, we compare the signal MC samples at particle level, and we present the final event yield. First we discuss various distributions related to the truth-level selection,
highlighting some of the larger differences between the samples. In this section only the electron results are presented; at this stage there is practically no visible difference between the electron and muon distributions and therefore the latter is not shown.

As in the previous chapter, there are two signal samples from which $N_{sel,\text{data}}$ is extracted:

- Powheg + Pythia
- AcerMC + Pythia

For additional studies another sample with lower statistics is used as well:

- aMC@NLO + Herwig + Jimmy

A fourth sample is also shown:

- Powheg + Herwig + Jimmy

For this last sample, only events at truth-level are available. This sample is nevertheless shown for completeness in the figures of this section. Figure 8.2a shows the legend, used in the corresponding plots.

In figure 8.2b, the jet multiplicity after the fiducial requirements for jets (but before the 2-jet requirement) is shown. Here, a clear difference of almost 10% between LO (AcerMC) and NLO predictions can be seen. This difference suggests that for proper modeling of the jet
Figure 8.3: The $E_T$ (a) and the $m_W$ (b) obtained at particle level of the pre-selection fiducial selection events in the electron stream, for the different MC samples.

Figure 8.4: The $H_T$ (a), and the $\eta$ of the light jet (b) obtained at particle level of the pre-selection fiducial selection events in the electron stream, right before their respective cuts, for the different MC samples.

multiplicity, NLO corrections are important to take into account. Thus, for future analyses, the signal sample should be generated with an NLO MC generator.

Figure 8.3a shows, for 2-jet events, that the prediction for the $E_T$ distribution is comparable for all samples. This is also the case for the $m_W$ distribution in figure 8.3b.
After the pre-selection, the same cuts as in the previous chapter are applied in order to select the signal region. Figure 8.4a shows the $H_T$ distribution before the $H_T$-cut (equation 7.3). In figure 8.4b, the light jet $\eta$ distribution is shown before the cut (equation 7.4). This distribution exhibits differences between the generators around 10% in the relevant regions.

Next, the $\Delta\eta$ between the light and $b$-jet is required to be larger than 1.0 (equation 7.5), as shown in figure 8.5a. In the selected regions the difference between the generators is about 10%.

This results in the reconstructed top mass distribution shown in figure 8.5b. It should be noted that the peak in this particle level distribution is substantially sharper than the peak reconstructed in the detector-level analysis, as shown in figure 7.6. If this quantity would be used in the particle level cuts as was done in previous analyses,\textsuperscript{278} a large difference between the particle level selection and that on simulated events can be expected. (This cut has been replaced with the fitting procedure described in section 7.3.)

**8.4 - Event yields and efficiencies**

In this section, the event yields and the efficiencies defined in section 8.1 are given after the fiducial cuts. The total number of events before any requirements ($N_{\text{total}}$) is $5.8\times10^5$ for all samples.
The event yields are given in table 8.2 (table 8.3) for the electron (muon) stream. In the tables $\sigma_{\text{fid,MC}}$ is the fiducial cross section as from the MC prediction. This comparison shows that the AcerMC, which is the only LO generator in this comparison, predicts significantly less events (~15%) in the fiducial region ($N_{\text{true}}$). This is the main reason for the large systematic uncertainty discussed in the previous chapter. For all generators, the $N_{\text{det}}$ yield deviates from $N_{\text{det}}$ significantly, indicating that only about 65% of the events that pass the non-fiducial selection also pass the fiducial cuts, as expressed by $\epsilon_{\text{corr,sel}}$. The values for $\epsilon_{\text{corr,sel}}$ appear quite comparable between the generators.

The efficiencies are calculated through the relations discussed in section 8.1. The difference in efficiencies $\epsilon_{\text{corr,sel}}$ and $\epsilon_{\text{corr,fid}}$ between the samples is smaller than 10%, consistent with the variations of similar order that are also observed in the comparison of distributions presented in section 8.3. The quantity $R_f$ shows similar variation and is used in the fiducial cross section calculation, which we will discuss next.

<table>
<thead>
<tr>
<th>Sample</th>
<th>$N_{\text{true}}$</th>
<th>$N_{\text{det}}$</th>
<th>$N_{\text{det &amp; true}}$</th>
<th>$\epsilon_{\text{fid}}$</th>
<th>$\epsilon_{\text{det}}$</th>
<th>$\epsilon_{\text{true}}$</th>
<th>$R_f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Powheg</td>
<td>13734</td>
<td>2910</td>
<td>1920</td>
<td>0.0239</td>
<td>0.660</td>
<td>0.140</td>
<td>4.72</td>
</tr>
<tr>
<td>aMC@NLO</td>
<td>14600</td>
<td>2917</td>
<td>1889</td>
<td>0.0253</td>
<td>0.648</td>
<td>0.129</td>
<td>5.00</td>
</tr>
<tr>
<td>AcerMC</td>
<td>11966</td>
<td>2461</td>
<td>1610</td>
<td>0.0208</td>
<td>0.654</td>
<td>0.135</td>
<td>4.86</td>
</tr>
</tbody>
</table>

Table 8.2: Fiducial event yields and efficiencies of the electron stream for the various samples.

<table>
<thead>
<tr>
<th>Sample</th>
<th>$N_{\text{true}}$</th>
<th>$N_{\text{det}}$</th>
<th>$N_{\text{det &amp; true}}$</th>
<th>$\epsilon_{\text{fid}}$</th>
<th>$\epsilon_{\text{det}}$</th>
<th>$\epsilon_{\text{true}}$</th>
<th>$R_f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Powheg</td>
<td>14360</td>
<td>3837</td>
<td>2523</td>
<td>0.0250</td>
<td>0.658</td>
<td>0.176</td>
<td>3.74</td>
</tr>
<tr>
<td>aMC@NLO</td>
<td>14712</td>
<td>3590</td>
<td>2424</td>
<td>0.0255</td>
<td>0.675</td>
<td>0.165</td>
<td>4.10</td>
</tr>
<tr>
<td>AcerMC</td>
<td>12500</td>
<td>3283</td>
<td>2146</td>
<td>0.0217</td>
<td>0.654</td>
<td>0.172</td>
<td>3.81</td>
</tr>
</tbody>
</table>

Table 8.3: Fiducial event yields and efficiencies of the muon stream for the various samples.

Chapter 8
8.5 - Fiducial cross section measurement

<table>
<thead>
<tr>
<th>Sample</th>
<th>$\sigma_{\text{fid},\text{MC}}$ (Electron)</th>
<th>$\sigma_{\text{fid},\text{MC}}$ (Muon)</th>
<th>$\sigma_{\text{fid}}$ (Electron)</th>
<th>$\sigma_{\text{fid}}$ (Muon)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Powheg</td>
<td>0.70 pb</td>
<td>0.73 pb</td>
<td>0.614 pb</td>
<td>0.706 pb</td>
</tr>
<tr>
<td>aMC@NLO</td>
<td>0.74 pb</td>
<td>0.75 pb</td>
<td>0.651 pb</td>
<td>0.773 pb</td>
</tr>
<tr>
<td>AcerMC</td>
<td>0.61 pb</td>
<td>0.63 pb</td>
<td>0.633 pb</td>
<td>0.719 pb</td>
</tr>
</tbody>
</table>

Table 8.4: Fiducial cross sections for the various MC samples and for data using corrections based on the corresponding MC simulation. These results are used to estimate the systematic uncertainties, as described in the text.

We take the $N_{\text{sel, data}}$ of 2639 (3827) for electrons (muons) as determined in chapter 7 and calculate the fiducial cross section using equation 8.6 and the three predictions of $R_f$ in table 8.2 (table 8.3). The differences are of order 5%, which is a modest variation.

Table 8.4 shows the various MC predictions for the fiducial cross section, $\sigma_{\text{fid,MC}}$, and the measured $\sigma_{\text{fid}}$ for the electron and muon stream. The central value and generator uncertainties of the fiducial cross section are obtained by taking the largest variation in $\sigma_{\text{fid}}$ between the various samples and symmetrizing it, which results in ±3.0% (±4.5%) for the electron (muon) stream.

This systematic uncertainty is combined with the experimental and background uncertainties as have been discussed in chapter 7, resulting in a total uncertainty of ±10.2% (±9.9%) for the electron (muon) stream.

The results for the fiducial cross section are:

$$\sigma_{t\text{-channel, fid}} (8 \text{ TeV, el}) = 0.614 \pm 0.022 \text{ (stat.)}^{+0.059}_{-0.059} \text{ (syst.)} \text{ pb}, \quad (8.8)$$

$$\sigma_{t\text{-channel, fid}} (8 \text{ TeV, mu}) = 0.706 \pm 0.026 \text{ (stat.)}^{+0.066}_{-0.066} \text{ (syst.)} \text{ pb}. \quad (8.9)$$

These cross sections correspond to all t-channel events that lead to an isolated electron or muon event and thus include small contributions from $W\rightarrow\tau\rightarrow e (\mu)$ events. For the Powheg generator, these contributions are predicted as 1.9% (1.8%) for the fraction of $W\rightarrow\tau$ events in electron (muon) events. The inclusive fiducial cross section is obtained from the combined results for the electron and muon channel. After a correction for the small fraction of tau events, followed by the trivial correction for the branching fractions of the $W$, this results in:

$$\sigma_{t\text{-channel, fid}} (8 \text{ TeV}) = 4.07 \pm 0.09 \text{ (stat.)}^{+0.37}_{-0.34} \text{ (syst.)} \text{ pb}. \quad (8.10)$$

153
The correlation between systematic sources in the electron and muon channel are taken into account in this combination. The results for electrons and muons are consistent and also in agreement with the predictions for $\sigma_{\text{fd,MC}}$ in table 8.4.

The uncertainties from the signal MC generator of typical 4% are significantly smaller than the corresponding variation of ±7.5% observed in the total t-channel cross section measurement; see table 7.2 of the previous chapter. This demonstrates that a substantial fraction of the differences between MC generators originates from predictions outside of the fiducial volume. Using a fiducial volume to determine the systematic uncertainty due to the choice of MC generator reduces this uncertainty.

8.6 - Discussion

The measured fiducial cross section is compatible with the Monte Carlo predictions. No indication for new physics is found. The 2-sigma uncertainty of the fiducial cross section is about 0.13 pb for electrons and 0.14 pb for muons, which is an estimate for the upper limit for the new physics cross section at 95% CL in these channels. Hence, future MC generators for new physics can compare their predictions with this limit (under the assumption that its selection efficiency $\epsilon_{\text{corr,fd}}$ is similar). Such comparison is not directly possible with the extrapolated (full phase space) cross section discussed in the previous chapter, because the new physics may be abundantly present outside the fiducial region, outside the sensitivity of the actual measurement.

The LHC will start in 2015 with an increased proton-proton collision energy of 13 TeV which will increase the top production, as suggested by figure 2.18. Together with more integrated luminosity (around 100 fb$^{-1}$ is expected in Run 2, a factor 5 increase) this will provide much more data for top-quark measurements than ever before. Additionally, with an improved understanding of the ATLAS detector, and the upgrades of the long shutdown (including adding another detection layer to the inner detector, the insertable B-layer (IBL)), this will all lead to a further reduction of the statistical and systematic uncertainties. As analyses will employ harder and more efficient background-rejecting cuts as well, the fiducial approach combined with the fitting procedure as presented in this thesis can be used as a basis for measuring the single top t-channel cross section even more accurately and precisely.

Currently, both the $s$-[281] and Wt-channel[282] are also measured to be consistent with the Standard Model prediction. However, if these are re-measured under the same improved circumstances, new physics which differently affects the several top production channels may yet reveal clues of its fundamental features.
Appendix A - W+jets sample definition

Table A.1 gives the exact definition of the $W$+jets samples introduced in section 3.3.2 by listing the values of the Alpgen input parameters. The 'ihvy' parameter mentioned gives the flavor of the quark that is treated as massive.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Alpgen process</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wq</td>
<td>$W + \text{jets}$</td>
<td>$\text{ptjmin} 15.0$ $\text{ptlmin} 0.0$ $\text{metmin} 0.0$ $\text{etajmax} 6.0$ $\text{etalmx} 10.0$ $\text{drjmin} 0.7$ $\text{drlmin} 0.0$</td>
</tr>
<tr>
<td>Wbb</td>
<td>$W + Q\overline{Q} + \text{jets}$</td>
<td>$\text{ihvy} 5$ $\text{ptjmin} 15.0$ $\text{ptbmin} 0.0$ $\text{ptlmin} 0.0$ $\text{metmin} 0.0$ $\text{etajmax} 6.0$ $\text{etabmax} 6.0$ $\text{etalmx} 10.0$ $\text{drjmin} 0.7$ $\text{drcmin} 0.0$ $\text{drlmin} 0.0$</td>
</tr>
<tr>
<td>Wc</td>
<td>$W + c + \text{jets}$</td>
<td>$\text{ptmin} 15.0$ $\text{ptcmin} 10.0$ $\text{ptlmin} 0.0$ $\text{metmin} 0.0$ $\text{etajmax} 6.0$ $\text{etacmax} 6.0$ $\text{etalmx} 10.0$ $\text{drjmin} 0.7$ $\text{drcmin} 0.0$ $\text{drlmin} 0.0$</td>
</tr>
<tr>
<td>Wcc</td>
<td>$W + Q\overline{Q} + \text{jets}$</td>
<td>$\text{ihvy} 4$ $\text{ptjmin} 15.0$ $\text{ptcmin} 0.0$ $\text{ptlmin} 0.0$ $\text{metmin} 0.0$ $\text{etajmax} 6.0$ $\text{etacmax} 6.0$ $\text{etalmx} 10.0$ $\text{drjmin} 0.7$ $\text{drcmin} 0.0$ $\text{drlmin} 0.0$</td>
</tr>
</tbody>
</table>

Table A.1: A table listing all Alpgen samples that contribute to the $W$+jets processes, and their phase space cuts.
Appendix A
### Appendix B - Fit results

In this appendix the full result tables for section 7.4 are given. Table B.1 is for the electron stream, and table B.2 for the muon stream.

<table>
<thead>
<tr>
<th>Variation</th>
<th>W</th>
<th>$t\bar{t}$</th>
<th>Single top $t$-channel</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Events</td>
<td>Fitfactor</td>
<td>Events</td>
<td>Fitfactor</td>
</tr>
<tr>
<td>Nominal</td>
<td>1590±114</td>
<td>0.892±0.064</td>
<td>1920±105</td>
<td>1.042±0.057</td>
</tr>
<tr>
<td><strong>Jet variations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>jeff</td>
<td>1573±114</td>
<td>0.884±0.064</td>
<td>1929±105</td>
<td>1.046±0.057</td>
</tr>
<tr>
<td>jer</td>
<td>1512±115</td>
<td>0.816±0.062</td>
<td>1953±105</td>
<td>1.046±0.056</td>
</tr>
<tr>
<td>jes_down</td>
<td>1712±117</td>
<td>1.057±0.073</td>
<td>2013±107</td>
<td>1.053±0.056</td>
</tr>
<tr>
<td>jes_up</td>
<td>1584±113</td>
<td>0.791±0.056</td>
<td>1824±101</td>
<td>1.025±0.057</td>
</tr>
<tr>
<td>jvf_down</td>
<td>1602±114</td>
<td>0.908±0.065</td>
<td>1889±103</td>
<td>1.042±0.057</td>
</tr>
<tr>
<td>jvf_up</td>
<td>1588±114</td>
<td>0.878±0.063</td>
<td>1937±105</td>
<td>1.043±0.057</td>
</tr>
<tr>
<td><strong>b-tagging variations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>btag_btag_down</td>
<td>1628±113</td>
<td>0.856±0.059</td>
<td>1852±101</td>
<td>1.042±0.057</td>
</tr>
<tr>
<td>btag_btag_up</td>
<td>1541±115</td>
<td>0.926±0.069</td>
<td>1995±108</td>
<td>1.045±0.057</td>
</tr>
<tr>
<td>btag_ctau_tag_down</td>
<td>1599±114</td>
<td>0.902±0.065</td>
<td>1920±104</td>
<td>1.043±0.057</td>
</tr>
<tr>
<td>btag_ctau_tag_up</td>
<td>1574±114</td>
<td>0.883±0.064</td>
<td>1926±105</td>
<td>1.044±0.057</td>
</tr>
<tr>
<td>btag_mistag_down</td>
<td>1608±114</td>
<td>0.869±0.062</td>
<td>1912±105</td>
<td>1.038±0.057</td>
</tr>
<tr>
<td>btag_mistag_up</td>
<td>1548±113</td>
<td>0.914±0.067</td>
<td>1944±104</td>
<td>1.054±0.057</td>
</tr>
<tr>
<td><strong>Electron variations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eer_down</td>
<td>1571±112</td>
<td>0.884±0.063</td>
<td>1955±103</td>
<td>1.058±0.056</td>
</tr>
<tr>
<td>eer_up</td>
<td>1555±114</td>
<td>0.875±0.065</td>
<td>1958±104</td>
<td>1.060±0.056</td>
</tr>
<tr>
<td>ees_down</td>
<td>1612±113</td>
<td>0.917±0.065</td>
<td>1921±105</td>
<td>1.027±0.056</td>
</tr>
<tr>
<td>ees_up</td>
<td>1583±111</td>
<td>0.872±0.061</td>
<td>1901±102</td>
<td>1.047±0.056</td>
</tr>
<tr>
<td><strong>Lepton variations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sf_ld_down</td>
<td>1623±113</td>
<td>0.875±0.061</td>
<td>1876±102</td>
<td>1.042±0.057</td>
</tr>
<tr>
<td>sf_ld_up</td>
<td>1554±115</td>
<td>0.908±0.067</td>
<td>1967±107</td>
<td>1.043±0.057</td>
</tr>
<tr>
<td>sf_reco_down</td>
<td>1595±114</td>
<td>0.891±0.064</td>
<td>1915±104</td>
<td>1.042±0.057</td>
</tr>
<tr>
<td>sf_reco_up</td>
<td>1584±114</td>
<td>0.892±0.064</td>
<td>1926±105</td>
<td>1.042±0.057</td>
</tr>
<tr>
<td>trigger_sf_down</td>
<td>1601±114</td>
<td>0.888±0.063</td>
<td>1908±104</td>
<td>1.042±0.057</td>
</tr>
<tr>
<td>trigger_sf_up</td>
<td>1579±114</td>
<td>0.896±0.065</td>
<td>1933±105</td>
<td>1.042±0.057</td>
</tr>
</tbody>
</table>
### Appendix B

#### Table B.1: The results of each variation for the fitting procedure, in the electron stream. The number of events contain the fitfactor.
<table>
<thead>
<tr>
<th>Variation</th>
<th>Events</th>
<th>Fitfactor</th>
<th>$\sigma$</th>
<th>Fitfactor</th>
<th>Events</th>
<th>Fitfactor</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>2725±135</td>
<td>0.811±0.040</td>
<td>1.100±0.056</td>
<td>3784±126</td>
<td>0.986±0.033</td>
<td>0.0067</td>
<td></td>
</tr>
<tr>
<td><strong>Jet variations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>jeff</td>
<td>2746±136</td>
<td>0.818±0.041</td>
<td>1.093±0.056</td>
<td>3772±126</td>
<td>0.984±0.033</td>
<td>0.0067</td>
<td></td>
</tr>
<tr>
<td>jer</td>
<td>2696±141</td>
<td>0.763±0.040</td>
<td>1.062±0.056</td>
<td>3879±128</td>
<td>1.021±0.034</td>
<td>0.0066</td>
<td></td>
</tr>
<tr>
<td>jes_down</td>
<td>2744±137</td>
<td>0.907±0.045</td>
<td>1.133±0.055</td>
<td>3572±128</td>
<td>0.992±0.036</td>
<td>0.0063</td>
<td></td>
</tr>
<tr>
<td>jes_up</td>
<td>2656±135</td>
<td>0.716±0.037</td>
<td>1.094±0.056</td>
<td>3953±125</td>
<td>0.976±0.031</td>
<td>0.0070</td>
<td></td>
</tr>
<tr>
<td>jvf_down</td>
<td>2770±135</td>
<td>0.825±0.040</td>
<td>1.091±0.056</td>
<td>3811±126</td>
<td>1.002±0.033</td>
<td>0.0066</td>
<td></td>
</tr>
<tr>
<td>jvf_up</td>
<td>2712±136</td>
<td>0.806±0.041</td>
<td>1.091±0.056</td>
<td>3752±127</td>
<td>0.964±0.033</td>
<td>0.0068</td>
<td></td>
</tr>
<tr>
<td><strong>b-tagging variations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>btag_btag_down</td>
<td>2750±135</td>
<td>0.772±0.037</td>
<td>1.107±0.056</td>
<td>3856±125</td>
<td>1.042±0.034</td>
<td>0.0064</td>
<td></td>
</tr>
<tr>
<td>btag_btag_up</td>
<td>2696±137</td>
<td>0.853±0.044</td>
<td>1.095±0.056</td>
<td>3712±127</td>
<td>0.934±0.032</td>
<td>0.0069</td>
<td></td>
</tr>
<tr>
<td>btag_ctautag_down</td>
<td>2767±136</td>
<td>0.821±0.040</td>
<td>1.096±0.056</td>
<td>3769±126</td>
<td>0.983±0.033</td>
<td>0.0067</td>
<td></td>
</tr>
<tr>
<td>btag_ctautag_up</td>
<td>2684±135</td>
<td>0.805±0.041</td>
<td>1.105±0.056</td>
<td>3796±126</td>
<td>0.989±0.033</td>
<td>0.0067</td>
<td></td>
</tr>
<tr>
<td>btag_mistag_down</td>
<td>2728±135</td>
<td>0.783±0.039</td>
<td>1.099±0.056</td>
<td>3793±126</td>
<td>0.989±0.033</td>
<td>0.0067</td>
<td></td>
</tr>
<tr>
<td>btag_mistag_up</td>
<td>2705±135</td>
<td>0.850±0.043</td>
<td>1.106±0.056</td>
<td>3780±126</td>
<td>0.985±0.033</td>
<td>0.0067</td>
<td></td>
</tr>
<tr>
<td><strong>Muon variations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>muid_down</td>
<td>2711±135</td>
<td>0.809±0.040</td>
<td>1.096±0.056</td>
<td>3797±126</td>
<td>0.990±0.033</td>
<td>0.0067</td>
<td></td>
</tr>
<tr>
<td>muid_up</td>
<td>2690±135</td>
<td>0.797±0.040</td>
<td>1.110±0.056</td>
<td>3789±126</td>
<td>0.989±0.033</td>
<td>0.0067</td>
<td></td>
</tr>
<tr>
<td>mums_down</td>
<td>2727±134</td>
<td>0.784±0.039</td>
<td>1.085±0.056</td>
<td>3821±125</td>
<td>1.132±0.037</td>
<td>0.0067</td>
<td></td>
</tr>
<tr>
<td>mums_up</td>
<td>2684±134</td>
<td>0.766±0.038</td>
<td>1.106±0.056</td>
<td>3795±126</td>
<td>1.122±0.037</td>
<td>0.0067</td>
<td></td>
</tr>
<tr>
<td>musc_down</td>
<td>2682±134</td>
<td>0.766±0.038</td>
<td>1.116±0.056</td>
<td>3783±126</td>
<td>1.120±0.037</td>
<td>0.0067</td>
<td></td>
</tr>
<tr>
<td>musc_up</td>
<td>2611±133</td>
<td>0.747±0.038</td>
<td>1.130±0.056</td>
<td>3801±125</td>
<td>1.124±0.037</td>
<td>0.0067</td>
<td></td>
</tr>
<tr>
<td><strong>Lepton variations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sf ld_down</td>
<td>2730±135</td>
<td>0.807±0.040</td>
<td>1.101±0.056</td>
<td>3793±126</td>
<td>0.994±0.033</td>
<td>0.0066</td>
<td></td>
</tr>
<tr>
<td>sf ld_up</td>
<td>2719±135</td>
<td>0.816±0.041</td>
<td>1.099±0.056</td>
<td>3775±126</td>
<td>0.979±0.033</td>
<td>0.0067</td>
<td></td>
</tr>
<tr>
<td>sf reco_down</td>
<td>2730±135</td>
<td>0.808±0.040</td>
<td>1.101±0.056</td>
<td>3790±126</td>
<td>0.991±0.033</td>
<td>0.0066</td>
<td></td>
</tr>
<tr>
<td>sf reco_up</td>
<td>2720±135</td>
<td>0.814±0.041</td>
<td>1.100±0.056</td>
<td>3778±126</td>
<td>0.981±0.033</td>
<td>0.0067</td>
<td></td>
</tr>
<tr>
<td>trigger sf down</td>
<td>2744±134</td>
<td>0.795±0.039</td>
<td>1.103±0.056</td>
<td>3817±126</td>
<td>1.013±0.033</td>
<td>0.0065</td>
<td></td>
</tr>
<tr>
<td>trigger sf_up</td>
<td>2705±136</td>
<td>0.828±0.042</td>
<td>1.098±0.056</td>
<td>3752±127</td>
<td>0.960±0.033</td>
<td>0.0068</td>
<td></td>
</tr>
<tr>
<td><strong>$\mathcal{E}_{T}$ variations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>res soft_down</td>
<td>2667±134</td>
<td>0.754±0.038</td>
<td>1.096±0.056</td>
<td>3821±127</td>
<td>1.119±0.037</td>
<td>0.0068</td>
<td></td>
</tr>
<tr>
<td>res soft_up</td>
<td>2688±136</td>
<td>0.768±0.039</td>
<td>1.093±0.056</td>
<td>3795±127</td>
<td>1.112±0.037</td>
<td>0.0068</td>
<td></td>
</tr>
<tr>
<td>sc soft_down</td>
<td>2847±135</td>
<td>0.823±0.039</td>
<td>1.060±0.056</td>
<td>3792±125</td>
<td>1.136±0.038</td>
<td>0.0066</td>
<td></td>
</tr>
<tr>
<td>sc soft_up</td>
<td>2701±135</td>
<td>0.755±0.038</td>
<td>1.103±0.056</td>
<td>3766±127</td>
<td>1.100±0.037</td>
<td>0.0068</td>
<td></td>
</tr>
<tr>
<td><strong>Monte Carlo variations - background estimation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SherpaW</td>
<td>2622</td>
<td>0.752±0.039</td>
<td>1.134±0.056</td>
<td>3795</td>
<td>0.989±0.033</td>
<td>0.0067</td>
<td></td>
</tr>
<tr>
<td>tbarMecAtNlo</td>
<td>2733±135</td>
<td>0.819±0.041</td>
<td>1.098±0.056</td>
<td>3781±126</td>
<td>0.986±0.033</td>
<td>0.0067</td>
<td></td>
</tr>
<tr>
<td>xsecvar_tbarPoPy_down</td>
<td>2830±130</td>
<td>0.769±0.036</td>
<td>1.108±0.056</td>
<td>3944±123</td>
<td>1.028±0.032</td>
<td>0.0067</td>
<td></td>
</tr>
<tr>
<td>xsecvar_tbarPoPy_up</td>
<td>2625±140</td>
<td>0.857±0.046</td>
<td>1.093±0.056</td>
<td>3638±129</td>
<td>0.948±0.034</td>
<td>0.0067</td>
<td></td>
</tr>
<tr>
<td><strong>Monte Carlo variations - signal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AcerMC</td>
<td>2586±135</td>
<td>0.740±0.039</td>
<td>1.082±0.056</td>
<td>3989±132</td>
<td>1.215±0.040</td>
<td>0.0057</td>
<td></td>
</tr>
<tr>
<td><strong>$H^\pm$-jets background variations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>knfacvar_Kbb_down</td>
<td>2728±135</td>
<td>0.822±0.041</td>
<td>1.100±0.056</td>
<td>3782±126</td>
<td>0.986±0.033</td>
<td>0.0067</td>
<td></td>
</tr>
<tr>
<td>knfacvar_Kbb_up</td>
<td>2721±135</td>
<td>0.801±0.040</td>
<td>1.101±0.056</td>
<td>3787±126</td>
<td>0.987±0.033</td>
<td>0.0067</td>
<td></td>
</tr>
</tbody>
</table>
### Appendix B

| kfacvar_Kc_down | 2723±135 | 0.813±0.040 | 2626±133 | 1.100±0.056 | 3786±126 | 0.987±0.033 | 0.0067 |
| kfacvar_Kc_up | 2727±135 | 0.810±0.040 | 2625±133 | 1.100±0.056 | 3783±126 | 0.986±0.033 | 0.0067 |
| kfacvar_Kll_down | 2725±135 | 0.811±0.040 | 2625±133 | 1.100±0.056 | 3784±126 | 0.986±0.033 | 0.0067 |
| kfacvar_Kll_up | 2725±135 | 0.812±0.040 | 2625±133 | 1.100±0.056 | 3784±126 | 0.986±0.033 | 0.0067 |
| kfacvar_WbbWccjet1_down | 2735±136 | 0.813±0.040 | 2622±133 | 1.099±0.056 | 3777±126 | 0.984±0.033 | 0.0067 |
| kfacvar_WbbWccjet1_up | 2713±135 | 0.809±0.040 | 2630±133 | 1.102±0.056 | 3792±126 | 0.988±0.033 | 0.0067 |
| kfacvar_Wcjet1_down | 2737±136 | 0.814±0.040 | 2621±133 | 1.098±0.056 | 3775±126 | 0.984±0.033 | 0.0067 |
| kfacvar_Wcjet1_up | 2713±135 | 0.809±0.040 | 2630±133 | 1.102±0.056 | 3793±126 | 0.988±0.033 | 0.0067 |

#### QCD variations

| el_fake_alternate | 2725±135 | 0.811±0.040 | 2625±133 | 1.100±0.056 | 3784±126 | 0.986±0.033 | 0.0067 |
| el_fake_mc_down | 2725±135 | 0.811±0.040 | 2625±133 | 1.100±0.056 | 3784±126 | 0.986±0.033 | 0.0067 |
| el_fake_mc_up | 2725±135 | 0.811±0.040 | 2625±133 | 1.100±0.056 | 3784±126 | 0.986±0.033 | 0.0067 |
| el_real_alternate | 2725±135 | 0.811±0.040 | 2625±133 | 1.100±0.056 | 3784±126 | 0.986±0.033 | 0.0067 |
| mu_fake_alternate | 2729±135 | 0.813±0.040 | 2628±133 | 1.101±0.056 | 3775±126 | 0.984±0.033 | 0.0067 |
| mu_fake_mc_down | 2703±135 | 0.805±0.040 | 2618±133 | 1.097±0.056 | 3784±126 | 0.986±0.033 | 0.0067 |
| mu_fake_mc_up | 2734±135 | 0.814±0.040 | 2628±133 | 1.101±0.056 | 3787±126 | 0.987±0.033 | 0.0067 |
| mu_real_alternate | 2821±135 | 0.840±0.040 | 2634±133 | 1.104±0.056 | 3867±126 | 1.008±0.033 | 0.0067 |

Table B.2: The results of each variation for the fitting procedure, in the muon stream.
The number of events contain the fitfactor.
Appendix C - DannyLoop

All the analyses in this thesis have been run by a single codebase named DannyLoop.[283] The development of this package parallel with a group initiative called EventLoop (this not the same EventLoop as currently used by ATLAS). DannyLoop is written in C++ and uses ROOT extensively. It is BSD-licensed and available for public use. However, the current codebase should be considered an experimental release, as many imperfections remain.

The simplest form of a commandline to run DannyLoop is:

```
./DannyLoop -i <input> -o <output> -s <script>
```

where 'input' is a name of a text-file listing the input filenames, 'output' is the filename of the output ROOT file (which can be omitted if there is no output), and 'script' is the filename of a plain-text file describing the flow of the program.

The basic structure of DannyLoop consists of a loop iterating over all the input events. Any number of input ROOT files can be read, and the program transparently switches between the files when necessary. Then, per event, a set of modules is called in user-specified order. Which modules to use and their settings is defined in the script file. This set-up allows for easy configuration of the run, without the need to recompile.

The configuration file also contains the variable mapping. Internally, variables are reserved up-front, and can be mapped to both a TBranch of the input files, and a TBranch in the output file. The name of the variable in DannyLoop does not need to match that of the input TBranch, nor that of the output TBranch. This variable mapping allows for easy adaptation to different input and output file conventions.

DannyLoop was designed specifically to reduce the number of memory-copies. By using pointers to variables, DannyLoop avoids the problem of copying variables between modules during the run, increasing memory usage and slowing performance. This is possible because all modules register the variables they need with a central 'EventMap', which returns pointers to the variables. All this is done before the run starts, and the pointers are never
Appendix C

changed during the run, improving the results of caching done by the operating system and hardware. Another benefit of the up-front registration is that it allows DannyLoop to only read in the TBranch's that are actually used; unused branches are never loaded, and this saves processing time.

The up-front registration strategy extends to vector-objects. All vectors are sorted in-place; no copy of a vector is ever needed, not even a temporary one. This is more flexible than making all variables global, which is generally be regarded as bad coding practice. For example, each module dealing with jets will require a pointer to several variables describing the $p_x$, $p_y$, etc. of the jet. If these were global variables, all modules would have to use the same jet variables. But because these are pointers in DannyLoop, each modules' pointers can point to another set of jet variables. Additionally, because these pointers are set-up with information of the configuration file, the pointers can be made to point to another set of jet variables by mere modification of the configuration file; no search-replace of variables names followed by a recompilation is required. This means that for example the jet collection used for jet selection can be independently chosen from the module running the jet reconstruction; the jet collections used by both do not need to be the same.

Another powerful feature are the variable groups. In the layout currently used for D3PDs in ATLAS, the four-vectors of particles are stored as separate vectors of floating points, one containing all the $p_x$, another one for $p_y$, etc. This makes removing a single particle complicated: the N-th element of each vector needs to be removed. A module called 'VariableGroup' simplifies handling: any number of vectors can be added to it, and a single call to it can remove the N-th element of all the vectors associated to it. For sorting, a similar call for swapping two elements is used.

Similar to the way modules are implements, 'cutters' are implemented as well. These objects simply perform a cut (for example: $p_T > 25$ GeV), but because they inherit from a base-class, the module calling the cutter can be completely agnostic as to the kind of cut. Cutter-objects can be created through a factory that includes a string-parsing function, so text from the configuration file can be passed straight to the factory, resulting in a dynamic way of implementing arbitrary cuts. Currently, not only the usual operations (less than, equal, absolute value less than, etc.) are implemented, but also a range-cut that allows for checking whether a value lies within a certain interval (the crack region cut was implemented in a single cutter in this way).

Even though DannyLoop is build fully modular, sometimes some information exchange between modules is needed. For this reason, an inter-module event-based messaging system was added. An arbitrary number of modules can register an event handler, and any module in the program can call that handler. A void-pointer allows for arbitrary data to be sent with the message. The messages are identified by a string defined in the script-file (although technically, they can also be hardcoded in a module), which are internally translated to integers for optimization reasons.
The program also defines events to which modules can register through an overloadable function-call. These are events such as the program switching to a new input file, or a run starting or ending. Through these events, modules can set-up their own dedicated output file with any format required.

A few specific flags for controlling the flow of the program are available for modules. Currently, there are only three booleans: one to kill the event (further processing of the event stops, and it is not outputted to the output file), to kill the entire run, and a flag to force the saving of the event to the output file.

Several global settings in the configuration file determine the naming of the input TTree and the output TTree, and a few other options are available to change the way DannyLoop processes events. For example, the 'KeepAllKilledEvents' boolean flag ensures that even events that are dropped are written to the output file.

DannyLoop has been run on the Grid, through the Panda infrastructure. The commandline format was designed to be compatible with this kind of usage. Additionally, the code has been tested with Valgrind\[284\] on several occasions, and no major problems have shown up.

An interesting concept is that of modifying an event while it is being processed. This is implemented in the Debugger module. By defining breakpoints (both unconditional and conditional types are supported), the debugger can be called, interrupting the processing of the event at that point and providing a commandline interface where commands inspecting the values of loaded variables, and modifying those variables are available. Afterwards, processing of the event can be resumed with the modified values, or the event can be dropped.

Suggestions for improvement:

- Rename it.
- Better exception handling (using C++ exceptions instead of boolean return codes).
- Templating variable types (currently, all variables types are hardcoded, and duplicates of modules have to be created for each variable type).
- The output ROOT file code should be moved into a separate module.
- The addition of CopyVariable and SwapVariable modules to allow multiple streams of processing ("parallel" cuts, for example).
- Make the naming of various things consistent throughout the code. For example, currently sometimes the term 'script file' is used when referring to the 'config file'.
- Allocate the variables from a single block of memory. This will reduce memory fragmentation of these often used variables, resulting in less cache misses and thus a better performance. Ordering of variables (read-only versus read-write, variable sizes, etc.) could also lead to performance improvements (cachelines, padding and alignment, etc.).
Appendix C
References


References


References


[106] https://twiki.cern.ch/twiki/bin/view/Main/HforToolD3PD, HforToolD3PD.


References


[160] http://www.eu-emi.eu/releases/emi-3-montebianco/products/-/asset_publisher/5dKm/content/storm-se-2, StoRM service.


[168] http://xrootd.org/, XRootD.


[177] https://twiki.cern.ch/twiki/bin/view/PanDA/GenericPanDAPilot, *Generic PanDA Pilot*.


[194] https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TechnicalitiesForMedium1, *TechnicalitiesForMedium1*.
References


https://twiki.cern.ch/twiki/bin/view/AtlasProtected/BTaggingBenchmarks#b_tagging_Benchmarks_extending_M, BTaggingBenchmarks.

https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/TopCommonObjects, TopCommonObjects.

D0 Collaboration, *Measurement of the p$\bar{p} \rightarrow t \bar{t}$ production cross section at $\sqrt{s} = 1.96$ TeV in the fully hadronic decay channel.*, Phys. Rev. D76 (2007) 072007, hep-ex/0612040.


https://twiki.cern.ch/twiki/bin/view/AtlasProtected/CrackDefinition, Calorimeter "crack" definition.

https://twiki.cern.ch/twiki/bin/view/AtlasProtected/EtMiss, EtMiss.


D0 Collaboration, Measurement of the electron charge asymmetry in $p \bar{p} \rightarrow W + X \rightarrow e \nu + X$ events at $\sqrt{s} = 1.96$ TeV, Phys. Rev. Lett. 101 (2008) 211801, hep-ex/0807.3367.


CDF Collaboration, Measurement of the shape of the boson rapidity distribution for $p \bar{p} \rightarrow Z/\gamma^* \rightarrow e^+ e^- + X$ events produced at $\sqrt{s}$ of 1.96 TeV, Phys. Rev. D76 (2007), hep-ex/0702025.


References


[269] D.A. Mason, Measurement of the strange - antistrange asymmetry at NLO in QCD from NuTeV dimuon data, Oregon University, 2006, Ph.D. thesis.


[277] https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/SgTop-D3PDs8TeVLeptonJets#v11, SgTop-D3PDs for 8 TeV, Lepton+Jets Selection.


Summary

The single top t-channel fiducial cross section at 8 TeV measured with the ATLAS detector

This thesis describes my research work performed from 2010 to 2015 using 8 TeV proton-proton collision data collected with the ATLAS detector.

The ATLAS detector is one of the four major experiments of the LHC collider at CERN. It is a general-purpose detector, designed to record proton-proton collisions. Through the use of various types of subdetectors, the trajectories and energy deposits of particles created in the interaction are recorded. The raw data is processed and then sent to the Grid for further reconstruction and analysis.

The Grid is a distributed computing network, which connects many computing clusters located in various institutes worldwide. This gives the ATLAS collaboration the massive data-processing power that it needs in order to process the recorded data of the detector and simulated events. For this thesis, I heavily relied on the Nikhef Tier 1 Grid-site to analyse 50TB of data, using over a hundred CPU-years.

The ATLAS scientific programme focuses on measurements in a new energy and luminosity domain to understand the generation of mass of elementary particles. Meanwhile, the long elusive Higgs boson has been discovered which confirms the Standard Model prediction. The Standard Model is however considered as incomplete and new physics may be around the corner. Known processes may show subtle deviations between prediction and actual measurements.

The test-case which is pursuit in this thesis is the production of top-quarks. The top-quarks have a mass of about the electroweak scale and decay before the hadronization can start. Hence, the top-quark process offers a possibility to study a bare heavy quark for which accurate predictions exist which are confronted with data in this thesis.

In this thesis, a single top t-channel cross section measurement at 8 TeV is performed using the 2012 ATLAS data, corresponding to an integrated lumi of 20.3 fb. This cross section
measurement is a direct probe of the $V_{tb}$ CKM-matrix element, and the couplings that play a role for the $Wtb$-vertex.

To compare the data against the predictions of theory in the perturbative regime is not easy. First, events are generated using Monte Carlo (MC) techniques, which incorporate all non-perturbative effects by employing parton showering methods. In addition to that, the ATLAS detector response has to be simulated.

Besides the signal events, also all background events have to be generated. One of the main backgrounds is the $W^{+}\text{jets}$ process, which is generated in this thesis with the Alpgen software package. A study into the generator uncertainties of the Alpgen generator is performed. For the normalization of this background a data-driven extrapolation technique is used. The extrapolation uses the MC prediction for $W^{+}\text{jets}$ events with $b$- or $c$-jets in the final state. The use of MC information introduces uncertainties in the prediction. The study of these extrapolation uncertainties required the generation of almost 2.5 billion events with different settings, occupying over 10TB of disk-space. Previously, this uncertainty was estimated to be 25%, dominated mostly by the limited amount of event statistics in the sample. The study in this thesis leads to the relatively small uncertainty of 10%, which is now dominated by the actual theoretical uncertainty instead of statistical fluctuations.

**Fitting procedure**

Both the recorded data and the simulated Monte Carlo events are processed on the Grid to reconstruct physical objects like electrons, muons, and jets. By applying selection cuts on these objects, a region of phase space is selected where the single top t-channel signal is enhanced. Further event reconstruction leads to a distribution of the top-quark mass, which is used in a fitting procedure. The fit is able to separate the signal from the backgrounds, resulting in the measurement of the single top t-channel cross section:

$$\sigma_{t-channel} \ (8 \ \text{TeV, data}) = 83.4 \pm 2.1 \ (\text{stat.})^{+9.8}_{-9.6} \ (\text{syst.}) \ \text{pb} \ .$$

The total cross section as predicted by NNLO theory is:

$$\sigma_{t-channel} \ (8 \ \text{TeV, prod.}) = 87.2^{+2.8}_{-1.0} \ (\text{scale var.})^{+2.0}_{-2.2} \ (\text{PDF}) \ \text{pb} \ .$$

The measurement is compatible with this prediction.

The cross section measurement above uses a rather large correction to extrapolate from the phase space of the detected events to the total phase space using the MC prediction. In fact, this correction includes a large part of phase space that remains undetected by ATLAS and has large uncertainties. To reduce these uncertainties and to facilitate the comparisons of (new) predictions with the actual measurement a so-called fiducial cross
section measurement is also performed.
To define the fiducial sample, the selection is mimicked on truth particle level to provide an unambiguous definition of the fiducial region that can be reproduced independently of the MC generator used.

The uncertainty associated with the choice of signal MC generator is studied by using the differences between three generators: Powheg, aMC@NLO, and AcerMC. Their event predictions are compared at truth particle level, and the difference in their cross section measurements is used as the uncertainty on the signal MC choice. This results in a 8 TeV inclusive fiducial single top t-channel cross section measurement with a significant reduction of the systematic uncertainties:

\[
\sigma_{t\text{-channel},fid} (8 \text{ TeV}) = 4.07 \pm 0.09 \text{ (stat.)}^{+0.37}_{-0.34} \text{ (syst.) pb}.
\]

The corresponding predictions for Powheg, aMC@NLO, and AcerMC are 1.43 pb, 1.49 pb, and 1.24 pb respectively. The predictions are thus consistent with the observation, and there are no indications of new physics. However, the LHC has increased the proton-proton collision energies to 13 TeV, and more data than previously will be recorded, which allows to use more aggressive requirements to remove background. This may lead to the discovery of new physics through the top-quark sector.
Samenvatting

De single top t-channel fiduciële doorsnede op 8 TeV gemeten met de ATLAS detector

Deze manuscript beschrijft het onderzoek dat ik gedaan heb van 2010 tot 2015 gebruikmakend van de 8 TeV data van proton-proton botsingen verzameld door de ATLAS detector.

De ATLAS detector is een van de vier grote experimenten bij de LHC deeltjesversneller van CERN. Het is een detector die zo gebouwd is dat deze meerdere type deeltjes kan onderzoeken die bij de proton-proton botsingen geproduceerd worden. Door middel van verscheidene subdetectoren wordt en de banen en energie afzettingen van de deeltjes die in de botsing gecreëerd zijn vastgesteld. De ruwe data wordt daarna verwerkt en naar het Grid verzonden voor verdere reconstructie en analyse.

Het Grid is een gedistribueerd computer netwerk, waarbij vele computer clusters die wereldwijd bij verschillende instituten staan verbonden worden. Dit geeft de ATLAS colaboratie de computerkracht die het nodig heeft om alle opgenomen data en gesimuleerde botsingen te verwerken. Voor deze manuscript heb ik dankbaar gebruik gemaakt van de Nikhef Tier 1 Grid-site voor de analyse van 50TB aan data, waarbij ik meer dan een honderd CPU-jaar heb gebruikt.

Het wetenschappelijk programma van ATLAS richt zich op metingen met een hogere energie dan ooit tevoren, en heeft als doel het begrijpen van het proces dat de massa van elementaire deeltjes genereert. Ondertussen is het lang gezocht Higgs boson gevonden, en dit is weer een bevestiging dat Standaard Model accuraat is. Het Standaard Model wordt daarentegen als incompleet beschouwd, en nieuwe fysica kan zomaar binnen handbereik liggen. Daarom is het belangrijk om bekende fysische processen goed te onderzoeken of er toch subtiele afwijkingen zichtbaar zijn tussen de voorspelling en de meting.

Het onderzoeksgebied van deze manuscript is de productie van top-quarks. Top-quarks hebben een massa die dicht bij de electrozwakke schaal ligt, en ze vervallen voordat hadronisatie plaats kan vinden. Hierdoor vormt de studie van top-quark processen die in deze
manuscript gedaan wordt een goede mogelijkheid om een naakt zwaar quark te observeren om de voorspellingen over het gedrag te toetsen aan de data.

In dit manuscript wordt gekeken naar een single top t-channel werkzame doorsnede meting bij 8 TeV. Deze wordt uitgevoerd met 2012 ATLAS data, corresponderend met een geïntegreerde luminositeit van 20.3 fb. Deze doorsnede meting is een directe toets van het $V_{tb}$ CKM-matrix element, en de koppelingen die een rol spelen voor de $Wtb$-vertex.

Het is niet eenvoudig om de data met de voorspellingen in het perturbatieve regiën te vergelijken. Allereerst worden botsingen gesimuleerd door middel van Monte Carlo (MC) technieken, die alle niet-perturbatieve effecten door middel van parton showering methodes meenemen. Daar bovenop moet de response van de ATLAS detector ook gesimuleerd worden.

Naast signaal botsingen moeten ook achtergrond botsingen gesimuleerd worden om tot een volledig beeld te komen. Een van de prominente achtergronden is het $W^+\text{jets}$ proces, die in deze manuscript gegenereerd is met het Alpgen software pakket. Er wordt een studie gedaan naar de generator onzekerheden die Alpgen introduceert. Voor de normalisatie van deze achtergrond wordt een data-gedreven extrapolatie techniek gebruikt. De extrapolatie gebruikt de MC voorspelling voor $W^+\text{jets}$ botsingen met $b$- of $c$-jets in de eindtoestand. Het gebruik van MC informatie introduceert onzekerheden in de voorspelling.

De studie van deze extrapolatie onzekerheden vereist de generatie van bijna 2,5 miljard botsingen met verschillende configuraties, die bij elkaar meer dan 10TB schijfruimte in beslag nemen. Voorheen werd deze onzekerheid afgeschat zijnde 25%, voornamelijk gedomineerd door de gelimiteerde hoeveelheid statistiek die beschikbaar was in het sample. De studie in deze manuscript verbetert deze meting naar 10%, en deze wordt nu gedomineerd door daadwerkelijke intrinsieke onzekerheden in plaats van statistische fluctuaties.

**Fit procedure**

Zowel de data en de gesimuleerde Monte Carlo botsingen worden op het Grid gereconstrueerd tot fysische objecten zoals elektronen, muonen, en jets. Door selectie snedes toe te passen worden een region van de faseruimte geselecteerd waar het single top t-channel signaal sterk is. Verdere reconstructie leidt tot een distributie van de top-quark massa, die gebruikt wordt in een fit procedure. De fit is in staat om het signaal van de achtergronden te scheiden, wat resulteert in de meting van de single top t-channel doorsnede van:

$$\sigma_{t-\text{kanaal}} (8 \text{ TeV, data}) = 83.4 \pm 2.1 \text{ (stat.)}_{-9.6}^{+9.8} \text{ (syst.) pb} .$$

De totale doorsnede zoals deze voorspeld wordt door de NNLO theorie is:

$$\sigma_{t-\text{kanaal}} (8 \text{ TeV, voorsp.}) = 87.2_{-1.0}^{+2.8} \text{ (schaal var.)}_{-2.2}^{+2.0} \text{ (PDF) pb} .$$
De meting is compatibel met deze voorspelling.

Bovenstaande werkzame doorsnede meting heeft een nogal grote correctie om van de faseruimte van de gedetecteerde botsingen naar een volledige faseruimte te extrapoleren, waarvoor MC voorspellingen worden gebruikt. In het bijzonder bevat deze correctie een groot stuk faseruimte die niet met ATLAS gemeten kan worden, en hierdoor worden grote onzekerheden geïntroduceerd. Om deze onzekerheden te verkleinen en om een vergelijking tussen de meting en (nieuwe) voorspellingen mogelijk te maken, wordt een zogenaamde fiduciële doorsnede meting uitgevoerd. Om het fiduciële sample te definiëren wordt de selectie op een reconstructie-nivo uitgevoerd waarbij gebruik wordt gemaakt van de 'werkelijke' deeltjes-informatie van de MC generator. Dit wordt het 'truth particle level' genoemd.

Voor het bepalen van de onzekerheid geassocieerd met de keuze van de MC signaal generator wordt gebruik gemaakt van drie verschillende generatoren: Powheg, aMC@NLO, en AcerMC. Ieders voorspelde botsingen worden op deeltjes nivo met elkaar vergeleken, en het verschil in de doorsnede wordt gebruikt als de onzekerheid van de MC generator keuze. Dit resulteert in een fiduciële 8 TeV single top t-channel doorsnede meting met een gereduceerde onzekerheid, voor elektronen en muonen gecombineerd:

\[
\sigma_{t\text{-kanaal, fid}} (8 \text{ TeV}) = 4.07 \pm 0.09 \text{ (stat.)}^{+0.37}_{-0.34} \text{ (syst.)} \text{ pb} \cdot
\]

De voorspellingen voor Powheg, aMC@NLO, en AcerMC zijn 1.43 pb, 1.49 pb, en 1.24 pb respectievelijk. Omdat mijn resultaat compatibel is met de theoretische voorspelling, heb ik dus geen indicatie van nieuwe fysica ontdekt. Maar de LHC heeft de energie van proton-proton botsingen verhoogd (naar 13 TeV). Ook wordt er naar verwachting een hogere geïntegreerde luminositeit verzameld, waardoor er agressievere achtergrond snedes gebruikt kunnen worden. Dit kan leiden tot een ontdekking van nieuwe fysica in de top-quark sector.

**Samenvatting**

De meting is compatibel met deze voorspelling.
Wow, waar moet ik beginnen? Er zijn zoveel dingen waarover ik het hier kan hebben. Ik moet in ieder geval een aantal dingen vertellen over de elementaire deeltjes waarmee ik gewerkt heb, een aantal technieken die gebruikt worden, mijn resultaten... Maar ik zal bij het begin beginnen. Het meest prominente nieuwe speeltje wanneer je een promotie-traject start bij de ATLAS-group te Nikhef is, natuurlijk, de ATLAS detector.

De ATLAS detector is één van de vier grote experimenten van de LHC deeltjes versneller van CERN. ATLAS is een detector die gebouwd is om verschillende typen fysica die bij proton-proton botsingen gebeuren te onderzoeken. Door gebruik te maken van verschillende soorten subdetectoren kan zo veel mogelijk informatie over de deeltjes die vrijkomen bij zo'n botsing vergaard worden. Dit levert een gigantische hoeveelheid data op, maar hoe kan ik nou met zoveel data omgaan?

De data van ATLAS wordt verwerkt met het Grid; een gedistribueerd computing netwerk, waarbij de computer-kracht van instituten wereldwijd aan elkaar wordt geknoopt om zo een groot computer-cluster te vormen. Dit geeft de ATLAS analyses de hoeveelheid computerkracht die nodig zijn om alle data te verwerken. Voor dit proefschrift heb ik veelal gebruik gemaakt van de Nikhef Tier 1 Grid-site; ik heb meer dan 100 CPU-jaar verbruikt.

En waar ben ik dan naar op zoek in al die data?

Het doel van ATLAS is om scherpere metingen dan ooit tevoren te doen, en om nieuwe, verrassende deeltjesfysica te ontdekken. Een onderdeel hiervan is het testen van het Standaard Model. Het Standaard Model (SM) van deeltjesfysica beschrijft alle tot nu toe bekende elementaire deeltjes en hun interacties. Alle bekende elementaire deeltjes zijn in figuur 1.1 op bladzijde 2 afgebeeld. Er zijn er een heleboel verschillende, maar ik kijk voornamelijk naar die "r", de derde van links in de bovenste rij; dat is het "top-quark". De productie van top-quarks wordt door het Standaard Model beschreven. Top-quarks zijn ook speciaal door hun grote massa; het is veel zwaarder dan de meeste andere deeltjes. Dit maakt het top-quark speciaal, omdat een aantal theorieënn dingen voorspellen die zichtbaar zouden
moeten zijn in het gedrag van dit zware quark. Dat is waarom dat ik er naar kijk, om te zien of er een afwijking in het gedrag zit dat we niet verwachten. En omdat het top-quark ook gekoppeld is aan het nieuw ontdekte Higgs boson, zou een ontdekking van voorheen onbekend gedrag bij top-quarks ook invloed kunnen hebben op hoe Higgs bosonen zich gedragen.

In dit proefschrift voer ik een "single top" (productie van maar één top-quark) werkzame doorsnede (productie snelheid) meting uit in het zogenaamde "t-kanaal", gebruikmakend van de 2012 data van ATLAS.

Om de gemeten data met de theoretische voorspellingen te kunnen vergelijken moeten er eerst een heleboel virtuele botsingen gesimuleerd worden, met behulp van Monte Carlo (MC) technieken. Dit behelst ook het simuleren van de ATLAS detector zelf. Dit kost allemaal veel computerkracht, en ook hier wordt het Grid weer voor gebruikt. Ikzelf heb een specifiek achtergrondsproces genaamd $W^+$jets bestudeerd. Hierdoor heb ik een onzekerheid van een bepaalde voorspelling die gebruikt wordt bij top-quark metingen kunnen reduceren van 25% tot ongeveer 10%. Aangezien deze onzekerheid steeds belangrijker werd doordat andere onzekerheden steeds meer krompen was dit een belangrijk resultaat dat ondertussen door alle top-quark onderzoeken bij ATLAS gebruikt wordt.

En de volgende stap? Het meten van het top-quark zelf!

Mijn meting is op te splitsen in twee stappen. In de eerste stap wordt een meting van de werkzame doorsnede met behulp van de data gedaan. Hiervoor zijn relatief grote correcties nodig die berekend zijn met behulp van gesimuleerd signaal, waardoor het resultaat theoretische onzekerheden zal bevatten. Deze onzekerheden worden afgeschat door met meerdere verschillende generatoren gesimuleerde botsingen te maken, en het verschil ertussen als onzekerheid aan te wijzen. Nu blijkt dat deze onzekerheid vrij groot is, en dat dat deels komt doordat er relatief grote verschillen tussen de MC generatoren zitten in een gebied waar we niet kunnen meten. Oftewel: een deel van de onzekerheid die aan de meting wordt toegeweven komt van een regio waar we eigenlijk geen data van hebben. Dit vergroot nodeloos de onzekerheid, dus stap twee is het bepalen van deze onzekerheid zonder dit gebied mee te nemen.

Uiteindelijk levert dit het volgende resultaat op voor de zogeheten fiduciële single top t-kanaal werkzame doorsnede:

$$\sigma_{t-kanaal,fid}(8 \text{ TeV}) = 4.07 \pm 0.09 \text{ (stat.)}^{+0.37}_{-0.34} \text{ (syst.) pb} .$$

Deze meting is compatibel met de waarde die door het Standaard Model wordt voorspeld. Dit houdt dus in dat ik zie wat het Standaard Model voorspelt dat ik zou moeten zien, en ik dus geen duidelijke sporen van nieuwe, spannende natuurkunde ontdekt heb. Maar dat is zeker geen slecht nieuws: het is namelijk een sterke bevestiging dat de beschrijving van de
interacties tussen deeltjes zoals het Standaard Model geeft, een erg goede beschrijving van de werkelijkheid is!
Ondertussen is de LHC deeltjes versneller opgewaardeerd waardoor deze op een veel hogere energie protonen kan botsen. Hierdoor zullen er veel meer top-quarks geproduceerd worden, waardoor er meer data beschikbaar komt. Dat heeft weer tot gevolg dat er strengere snedes gebruikt kunnen worden, zodat er naar een puurder top-quark signaal gekeken kan worden. Gecombineerd met de technieken die ik toegepast heb in dit proefschrift kunnen er in de toekomst scherpere studies uitgevoerd worden naar top-quarks, wat zal leiden tot een nog beter begrip van hoe de wereld in elkaar zit.
There are way too many people that I should be thanking... So let me first of all thank all those whose names I haven't mentioned in this section! Sorry about that!

First of all, I gratefully thank both Stan Bentvelsen and Marcel Vreeswijk, my promotor and co-promotor respectively. Without their guidance and wisdom I would never have been able to finish my PhD, nor would I have learned even a fraction of that which I did. Stan, Marcel, thank you both for everything!

Without the support of my family I would probably not have survived it either. Their moral and practical support really helped me to pull through.

Thank you Hurng, for your assistance with my Grid service work. Thanks to Wouter Verkerke for helping me get Alpgen running smoothly, and for putting up with my C++ and ROOT questions. And much thanks to Duc for his invaluable help with the samples and with my analysis.

Rogier, thank you for working by my side during the last few stages of my analysis. You really helped pull me through. And thank you Gossie, for bestowing the role of IT-guy on me.

A special thanks to Wouter v/d W. for putting up with my coffee-breaks and conversational topics. ;)

Thanks to all those at the Nikhef Helpdesk who helped me with my computer issues and questions. A special thanks to David, Ronald, Ton, Dennis, Mischa, and Jeff. You've all helped me a lot by resolving all kinds of problems36 I ran into when working with Stoomboot, the Tier 1 site, and the Grid in general. And of course, I'd like to thank those last three computing clusters as well for enabling me to run ridiculously large amounts of code and data.

36. Most of which I probably caused myself...
And while I'm thanking inanimate objects anyway, I'd like to give my sincerest thanks to Genil, Melito, Venosa, Mella, and Muga, my desktop computers. Without you, I (literally!) couldn't have done all this work!

A quick thanks to Claire Gwenlan for the patience and understanding she displayed when I found anomalies\(^{37}\) in my Alpgen samples. I should also mention Michael Day and the team at Prince, for their quick responses every time I hit a problem with the Prince-software.

And of course, Tristan. How can I ever thank you enough for your help and distractions? Our projects together provided the outlet I often times needed when goings got tough. And I hope people will remember as fondly as I do all the havoc we caused.

Finally, Nicole. What can I say that does justice? It's always awesome to meet somebody who is pleasantly disturbed in the same way as yourself. So let me just remind you of one thing: watch out for 'stoeptegels'. You're one of the only ones that knows...

---

37. Ok, mistakes. I messed up, but nobody got hurt!
Index

A
AcerMC: page 35
AFS: page 92
Alpgen: page 36, 40
ALPHA algorithm: page 40
AMC@NLO: page 36
Analysis object data: page 95
Ariadne: page 36
Asymptotic freedom: page 13
Athena: page 87
ATLAS: page 72

B
b-tagging: page 106
Baryon: page 2
Beam remnants: page 23
Beampipe: page 69
Build job: page 94

C
CERN: page 1
CKM-matrix: page 10
Cloud: page 88
Cloud computing: page 95
Color confinement: page 14
Computing element: page 94
Conditions database: page 84
CP violation: page 8
Crack region: page 77
CT10: page 138
CTEQ6: page 17

D
DannyLoop: page 161
Data acquisition: page 79
Dataset: page 88
DaTRi: page 89
Detector-level: page 58
DGLAP equations: page 17
Diboson processes: page 40
Digitization: page 84
Distributed Data Management: page 90
Double parton interaction: page 38
DPD: page 96
DQ2: page 90
Drell-Yan process: page 40

E
Electromagnetic calorimeter: page 76
Electroweak interactions: page 7
Electroweak theory: page 7
Event Filter: page 80
Event summary data: page 95

F
Fake electrons: page 113
FAX: page 94
Fiducial cross section: page 143
Fiducial volume: page 143
File transfer service: page 90
Final state radiation: page 16, 23
Fine structure constant: page 7
Forward calorimeter: page 77
FullSim: page 84
**G**
- **Ganga**: page 93
- **GEANT4**: page 84
- **Good run list**: page 110
- **GridFTP**: page 90

**H**
- **Hadron**: page 2
- **Hadronic calorimeter**: page 77
- **Hadronization**: page 15, 24
- **Hard scattering**: page 15, 16
- **Heavy flavor**: page 43
- **Heavy flavor fraction**: page 54
- **Heavy Flavor Overlap Removal**: page 48
- **Herwig**: page 37
- **Herwig++**: page 37
- **Hit**: page 75

**I**
- **Impact parameter**: page 106
- **Initial state radiation**: page 16, 23
- **Inner detector**: page 73
- **Insertable B-layer**: page 154

**J**
- **Jet**: page 14
- **Jet matching**: page 44
- **Jet vertex fraction**: page 110
- **Jetbin**: page 54
- **Jimmy**: page 37
- **Job**: page 92

**K**
- **k-factor**: page 21

**L**
- **Large Hadron Collider**: page 69
- **LCG File Catalog**: page 90
- **Leading log**: page 23
- **Leading order**: page 20
- **Lepton dressing**: page 146
- **Level 1 trigger**: page 79
- **Level 2 trigger**: page 80
- **LHC Optical Private Network**: page 90
- **Logical file name**: page 90
- **LCG File Catalog**: page 90
- **Leading order**: page 20
- **Leading log**: page 23
- **Leading order**: page 20
- **Lepton dressing**: page 146
- **Level 1 trigger**: page 79

**M**
- **MadGraph**: page 36
- **Mandelstam variable**: page 28
- **Matrix element**: page 21
- **MC@NLO**: page 36
- **Meson**: page 2
- **Minbias**: page 83, 85
- **Missing transverse energy**: page 114
- **Monitored drift tube**: page 78
- **Monte Carlo**: page 14
- **MSTW2008**: page 17
- **Muid muon**: page 102
- **Multiple parton interaction**: page 38
- **Muon spectrometer**: page 77

**N**
- **Next-to-leading order**: page 20
- **NNPDF**: page 138

**O**
- **Offline processing**: page 87
- **Online processing**: page 87

**P**
- **PandaTools**: page 93
- **Particle identification**: page 100
- **Particle level**: page 146
- **Parton density function**: page 17
- **Parton distribution function**: page 15
- **Parton level**: page 45
- **Pile-up**: page 85
- **Pile-up vertices**: page 114
- **Pilot framework**: page 94
- **Pixel detector**: page 74
- **POOL**: page 87
- **Powheg**: page 36
- **Primary vertex**: page 114
- **Proto-jet**: page 104
- **Pseudo-rapidity**: page 73
- **Pythia**: page 37
Index

Q
QCD multijets: page 32
Quantum Chromo Dynamics: page 7, 10
Quantum Electro Dynamics: page 7
Quantum field theory: page 7

R
Rapidity: page 73
RAW: page 86
Resistive plate chamber: page 78
ROOT: page 96
RootCore: page 96
Rucio: page 90
Run job: page 94

S
Secondary vertex: page 106, 114
Segment: page 102
Semi-conductor tracker: page 75
Sherpa: page 36
Soft collision: page 79
Staco muon: page 102
Standalone muon: page 102
Standard Model: page 1
Storage element: page 90
Storage URL: page 91
Stream: page 81

T
Tile calorimeter: page 77
TopRootCore: page 97
Transfer URL: page 91
Trigger chain: page 81
Trigger menu: page 81
Trigger system: page 79
Truth-level: page 57

U
Underlying event: page 15, 24

V
Vector boson fusion: page 20

W
W+jets: page 39
Worker node: page 94

Z
Z+jets: page 40