Universität Bonn

Physikalisches Institut

Search for single top-quark production in the *Wt*-channel with 1 lepton and 2 jets at ATLAS

Pienpen Seema

A search for single top-quark production and an associated W boson is presented. The search is conducted in the lepton+jets channel requiring one charged lepton, two quark jets, one of which is a b-tagged jet, and missing transverse energy. Collision data collected by the ATLAS detector in 2011 with an integrated luminosity of 4.7 fb^{-1} of proton-proton collisions at a center-of-mass energy of 7 TeV were used. The study of the Wt channel is challenging since it suffers strongly from large backgrounds – mainly top-quark pair production and W+jets production. To discriminate the signal from the backgrounds a neural network was used. Three different approaches to signal discrimination were examined: training signal against a mixture of all backgrounds; training signal against each background separately; and a two-dimensional distribution of the neural network outputs of signal trained against each of the two main backgrounds. The first two approaches do not provide good separation against top-quark pair production. The use of the two-dimensional distribution strategy provides a better separation power. In order to extract the signal, a binned maximum likelihood fit was performed. A cross section for the Wt signal of $\sigma = 7.9 \pm 3.5$ pb is obtained with an expected significance of 4.4σ and an observed significance of 2.4σ, where any statistical uncertainties have been taken into account.

Physikalisches Institut der
Universität Bonn
Nußallee 12
D-53115 Bonn

BONN-IB-2013-01
April 2013

Universität Bonn

Physikalisches Institut

Search for single top-quark production in the *Wt*-channel with 1 lepton and 2 jets at ATLAS

Pienpen Seema

Dieser Forschungsbericht wurde als Masterarbeit von der Mathematisch-Naturwissenschaftlichen Fakultät der Universität Bonn angenommen.

Angenommen am:	09.10.2012
Referent:	Prof. Dr. Ian C. Brock
Koreferentin:	Prof. Dr. Jochen Dingfelder

Contents

Chapter 1

Introduction

Understanding of elementary particle physics is beneficial to all fields of science since it might answer the simple question about the building blocks of matter. The Standard Model of elementary particles and the interactions between them is so far the best theory to understand the whole picture of the subatomic world. It is based on relativistic quantum field theories. Over the last 40 years, the Standard Model has proven to be a well-tested theory of physics through many experiments. Furthermore, the last missing piece of the Standard Model particles, the "Higgs boson", has probably been found, as announced on 4 July 2012 after 48 years of "Higgs searching". The ATLAS and the CMS collaborations at the LHC announced the discovery of the Higgs-like particle using collision data with an integrated luminosity of about 10 fb^{-1} of proton-proton collisions collected at a center-of-mass energy, $\sqrt{s}$ = 7-8 TeV.

In the Standard Model, the heaviest among the observed elementary particle is the top quark. It has a large mass of about 173 GeV, that lead to its late discovery in 1995 at the Tevatron proton-antiproton collider. Many measurements of the top quark have been made since then. Furthermore, the mass of the top quark can be used to constrain the mass of the Higgs boson and some measurements of top-quark production are also sensitive to new physics beyond the Standard Model. The top quark can be produced via either the strong interaction or the weak interaction. The major production mechanism of the top quark both at the Tevatron and the LHC is in pairs of a top and an antitop quark via the strong interaction. Another source is single top-quark production via the weak interaction; however, it contributes less than top-quark pair production. Searches for single top-quark production were first performed at the Tevatron; however, it took another 14 years from discovery of the top quark to that of single top-quark production. The main reason was that the beam energy was not so high at the Tevatron. For the final operation, it was at $\sqrt{s}$ = 1.96 TeV. At the LHC, the operation was at $\sqrt{s}$ = 7 TeV in 2010-2011 and top-quark production via the weak interaction was first measured in 2011. Studies of single top-quark production can probe the Standard Model in several ways and provide information for new physics beyond the Standard Model; therefore, measurements of single top-quark production are absolutely advantageous.

This thesis is part of the efforts of the Bonn single top-quark group in their search for Wt-channel production in the lepton+jets mode in ATLAS. The lepton+jets mode normally contains one charged lepton, one b-quark jet, two light-quark jets and missing transverse energy from one neutrino in the final state measured with the ATLAS detector. In reality one quark jet can be missed in the final state; but on the other hand, one additional light-parton jet can be produced. The analysis presented in this thesis is related to the final state of the lepton+jets mode that misses one of the two light-quark jets. This topology contributes about 42% in all. It has a large background especially from W+jets production. Furthermore, the two main backgrounds: top-quark pair production and W+jets production have very similar signatures to that of the signal. As a result, it is difficult to separate the signal from the large irreducible backgrounds.

A neural network is used to combine many variables into one powerful discriminator. The aim of the thesis is to find a method which provides the best separation power from both main backgrounds at the same time. Training signal against a mixture of all backgrounds is first introduced as the standard approach. Then two additional approaches are investigated. Training a network of signal against each background separately is one of the two additional approaches. The second method is to combine the outputs of signal trained against each of the two main backgrounds into a two-dimensional distribution. In the last step, the tool Bill is used to perform a binned maximum likelihood fit to the final output discriminator in order to measure the cross section of the *Wt*-channel production and compute the significance of the observed signal using hypothesis testing.

This thesis is structured as follows. An overview of the Standard Model of elementary particle physics including a general introduction to top-quark physics is given in Chapter 2. The topic of Chapter 3 is the ATLAS experiment. It consists of a description of the LHC as well as of all main components of the ATLAS detector. In Chapter 4, *Wt*-channel production in the lepton+jets mode is described in more detail. The exact final state of the signal for this thesis and all relevant background processes are explained in the second half of the chapter. The general analysis setup containing some details of the used datasets, the definitions of physical objects and the event selection is described in Chapter 5. In this chapter, the methods of estimating multijet background using data-driven methods are also discussed. In Chapter 6, the strategies of the analysis based on the multivariate technique are reported. In addition to the main analysis, a study of the *b*-tagging working point is also given in this chapter. Chapter 7 describes the signal extraction using a likelihood fitting package Bill. A measurement of the cross section and a calculation of the significance of the observed signal are presented here. In Chapter 8, the results of the thesis are summarised.

Chapter 2

Theoretical Basics

The current theory of elementary particles and the interactions between them is called the Standard Model. In the first part of this chapter, the basic theoretical knowledge regarding the Standard Model of elementary particle physics is described. Since the top quark is the key for this analysis, the second half of this chapter details its properties and its production as well as its decay. Before that, some quantities used in elementary particle physics are introduced.

2.1 Physical Quantities

Natural units. Two fundamental constants always appear in particle physics: the Planck constant, h, (or the reduced Planck constant, $\hbar$) and the speed of light, c. Their values in SI units are

$$\hbar = \frac{h}{2\pi} = 1.055 \times 10^{-34}\ \text{Js},$$

$$c = 2.998 \times 10^{8}\ \text{m/s}.$$

In particle physics, it is convenient to use the so-called "natural units" defined by choosing

$$\hbar = c = 1. \tag{2.1}$$

Using natural units, one may drop $\hbar$ and c from all units and formulae together, a practice that will be adopted for the rest of this thesis. For example, mass, momentum and energy are all expressed in terms of energy units, or length and time are in units of inverse energy. Furthermore, the electron volt (eV) is common to use in particle physics as the unit of energy; 1 eV is equal to $1.6 \cdot 10^{-19}$ J.

Decay rate, lifetime and branching ratio. The decay rate (also known as the decay width), Γ, is defined as the probability per unit time that any given particle will disintegrate. Often particles can decay in several ways called the partial decay rates, Γ_i. The sum of each partial decay rate is known as the total decay rate, $\Gamma_{\text{tot}} = \sum \Gamma_i$. The unit of the decay rate is expressed in eV. The mean lifetime of a particle, τ, is directly related to its total decay rate as

$$\tau = \frac{1}{\Gamma_{\text{tot}}}. \tag{2.2}$$

One can determine the probability of a particle to decay into a particular decay mode in terms of the

branching ratio, $\mathcal{B}$, which is defined as

$$\mathcal{B} = \frac{\Gamma_i}{\Gamma_{\text{tot}}}. \tag{2.3}$$

Therefore, one can easily calculate the lifetime and branching ratios of particles from the decay rates.

Cross section. In a scattering process the cross section, σ, represents the effective area in which the interaction takes place. In a case of particle-particle colliders, one can consider the cross section of the process as the effective area of a target hit by a probe. The cross section is expressed in barn (b), $1\text{b} = 10^{-28}\ \text{m}^2$. The cross section can be calculated using the Feynman rules as

$$\sigma \propto |M_{ij}|^2 \cdot \rho, \tag{2.4}$$

where M_{ij} is the matrix element which refers to the transition amplitude of the initial state, i, to the final state, j, of the process and ρ is the phase space.

Luminosity. Luminosity, $\mathcal{L}$, is defined as the number of particles per unit area, per unit time. In the case of incoming particles with uniform luminosity, the number of particles per unit time hitting on the effective area, σ, is

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \mathcal{L} \cdot \sigma. \tag{2.5}$$

This luminosity is normally expressed in $\text{cm}^{-2}\text{s}^{-1}$ or $\text{b}^{-1}\text{s}^{-1}$. Besides the uniform (instantaneous) luminosity, the integrated luminosity defined as $\int \mathcal{L}\mathrm{d}t$ is also introduced. The integrated luminosity is therefore expressed in b^{-1} (commonly used in particle physics as pb^{-1} and fb^{-1}). The luminosity is one of two parameters (the other one is the beam energy) which is used to qualify the performance of a collider.

2.2 Standard Model

The principles of symmetries are extremely important and they guide the construction of the Standard Model. However, many phenomena in particle physics signify the need of symmetry breaking, so that both symmetries and symmetry breaking play important roles in the theory. The Standard Model is a gauge field theory with the group symmetries

$$SU(3)_C \times SU(2)_L \times U(1)_Y. \tag{2.6}$$

These symmetries imply that the Standard Model Lagrangian density, L, (simply Lagrangian from here on) is invariant under a local $U(1)_Y$ transformation, a local $SU(2)_L$ transformation and a local $SU(3)_C$ transformation as well as under the $U(1)_Y \times SU(2)_L$ transformation. According to the Glashow-Salam-Weinberg (GSW) theory, the electromagnetic and weak interactions are unified in the electroweak interaction, that is a significant step towards the unification of all fundamental forces in nature. The electroweak interaction is described by the $U(1)_Y \times SU(2)_L$ gauge group which allows the mixing between the gauge bosons of both groups. Because of the mixing, the physical $W^{\pm}$ and Z bosons are the force carriers of the weak interaction while the photon is the force carrier of the electromagnetic interaction.

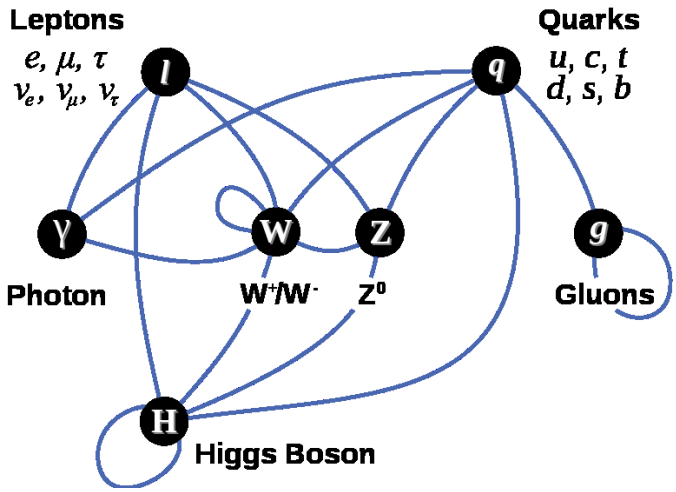

Figure 2.1: Overview of particles and interactions in the Standard Model picture.

Quantum chromodynamics, QCD, the theory of the strong interaction, is described by the $SU(3)_C$ group. The mediators of the strong interaction are the so-called gluons which are massless and electrically neutral. In this theory, there are three basic states of color charge, named red, green and blue. Each quark carries one unit of the color charge while each antiquark carries one of three anticolors, denoted antired, antigreen and antiblue. Each gluon carries a mixture of a color state and an anticolor state. There are nine possibilities of gluons, but only eight gluons remain independent color states. Besides the quark-gluon interactions, gluon-gluon interactions are a part of the strong interaction, since the gluons themselves carry color charge.

2.2.1 Particles and their interactions

In the Standard Model there are two fundamental classes of particles, which are called fermions and bosons. The fermions include six quarks and six leptons. They have spin $\frac{1}{2}$; therefore, they are subjected to the Pauli exclusion principle. Regarding to the weak isospin states, the quarks and leptons are grouped separately into doublets in three generations as

$$\begin{pmatrix} u \\ d \end{pmatrix}_L \quad \begin{pmatrix} c \\ s \end{pmatrix}_L \quad \begin{pmatrix} t \\ b \end{pmatrix}_L$$
$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L .$$

Each up-type quark carries electric charge $Q = +\frac{2}{3}e$ and each down-type quark carries, $Q = -\frac{1}{3}e$. While each of the charged leptons carries $Q = -1e$, all neutrinos are neutral. Furthermore, each Standard Model particle has a corresponding antiparticle with Z, H, γ being their own antiparticle. Antiparticles have the electric charge and other charges of opposite sign to their corresponding particles, but particles and antiparticles have the same mass, lifetime and spin. Normally the antiparticles are indicated by a bar over the symbol of their corresponding particles except for charged leptons and charged bosons. All those doublets are in left-handed states while the right-handed states only form singlets which do not

interact via charged weak interactions. The fermions, except neutrinos, can occur in both left-handed and right-handed states; however, the left-handed states are normally preferred due to chirality suppression. The quarks are arranged in the upper doublets and the leptons in the lower doublets. The up-type quarks (u,c,t) and neutral leptons (ν_e, ν_μ, ν_τ) have the third component of the weak isospin, $T_3 = +\frac{1}{2}$. The down-type fermions have $T_3 = -\frac{1}{2}$. Table 2.1 shows all elementary particles in the Standard Model and their masses.

Quark		Lepton		Boson	
Flavor	Mass [GeV]	Flavor	Mass [GeV]		Mass [GeV]
u	$(2.3^{+0.7}_{-0.5}) \cdot 10^{-3}$	ν_e	≈ 0	γ	0
d	$(4.8^{+0.7}_{-0.3}) \cdot 10^{-3}$	e	$0.511 \cdot 10^{-3}$	g	0
c	1.275±0.025	ν_μ	≈ 0	$W^\pm$	80.385±0.015
s	$(95\pm5) \cdot 10^{-3}$	μ	0.105	Z	91.188±0.002
t	173.5±0.6±0.8	ν_τ	≈ 0	H^0	> 115.5
b	4.18±0.03	τ	1.777		

Table 2.1: Elementary particles in the Standard Model and their masses from the 2012 PDG particle physics booklet.

Quarks. Since all quarks carry color charge, they cluster together to form color neutral composite particles named hadrons due to color confinement. According to the color confinement phenomenon, the quarks cannot be isolated. They are held together by the strong interaction. Besides the strong interaction, quarks can interact with other fermions via both the electromagnetic and weak interactions. There are two classifications of hadrons: baryons containing three quarks and mesons containing a quark and an antiquark. Baryons are half-integer spin particles and mesons are integer spin particles.

Leptons. The best-known of all leptons is the electron, e. Leptons cannot interact via the strong interaction since they do not carry the color charge. The three neutrinos are subjected only to the weak interaction. They interact rarely with matter such that they are hardly detectable. On the other hand, all charged leptons, electron muon and tau, interact also electromagnetically.

The higher generation a fermion, the greater mass it has; however, all neutrinos are theoretically assumed to be massless.[1] The particles in the first generation do not decay and all ordinary matter in the universe is made of them. The higher generation charged particles decay with very short lifetime and they might be observed only with very high energy scenarios, e.g. in collider experiments.

Gauge bosons. As already described, there are three types of gauge boson which are defined as the force mediators. The electromagnetic force is carried by the so-called photon, γ. The gluons, g mediate the strong interactions and the charged W and neutral Z bosons are the force carriers of the weak interaction. The gauge bosons all have spin 1. The electromagnetic interaction acts on all electrically charged particles: all quarks and leptons, except neutrinos, and on the W bosons. The strong interaction acts on all color charged quarks via the eight gluons which also carry a combination of color and anticolor charge. The weak interaction acts on all quarks and all leptons via the $W^\pm$ and Z bosons. The charged

[1] In neutrino oscillations, neutrinos must have nonzero masses.

$W^{\pm}$ bosons act on left-handed particles or right-handed antiparticles only, while the neutral Z boson acts on right-handed particles, albeit with different strength. Because the W and Z bosons have high masses, they can decay into other particles as

$$W^+ \to e^+\nu_e, \mu^+\nu_\mu, \tau^+\nu_\tau, q\bar{q}'(\to \text{hadrons})$$
$$Z \to e^+e^-, \mu^+\mu^-, \tau^+\tau^-, q\bar{q}(\to \text{hadrons}),$$

and the opposite charge configuration for the W^- decay. The unification of the electromagnetic and weak interactions into the electroweak interaction is the first step of the unification of all four fundamental forces including the gravitational force that is predicted to occur in the Planck era, about 10^{-43} s after the Big Bang.

Higgs boson. The Higgs boson of the Standard Model is a massive scalar particle. It is unstable and decays almost immediately into other particles. Also, the Higgs boson has no electric charge nor color charge. It plays a master role in the theory because it is predicted as a mass generator for all other massive elementary particles. On the 4$^{\text{th}}$ of July 2012, observation of "Higgs-like particle" was announced at the LHC after 48 years of the Higgs searching. Several analyses are ongoing to confirm the existence of the Higgs boson.

2.2.2 Feynman diagrams

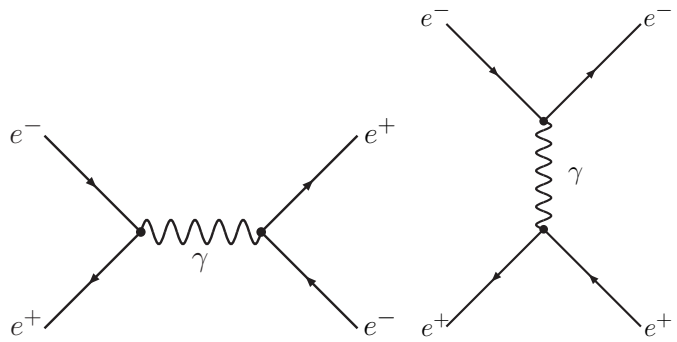

Figure 2.2: The leading order Feynman diagrams for electron-positron scattering via the electromagnetic interaction.

The Feynman diagrams introduced in perturbation theory are graphical representations of particle interactions. Figure 2.2 shows two examples of Feynman diagrams for electron-positron scattering via the electromagnetic interaction. The left diagram is interpreted as the e^+e^- annihilation process in which an electron and a positron annihilate into a virtual photon generating another e^+e^- pair. The second diagram shows a different process by rotating the first diagram by 90°. An electron interacts with a positron by emitting a photon but they do not touch each other directly; hence, the final products remain the electron and the positron.

In general Feynman diagrams are space-time diagrams. The time axis pointing to the right and the space axis pointing upward is commonly used in particle physics. Particles are represented by lines with arrows to the right, while the arrows of antiparticles point in the opposite direction. The wiggly or dashed lines without any arrow represent the gauge bosons. Observable particles are shown by only lines entering or leaving the diagram. The internal lines, which have the beginning and the ending in

the diagram represent the virtual particles. At each joining point called a vertex, all conservation laws corresponding to the interaction of each process have to be taken into account. This implies that electric charge, baryon number and lepton number must be conserved at each vertex.

In the above example of electron–positron annihilation, two electrons are represented by the upper lines with arrows to the right and two antielectrons, known as positrons, by the lower lines with arrows to the left. The photon is represented by the ended wavy line. This process happens via the electromagnetic interaction, the total electric charge of the initial state system is $0e$. Therefore, the final states are not only another electron and positron, but can also be $\mu^+\mu^-$ or another particle–antiparticle pair. Also, energy and momentum conservation have to be considered.

2.2.3 Higgs mechanism

In gauge theories, the gauge bosons are massless particles; however, the gauge bosons are in fact the massive particles of the Standard Model. A mass term in the Lagrangian of a quantum field theory is required to give the masses to the gauge bosons; nevertheless this idea destroys the gauge invariance of the Lagrangian and makes this theory meaningless because it is unrenormalisable. A solution for that is spontaneous symmetry breaking of a local gauge theory by introducing the Higgs mechanism. The Higgs mechanism allows the gauge bosons to gain mass as mathematically described below. Considering a Lagrangian:

$$L = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2, \tag{2.7}$$

with an $SU(2)$ doublet of complex scalar fields which are introduced as

$$\phi = \begin{pmatrix} \phi_\alpha \\ \phi_\beta \end{pmatrix} = \sqrt{\frac{1}{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \tag{2.8}$$

In the case of $\mu^2 < 0$ and $\lambda > 0$, the potential $V(\phi)$ as

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda^2 (\phi^\dagger \phi)^2 \tag{2.9}$$

has a non-zero minimum called the vacuum expectation value (VEV) at a value of $|\phi|$ where

$$\phi^\dagger \phi = \frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = -\frac{\mu^2}{2\lambda}. \tag{2.10}$$

This VEV leads to spontaneous symmetry breaking, since the scalar field ϕ has a non-zero ground state. The ϕ can also be expanded around a particular minimum which gives $V(\phi) = 0$, e.g. $\phi_1 = \phi_2 = \phi_4 = 0$ along with $\phi_3^2 = -\frac{\mu^2}{\lambda} \equiv v^2$. Due to local gauge invariance, the scalar field can be parametrised in terms of four real fields: $\theta_1(x)$, $\theta_2(x)$, $\theta_3(x)$ and $h(x)$ which have zero VEV:

$$\phi(x) = e^{i\tau \cdot \theta(x)/v} \begin{pmatrix} 0 \\ \frac{v+h(x)}{\sqrt{2}} \end{pmatrix}. \tag{2.11}$$

Three massless Goldstone bosons are represented by three $\theta(x)$ fields. In the process of giving mass, the $\theta(x)$ fields are gauged away from the scalar field sector to represent in the gauge field sector as longitudinal components of the $W^\pm$ and Z bosons. However, the scalar field, $h(x)$, remains, that leads to the massive scalar particle, the so-called Higgs boson. The photon cannot interact with the Higgs field due to the remaining $U(1)_{\text{em}}$ symmetry of QED; therefore, it remains massless. Since the spontaneous

symmetry breaking happens only to the $SU(2)_L \times U(1)_Y$ symmetry, the gluons which are the gauge bosons of $SU(3)_c$ also remain massless.

In a naive picture, the Higgs mechanism adds an extra scalar field, the Higgs field to the gauge theory. The Higgs field occupies all space and the gauge bosons gain mass when they couple to the Higgs field. Also, the Higgs field can fluctuate itself and form the Higgs boson. The photons and gluons do not couple to the Higgs field; hence, they are still massless. The Higgs mechanism gives mass to the fermions in a similar way as for the gauge bosons.

2.2.4 CKM quark mixing matrix

The Cabibbo–Kobayashi–Maskawa (CKM) matrix was introduced by Makoto Kobayashi and Toshihide Maskawa. They added the third generation of quarks into the matrix which was previously introduced by Nicola Cabibbo in order to describe that both a d quark and a s quark can decay into a u quark via the charged current interaction. Meanwhile, the Glashow-Iliopoulos-Maiani (GIM) mechanism also predicted the c quark as the up-type quark of the second generation. That implies that the mass eigenstates of the quarks (d, s, b) [2] are different from their weak eigenstates (d', s', b'). They are related to each other via the CKM matrix, V_{CKM}, as

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \cdot \begin{pmatrix} d \\ s \\ b \end{pmatrix}.$$

The CKM matrix is a unitary matrix. It can be parametrised by three mixing angles and one CP - violating phase and it is common to exhibit its hierarchy using the Wolfenstein parametrisation. Unitarity implies that $\Sigma_i V_{ij} V_{ik}^* = \delta_{jk}$ and $\Sigma_j V_{ij} V_{kj}^* = \delta_{ik}$. The CKM matrix elements describe the probability of a transition from one up-type quark, i, to another down-type quark, j, (or vice versa). The values of all nine CKM elements are estimated using the results of all available experiments from a global fit with theoretical constraints. The current magnitudes of the CKM matrix elements [1] are

$$V_{\mathrm{CKM}} = \begin{pmatrix} 0.97428 \pm 0.00015 & 0.2253 \pm 0.0007 & 0.00347^{+0.00016}_{-0.00012} \\ 0.2252 \pm 0.0007 & 0.97345^{+0.00015}_{-0.00016} & 0.0410^{+0.0011}_{-0.0007} \\ 0.00862^{+0.00026}_{-0.00020} & 0.0403^{+0.0011}_{-0.0007} & 0.999152^{+0.000030}_{-0.000045} \end{pmatrix}.$$

The representation of the CKM matrix shows that the probability of the coupling within the same generation is close to one as seen on the diagonal. On the other hand, the off-diagonal terms, called the CKM suppressed terms, are rather small. Quarks therefore mostly prefer to decay within the same generation. Moreover, the second most probable transition is between the first and second generations, the third most probable transition is between the second and third generations. Finally, the transition between the first and third generations has the smallest probability.

2.3 Top Quark Physics

The heaviest elementary particle observed to date is the top quark (sometimes called *truth* quark as well) which is a member of the third generation of quarks. Its mass of about 173 GeV is very high among all other quarks, roughly 60,000 times the u quark's mass and 40 times the b quark's mass. The two experiments, at the Tevatron (a $p\bar{p}$ collider at a center-of-mass energy of 1.96 TeV at Fermilab), CDF

[2] The convention is to mix down-type quarks.

and D0 discovered the top quark in 1995. Since that big discovery various properties of the top quark have been explored in order to understand the subatomic world. Furthermore, new physics beyond the Standard Model might be revealed using the top quark as a probe.

2.3.1 Top-quark properties

The third generation of quarks was predicted by Makoto Kobayashi and Toshihide Maskawa in order to explain the observed CP violations in K meson decay in 1973 [2]. The top quark is identified as the up-type quark of the third generation and its weak isospin partner the b quark. Because of its very large mass, its discovery took about 23 years after its prediction; the b quark was observed already in 1977. At the LHC, the pp collider at a center-of-mass energy of 7 TeV was successfully operated in 2010 and first top-quark measurements were also achieved in that year. In 2011, single top-quark production via the weak interaction was first measured. The LHC is often called a top factory, allowing many top-quark physics studies. In the following, the current properties of the top quark will be briefly summarised.

The most recent direct measurement of the top-quark mass from Tevatron is 173.5 ± 0.6 (stat.) ± 0.8 (syst.) GeV (with a precision of 0.5%). Using only the indirect constraints on the top-quark mass from the recent precision measurements of the parameters of the electroweak theory indicate that the top quark mass is $179.7^{+11.7}_{-8.7}$ GeV [3] which is in very good agreement with the direct measurements. Corresponding to the weak isospin doublet, the top quark has $T_3 = +1/2$, and a charge, Q, of $+2/3\ e$.

In the Standard Model, top quarks decay almost 100% into a b quark and a W boson according to the CKM matrix element $|V_{tb}| \gg |V_{ts}|$ and $|V_{td}|$. Considering only this decay mode and using the top-quark mass of 172.5 GeV, the predicted width of the top quark is $\Gamma_t = 1.33$ GeV. This width corresponds to a very short lifetime, $\tau_t = 1/\Gamma_t \sim 5 \times 10^{-25}$ s. Its observed width from the t-channel single top-quark cross-section and the branching fraction of $t \to Wb$ measurements yields $\Gamma_t = 2.00^{+0.47}_{-0.43}$ GeV [4] corresponding to $\tau_t \sim 3.29^{+0.90}_{-0.63} \times 10^{-25}$ s, in good agreement with the Standard Model prediction. Because the lifetime of the top quark is smaller than the typical time to form hadrons, $\tau_{\text{had}} \sim 1$ fm/$c \sim 3 \times 10^{-24}$ s, the top quarks decay before they can form bound states. Another consequence of its very short lifetime is that the spin of the top-quark decay is not depolarised due to hadron formation. Thus its spin properties are able to be inferred from the spin information of its decay products. A possible difference in mass between the top and antitop quarks could be measured directly, since the top quarks have a very short lifetime and they decay before they are hadronised. The recent result is Δm_t = -0.44 ± 0.46 (stat.) ± 0.27 (syst.) GeV [5]. In principle, a mass difference between a quark and its corresponding antiquark would imply CPT symmetry violation.

2.3.2 Top-quark production at the LHC

At hadron colliders top quarks are produced via two mechanisms. The production of top–antitop ($t\bar{t}$) pairs via the strong interaction is the dominant source of top quarks at both the Tevatron and the LHC. Tens of thousands of $t\bar{t}$ pairs were created at the Tevetron. Millions of $t\bar{t}$ pairs have been produced and part of them has been studied at the LHC. Another top-quark production mechanism is single top-quark production via the weak interaction. The cross section of single top-quark production is small compared to that of $t\bar{t}$ production but this source is rather important for the electroweak theory studies.

Top-quark pair production

Partons are the point-like constituents of hadrons, i.e. quarks, antiquarks and gluons. Differently from the naive parton model picture, protons do not only consist of two up quarks and one down quark, called

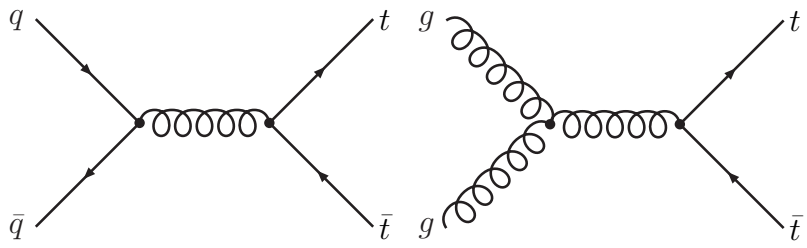

Figure 2.3: Examples of Feynman diagrams for $t\bar{t}$ production at leading order: $q\bar{q}$ annihilation (left) and gg fusion (right).

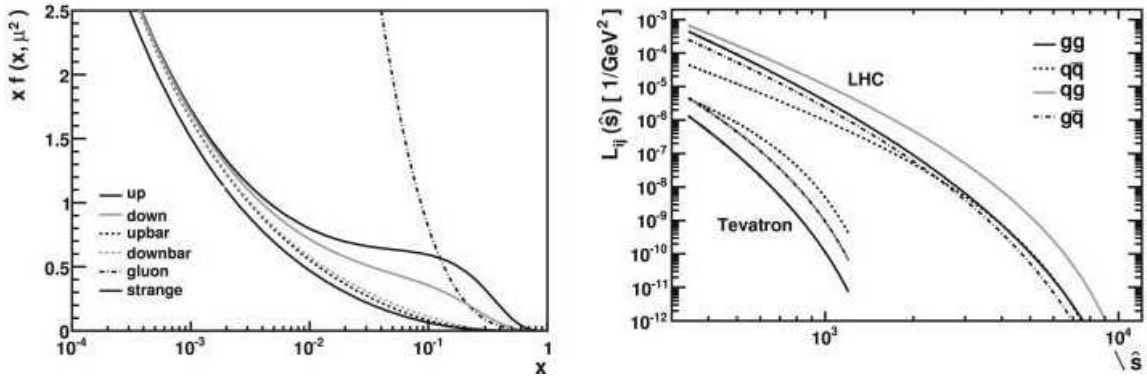

Figure 2.4: PDFs of quarks and gluons inside the proton from the CTEQ collaboration [6] (left). Parton luminosities for gg, $q\bar{q}$, qg and $g\bar{q}$ interactions at the Tevatron and the LHC [7] (right).

valence quarks, but also quark–antiquark pairs from gluons which are the so-called sea quarks. In fact, the interaction from the pp or $p\bar{p}$ collisions is the hard scattering process of the partons inside protons. The proton structure is described by the parton distribution function (PDF) which depends on the scale Q^2, where $Q^2 = -q^2$, q is the four-momentum of the mediator.

The top-quark mass is much higher than the QCD scale, $\Lambda_{QCD} = 217\pm25$ MeV [8]; hence, the cross section of $t\bar{t}$ pairs can be determined using the QCD-improved parton model [9]. The factorisation theorem says that the inclusive cross section of the process $pp \to t\bar{t}$ at the LHC in terms of parton distribution functions (PDF), f_i and a partonic cross section, $\hat{\sigma}$, is

$$\sigma_{pp\to t\bar{t}}(s, m_t) = \sum_{i,j=q,\bar{q},g} \int dx_i dx_j f_i(x_i, \mu_f^2) f_j(x_j, \mu_f^2) \cdot \hat{\sigma}_{ij\to t\bar{t}}(\hat{s}, m_t, \mu_f, \mu_r, \alpha_s). \tag{2.12}$$

The cross section depends on the square of the center-of-mass energy of the collider, $s = 4E^2_{\text{beam}}$ and m_t. x_i are the parton momentum fractions, $\mu_{f(r)}$ are the factorisation and renormalisation scales and $\hat{s} \sim x_i x_j s$ is the center-of-mass energy of the partons. The PDFs $f_i(x_i, \mu_f^2)$ are generally determined from global fits to experimental data of deep-inelastic scattering. The typical momentum fractions for both pp and $p\bar{p}$ colliders can be estimated from $x \approx 2m_t/\sqrt{s}$ due to two assumptions: the need of the

minimal energy, $\sqrt{\hat{s}} \geq 2m_t$, and setting $x_i \approx x_j \approx x$.

At leading order QCD, the top-quark pairs are produced by quark–antiquark annihilation, $q\bar{q} \rightarrow t\bar{t}$ and gluon-gluon fusion, $gg \rightarrow t\bar{t}$ shown in figure 2.3. The PDFs of u, d, $\bar{u}$, $\bar{d}$, s quarks and gluons inside the proton are shown in figure 2.4 (left). In the x region less than 0.15, gluons are dominant. Figure 2.4 (right) shows the parton luminosity distributions at the Tevatron and the LHC. The parton luminosity is used to compare the same process at different collider setups and it is defined as:

$$L_{ij\rightarrow t\bar{t}}(\hat{s}; s, \mu_f) = \frac{1}{s} \int \frac{dx_i}{x_i} f_i(x_i, \mu_f^2) f_j(\frac{\hat{s}}{x_i s}, \mu_f^2). \tag{2.13}$$

At the Tevatron, the beam energy is about 10 times the top quark mass. That leads to the large corresponding x value, $x \approx 0.18$ with $\sqrt{s} = 1.96$ TeV. The quark–antiquark annihilation contributes about 85% of the total cross section there and the remainder is mostly gg fusion. While at the center-of-mass energy of $\sqrt{s} = 7$ (14) TeV, the corresponding x of 0.05 (0.025), the lower x region becomes important, around 80% (90%) of the top quark pair production is therefore via gluon-gluon interactions at the LHC. The rest is mostly via $q\bar{q}$ annihilation. The qg and $g\bar{q}$ initial states are generally suppressed by the coupling constant, α_s, thus the cross section from their contributions is only a few percent at both the Tevatron and the LHC. At the LHC the predicted cross section of the top quark pair production at $\sqrt{s} = 7$ TeV is 165^{+11}_{-16} pb assuming $m_t = 172.5$ GeV [1].

Single top-quark production

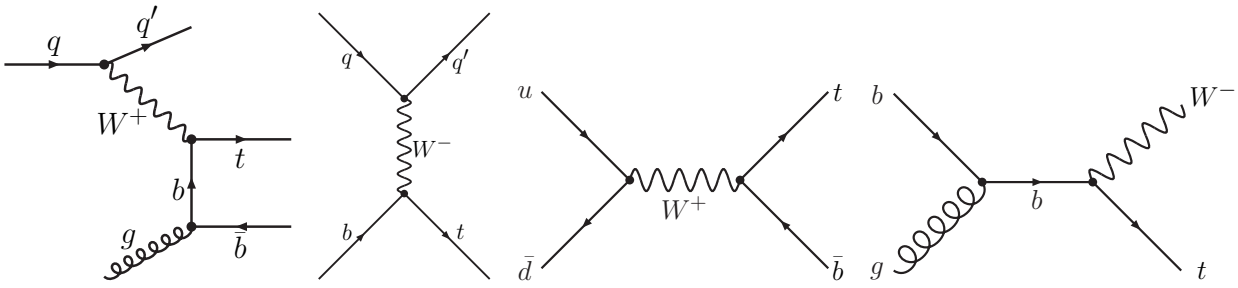

Figure 2.5: Examples of Feynman diagrams for single top-quark production at leading order. From left to right: the t-channel as Wg fusion and as flavor excitation, the s-channel and the Wt-channel.

Single top-quarks can be produced via the electroweak interaction, which involves almost exclusively the Wtb vertex, since the production involving Wtd and Wts vertices is strongly suppressed as described above. There are three different production modes: the t-channel, as Wg fusion and flavor excitation, the s-channel and the Wt-channel shown in figure 2.5. They are distinguished according to the virtuality, $-q^2$, of the produced W boson, where q is the four-momentum of the W boson. Table 2.2 shows theoretical cross section predictions for all three single top-quark production channels at the Tevatron and the LHC. The predicted cross sections for the Tevatron and the LHC at $\sqrt{s} = 14$ TeV are taken from [10]. The calculations assumed $m_t = 175$ GeV. The first uncertainty corresponds to the scale uncertainty and the second one is the PDF uncertainty. The predicted values for the LHC at $\sqrt{s} = 7$ TeV are from [11] using $m_t = 173$ GeV. Their error is from both scale and PDF uncertainties.

Production mode	$\sqrt{s}$ [TeV]	σ_t [pb]	σ_s [pb]	σ_{Wt} [pb]
$p\bar{p} \to t/\bar{t}$	1.96	$1.98^{+0.28}_{-0.22}$	1.02 ± 0.08	0.25 ± 0.03
$pp \to t$	7	$41.7^{+1.6}_{-0.2} \pm 0.8$	$3.17 \pm 0.06^{+0.13}_{-0.10}$	$7.8 \pm 0.2^{+0.5}_{-0.6}$
$pp \to \bar{t}$	7	$22.5 \pm 0.5^{+0.7}_{-0.9}$	$1.42 \pm 0.01^{+0.06}_{-0.07}$	$7.8 \pm 0.2^{+0.5}_{-0.6}$
$pp \to t$	14	156 ± 8	$7.23^{+0.55}_{-0.47}$	41.1 ± 4.2
$pp \to \bar{t}$	14	91 ± 5	$4.03^{+0.14}_{-0.16}$	41.1 ± 4.2

Table 2.2: Predicted total cross sections for three single top quark production modes at the Tevatron with $\sqrt{s}$ = 1.96 TeV and the LHC $\sqrt{s}$ = 7 (14) TeV. The values are the sum of t quark and $\bar{t}$ quark production for the Tevatron, while the values are separated for the LHC

At the LHC with $\sqrt{s}$ = 7 TeV, the total cross section of single top-quark production is about 84.4 pb which is approximately 50% of the top-quark pair production cross section. The mass of the top quark is of the same order of magnitude as the scale of the electroweak symmetry breaking. Single top-quark production is therefore important as a probe to the electroweak theory together with the discovery of new physics. For example, single top-quark production in the t-channel via $ug \to t$ allows to search for flavor-changing neutral currents (FCNC).

***t*-channel production.** The t-channel is the dominant mode for single top-quark production at both the Tevatron (about 61%) and the LHC (about 76%). In the t-channel a virtual W boson scatters off a sea b quark inside the proton, and gives a t quark in the final state. The b quark originates from a gluon which splits into a $b\bar{b}$ pair as shown in the first graph of figure 2.5. This scheme is known as W-gluon fusion. A b quark in the proton can also be considered as the leading order shown in the second graph of figure 2.5. Basically the Wg fusion is therefore a next-to-leading-order process which is important for the cross section calculations. This mode, $q + b \to q' + t$, is called flavor excitation. Because of its large cross section and its clean signature, the first study of the t-channel production was done at the Tevatron in 2009 [12] and several measurements of the cross section of the t-channel production have been performed since 2011 at the LHC [13].

***s*-channel production.** A time-like W boson is produced by a quark and an antiquark that belong to the same isospin doublet, mostly $u\bar{d}$, and further decays into $t\bar{b}$. Some of the leading order processes of the s-channel production have the same initial and final states as the Wg fusion of the t-channel; however, one can distinguish them via their color structure which does not allow them to interfere with each other. The $t\bar{b}$ pair in the s-channel forms a color singlet, since it is from a W boson, while the $t\bar{b}$ pair in the t-channel is a color octet state since it comes from a gluon. At the Tevatron, the predicted cross section of the s-channel production is about 31% of the total single top-quark cross section; on the other hand, it is only 5% at the LHC with $\sqrt{s}$ = 7 TeV. One reason for that is the $q\bar{q}'$ annihilation process does not play an important role at the LHC.

At the Tevatron, the combined cross section of the t-channel and s-channel was measured to be 3.9 ± 0.9 pb [14], in good agreement with the predicted value of 2.9 ± 0.4 pb. Separating s-channel production from t-channel production is the main effort. At the LHC, the cross-section measurement of single top-quark production in the s channel is challenging since the cross section of the s-channel is the smallest.

***Wt*-channel production.** In the *Wt*-channel, a t quark is produced in association with an on-shell (or close to on-shell) W boson. This mode is also known as associated production of the single top-quarks. At leading order, the W boson and the t quark are emitted from a virtual b quark shown in figure 2.5 (right) or t-channel-like as $bg \to Wt$ via a virtual t quark. Considering the first case but with an initial gg pair as shown in figure 2.6, this is an example of *Wt*-channel production at next-to-leading order. The final state corresponds to $t\bar{b}W$ which is the same as the final state of the $t\bar{t}$ production at the leading order but with different kinematics.

At the LHC, single top-quark production in the *Wt*-channel is the second most important production channel (about 19%) and expected to be observable there. On the other hand, its cross-section is too small by far to be observed at the Tevatron.

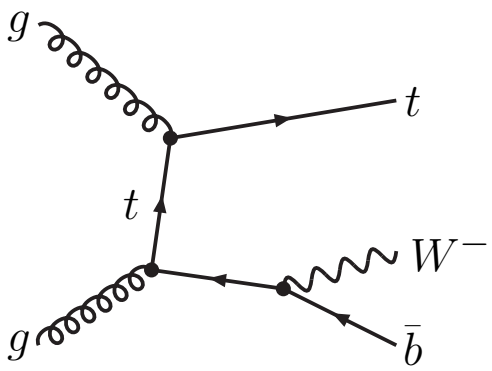

Figure 2.6: Feynman diagram of *Wt*-channel production at next-to-leading order.

2.3.3 Top-quark decay

The decays $t \to Wd$ and $t \to Ws$ are strongly suppressed and normally negligible. Considering only $t \to Wb$, the decay rate of the top quark in the Standard Model at next-to-leading order QCD is given by [7]

$$\Gamma_t = \frac{G_F m_t^3}{8\pi\sqrt{2}} |V_{tb}|^2 \left(1 - \frac{M_W^2}{m_t^2}\right)^2 \left(1 + 2\frac{M_W^2}{m_t^2}\right) \left[1 - \frac{2\alpha_s}{3\pi}\left(\frac{2\pi^2}{3} - \frac{5}{2}\right)\right], \tag{2.14}$$

where G_F is the Fermi constant and α_s the strong coupling constant. As discussed in Section 2.3.1, $\Gamma_t = 1.33$ GeV at $m_t = 172.5$ GeV.

The ratio of the branching fractions of the top quark decaying to a b quark and to any down-type quark is

$$R = \frac{\mathcal{B}(t \to Wb)}{\mathcal{B}(t \to Wq)} = \frac{|V_{tb}|^2}{|V_{tb}|^2 + |V_{ts}|^2 + |V_{td}|^2}. \tag{2.15}$$

Assuming unitarity of the CKM matrix, $\sum_{q_{\text{down-type}}} |V_{tq}|^2 = 1$, a measurement of R directly indicates the CKM matrix element $|V_{tb}|$. Furthermore, the only direct determination of $|V_{tb}|$ without assuming unitarity is from single top-quark production cross sections. The cross section measurements by CDF

and D0 of $2.76^{+0.58}_{-0.47}$ pb [12] imply that the absolute value of the element is $|V_{tb}| = 0.88 \pm 0.07$.

In general, the decay modes of the t quark can be classified by the decay modes of the W boson. The W boson decays into a quark–antiquark pair with the branching ratio $\mathcal{B} = 68\%$ or into a lepton and its corresponding neutrino with $\mathcal{B} = 32\%$. The first case is normally known as hadronic decay and the latter as leptonic decay. There are three cases for the decay of systems which have two W bosons, e.g. the Wt-channel and $t\bar{t}$ production:

- all-hadronic channel: both W bosons decay hadronically, with $\mathcal{B} = 46.2\%$,
- lepton+jets channel: one W boson decays hadronically while the other decays leptonically, with $\mathcal{B} = 43.5\%$,
- dilepton channel: both W bosons decay leptonically, with $\mathcal{B} = 10.3\%$.

A term "jet" is referred to as a hadron cluster produced by the hadronisation of a quark or gluon. Experimentally, the all-hadronic channel, which contains only jets in the final state, is hard to separate from the multijet background that will be discussed in Section 4.2.7. This mode is therefore difficult to isolate in analyses even if it has the largest branching ratio. On the other hand, the dilepton channel is the cleanest, but there are two problems in this channel: the smaller branching ratio and large missing transverse momentum due to two neutrinos, that one cannot reconstruct directly. The lepton+jets channel has the second largest branching ratio and separation of the signal from the background in this channel is easier than in the hadronic channel; therefore, it is a promising channel to have a look at for the Wt-channel production.

Chapter 3

The ATLAS Experiment

3.1 The Large Hadron Collider (LHC)

The most powerful particle accelerator to date is the LHC at CERN, the European Organization for Nuclear Research, located between the borders of France and Switzerland, near Geneva. The LHC beam pipe has a circumference of about 27 km and lies about 100 m underground. Two beams of hadrons are accelerated in opposite directions, then collide with each other. The LHC is expected to reach a center-of-mass energy of 14 TeV and a design luminosity of 10^{34} cm^{-2} s^{-1} in proton-proton collisions with 2808 proton bunches of $1.15 \cdot 10^{11}$ protons per beam and a bunch spacing of 25 ns. The aim is to take 100 fb^{-1} integrated luminosity of data per year. Products of proton-proton collisions are mainly explored in order to verify our current knowledge and clarify several remaining puzzles about the universe. Moreover, the LHC has a heavy-ion programme to study matter in the conditions that existed in 10^{-6} s after the Big Bang. The successful operation in 2010 was at a proton-proton center-of-mass energy of 7 TeV with an instantaneous luminosity of $2 \cdot 10^{32}$ cm^{-2} s^{-1}. In 2011, an instantaneous luminosity of up to $3.65 \cdot 10^{33}$ cm^{-2} s^{-1} at a bunch spacing of 50 ns was reached. An integrated luminosity of the full 2011 proton-proton data amounts to 5 fb^{-1}. This 2011 data was used in this thesis.

Figure 3.1 shows a schematic view of CERN's accelerator complex. Protons from a hydrogen-gas source are accelerated by radio-frequency (RF) cavities. The protons are firstly injected into an RF cavity accelerating them up to 750 keV, and then into the Linear Accelerator called LINAC2 to 50 MeV. In order to achieve an energy of 25 GeV, the protons are sent to the Proton Synchrotron (PS) by the PS Booster. Finally, before injecting them to the LHC ring, the protons are accelerated by the Super Proton Synchrotron (SPS) to 450 GeV. 1232 superconducting dipole magnets of 15 m length are used to bend the beams and up to 392 quadrupole magnets, each 5–7 m long, focus the beams. The dipoles must produce fields of 8.36 T in order to bend the beams. Superfluid helium is used to cool down temperatures of the dipole magnets to 1.9 K to achieve such a high field strength.

There are six detectors installed at the LHC's collision points. ATLAS and CMS are general purpose detectors to investigate particles produced by the proton-proton collisions so that they serve a wide spectrum of experiments. They also aim for fulfilling a gap in the Standard Model that is the searches for the Higgs boson. These two experiments are therefore on the lookout for predicted phenomena beyond the Standard Model as well as completely new scenarios in physics. The ATLAS and CMS detectors were designed independently in order to cross-check any result. LHCb is a medium-sized detector. It aims to understand why the dominant substances in the universe are particles rather than their corresponding antiparticles using b quarks. The detector is designed specially as a series of sub-detectors to catch the b quarks mainly in the forward direction. Another medium-sized detector is ALICE. The ALICE experiment focuses on the understanding of Pb-Pb collisions. These collisions are

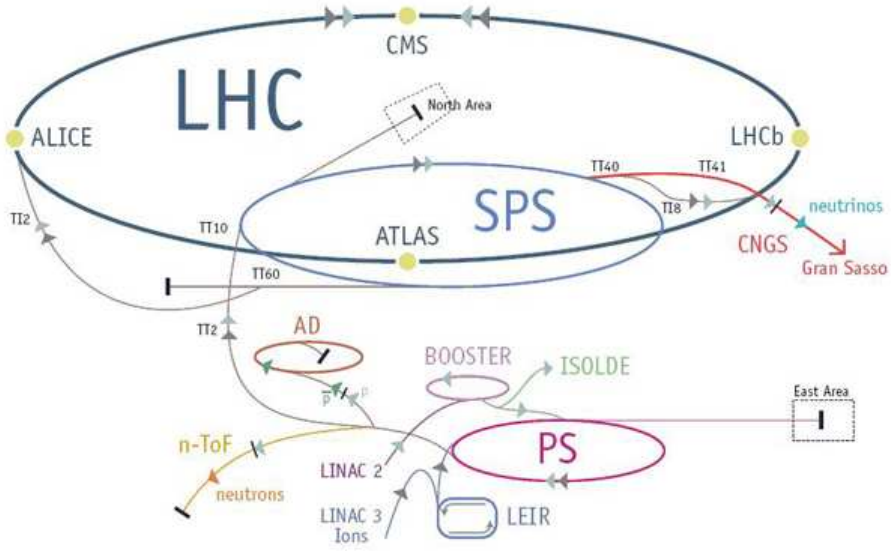

Figure 3.1: CERN's accelerator chain [15].

supposed to recreate a state of matter which existed just after the Big Bang, namely the quark-gluon plasma.

Besides the four main detectors of the LHC, there are two further experiments, TOTEM and LHCf, which are much smaller in size and weight. The detectors for the TOTEM experiment are placed near the CMS detector, those for the LHCf experiment are near the ATLAS detector. The TOTEM experiment concentrates studying on the proton structure in depth as well as the precise measurement of the proton-proton interaction cross section. Cosmic rays have been studied in the laboratory conditions by the LHCf. Both of them explore mainly the particles produced in the forward region of collisions. A seventh experiment at the LHC, named as MoEDAL, is designed to search for magnetic monopoles together with highly ionizing stable massive particles using nuclear track detectors.

3.2 ATLAS - A Toroidal LHC ApparatuS

ATLAS is one of the two multi-purpose particle detectors. The ATLAS detector, with the huge size of 46 m length, 25 m diameter and about 7,000 ton weight, is the largest particle detector ever built. Many particle physics experiments have been performed within the framework of the ATLAS collaboration. About 3,000 physicists at 175 institutions in 38 countries around the world participate in the project. Before some details of all components of the ATLAS detector are described, some kinematic parameters used at ATLAS will be given.

3.2.1 Kinematics at ATLAS

The standard coordinate system used at ATLAS is a right-handed (x, y, z) coordinate system. At the interaction point in the center of the ATLAS detector, the z-axis is parallel to the beam in counter-clockwise direction. The x-axis points towards the centre of the LHC ring. The y-axis points upwards.

The (r, θ, ϕ) coordinate system is also used to describe detector positions. The radial distance from the beam axis, r, is defined as

$$r = \sqrt{x^2 + y^2}. \tag{3.1}$$

The polar angle, θ, is defined as the angle between the particle direction and the beam axis and the azimuthal angle ϕ is defined as the angle in the x - y plane. The pseudorapidity, η, is defined using the polar angle, θ, as

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right). \tag{3.2}$$

The pseudorapidity is an important quantity of a particle and simply used instead of the rapidity since it depends upon only the polar angle θ. The rapidity, y, is defined as

$$y = \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right). \tag{3.3}$$

This parameter is useful since a difference in the rapidity of two particles is invariant under Lorentz boosts along the beam axis. The distance ΔR given in η - ϕ space is defined as

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \tag{3.4}$$

At hadron colliders, momenta are usually expressed in terms of transverse momentum, p_T, and η. The absolute value of momentum is defined as

$$|p| = p_T \cosh\eta, \qquad \text{with} \tag{3.5}$$

$$p_x = p_T \cos\phi,\ p_y = p_T \sin\phi,\ p_z = p_T \sinh\eta. \tag{3.6}$$

Due to the high center-of-mass energy of the pp collisions, one often works in the high energy limit, where the momentum is approximately the same as the energy. Therefore, the transverse energy, E_T, is calculated. Moreover, it is also possible to determine the other transverse quantities for example the missing transverse energy, E_T^{miss}.

3.2.2 ATLAS detector

The detector is structured in multiple layers of the individual components surrounding the interaction regime. The ATLAS detector is illustrated in figure 3.2. Each component is built up from different materials in such a way that corresponding particles interact with the material and deposit their energies during interactions as shown in figure 3.3. The detector comprises three major components: the inner detector, the calorimeter and the muon spectrometer. The following information gives a short overview of the ATLAS detector as well as the trigger and data acquisition system. The following details are mainly based on the ATLAS performance report [17].

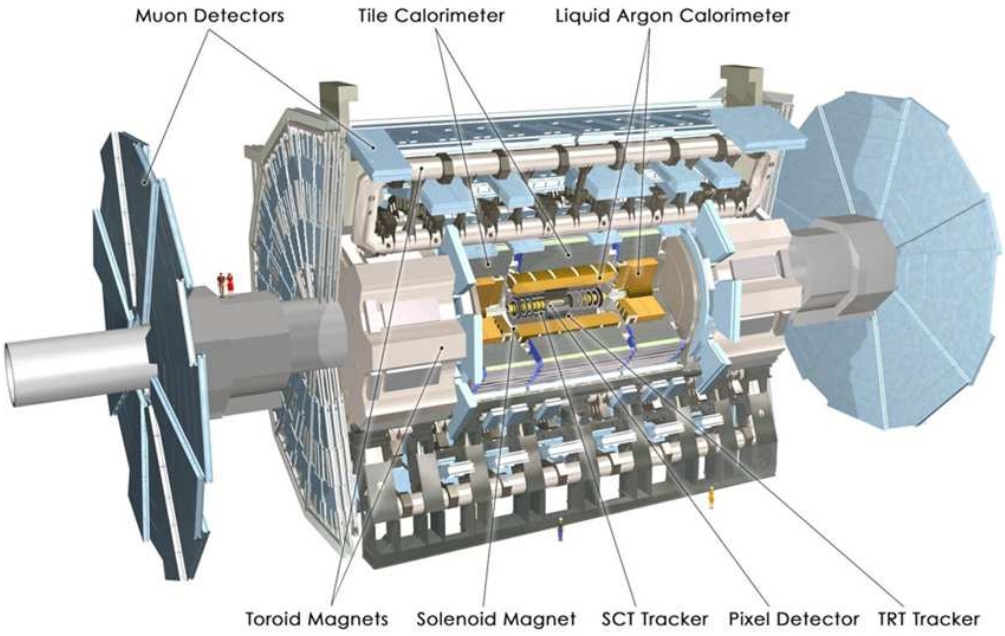

Figure 3.2: An overview of all components of the ATLAS detector [16].

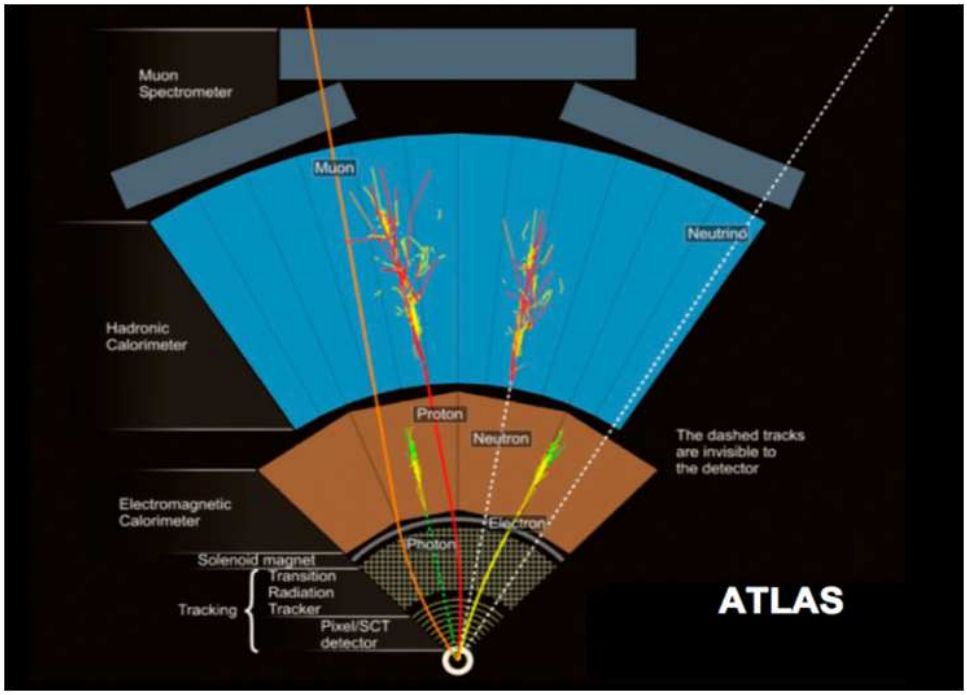

Figure 3.3: Interactions of particles in different detector components of the ATLAS detector.

Inner Detector

The ATLAS Inner Detector (ID) is the central tracking system of the ATLAS detector, that surrounds the LHC interaction point region. The ID comprises three main components, the Pixel Detector, the Semiconductor Tracker (SCT) and the Transition Radiation Tracker (TRT). All are contained in a cylinder with 1.15 m diameter and 7 m length, surrounded by a solenoid magnet which provides a field of 2 T. The ID covers a pseudorapidity of $|\eta| < 2.5$ with full coverage in ϕ. The barrel region of each detector is arranged in concentric cylinders around the beam axis while the end-cap regions are positioned on disks which are perpendicular to the beam axis. The detector provides information of the primary vertex together with secondary vertex reconstruction, transverse momentum measurement, track reconstruction and charge determination.

The Pixel Detector. The Pixel detector is the innermost component, with a radial range between 50.5 mm and 150 mm. It consists of 1,744 silicon pixel modules with 47,232 pixels each. Each pixel has an area of 50 μm × 400 μm. Only 46,080 pixels are read out by 16 radiation-hard front-end chips. The total number of readout channels therefore is 67 million pixels in the barrel region and 13 million pixels in the end-caps. The pixel modules are arranged in three concentric barrel layers and three end-cap disks each. The Pixel detector is able to provide information on short lived particles such as B-hadrons, determine the impact parameter using three precise measurements per track and find vertices.

The Semiconductor Tracker. The Semiconductor Tracker covers the intermediate radial range between 299 mm and 560 mm. The SCT provides eight precision measurements per track. There are 8448 silicon modules in the barrel and 6,944 modules in the forward region. The modules are arranged in four concentric layers in the barrel region as well as nine disks on each end-cap. Each of the silicon detectors consists of 768 readout strips of 80 μm pitch in the barrel while of variable pitch in the end-caps. The strips of approximately 12 cm in length are roughly parallel to the beam direction in the barrel region and radial in the forward region.

The Transition Radiation Tracker. The Transition Radiation Tracker consists of thin drift tubes filled with a gas mixture of 70% Xe, 27% CO_2 and 3% O_2. Each straw tube of 4 mm diameter is formed from two layers of multilayer films of 35 μm thickness. The barrel of the TRT is arranged into cylindrical layers covering the radial range from 56 cm to 107 cm and the end-caps are arranged in wheel-like structures. In the barrel region there are 73 layers of straws which are parallel to the beam pipe and have 144 cm in length. In the end-cap regions, 160 straw planes are radially arranged into wheels with 37 cm in length. The total number of readout channels is 420,000. The TRT provides electron identification using transition radiation photons which are created in polypropylene fibres (barrel) or foils (end-caps) between the straws. Charged particles with transverse momentum $p_T > 0.5$ GeV and pseudorapidity $|\eta| < 2.0$ are also expected to cross more than 30 straws in the TRT. At the outer radius the momentum measurements can be complemented via the straw hits with lower precision but longer measured track length.

Calorimeter

The ATLAS calorimeter system is located outside the solenoid magnet surrounding the inner detector. It is divided into two sub-systems, namely the electromagnetic (EM) calorimeter and the hadron calorimeter. The calorimeters are designed to measure the energies of particles by absorbing them. In addition, they contribute to particle identification and missing energy determination. The components

of the calorimeter system are metal absorbers and active layers. Incoming particles interact with absorbers and deposit their energies to generate particle showers which are detected by the active layers. Liquid argon (LAr) is filled in gaps throughout the calorimeter system. The EM calorimeter is designed to measure electromagnetic showers from electrons and photons while the hadron calorimeter measures hadronic showers from jets.

The Electromagnetic Calorimeter. The Electromagnetic Calorimeter is the closest part of the system to the beam pipe. It is designed to detect energy deposits of electrons and photons as electromagnetic showers in a large energy range between 5 GeV up to 5 TeV. The showers are produced by for example pair production, Compton scattering and bremsstrahlung processes. The EM calorimeter is divided into a barrel region and two end-cap regions. The barrel lies inside the barrel cryostat with a radial range between 1.5 m and 2 m and covers the range $|\eta| < 1.475$. Each end-cap calorimeter lies inside the end-cap cryostat on each side of the barrel and they cover the region $1.375 < |\eta| < 3.2$. However, precision measurements are possible only within $|\eta| < 2.5$. Because of its accordion structure with granularities, it can cover all ϕ-angles without any azimuthal cracks. The EM calorimeter uses a lead-liquid argon sampling technique. Lead plates bent in an accordion shape are applied as the absorbers and the LAr as active material. There are three layers made of copper-polymide between two absorbers, that provides the drift field due to high voltage fed to the outer plates and performs the capacitive readout on the inner layer. The thickness of the EM calorimeter is larger than 22 radiation lengths (X_0) in the barrel and larger than 24 X_0 in the end-caps.

A presampler detector is used to measure the energy corresponding to the low momentum particles which are created in the interaction of electrons/positrons and photons. The presampler detector was put in front of the barrel calorimeter over the full η range of the barrel. In the transition region between the barrel and end-cap calorimeters, $1.5 < |\eta| < 1.8$, a presampler detector is also used in order to improve the energy measurement in this range. The EM calorimeter is divided in depth into three longitudinal layers being in accord with different $\Delta\eta \times \Delta\phi$ granularities. The energy resolution depending on the particle type and position of $|\eta|$ can be expressed as [18]

$$\frac{\sigma(E)}{E} = \frac{a}{E} \oplus \frac{b}{\sqrt{E}} \oplus c. \tag{3.7}$$

The first term on the right-hand side of the equation is referred to as the noise term which is always ignored from the calculation. The second term is the sampling term and the last term is a constant. The energy resolution of electrons is expected to reach

$$\frac{\sigma(E)}{E} = \frac{10\%}{\sqrt{E}} \oplus 0.7\%. \tag{3.8}$$

The Hadron Calorimeter. The Hadron Calorimeter takes a major responsibility for jet reconstruction. The hadron calorimeter consists of the tile calorimeter, the end-cap calorimeter and the forward calorimeter. The tile calorimeter encloses the EM calorimeter. Steel is used as the absorber and scintillator as the active medium. It has a central barrel part together with two extensions covering the region $|\eta| < 1.7$. In the end-cap regions, the calorimeters extend the range of $1.5 < |\eta| < 3.2$. For these calorimeter wheels, copper plates are used as the absorbers and LAr is the active medium. The end-cap cryostats are shared between the EM and hadronic end-cap calorimeters as well as the forward calorimeter. The forward calorimeter covers the range $3.1 < |\eta| < 4.9$. It is split into one EM module

made of copper and two hadron modules made of tungsten. The active medium is still LAr which is used to fill the gaps.

The hadronic showers develop as a sequence of inelastic hadronic interactions. There are many produced particles with large momentum which can also create tertiary particles and so forth. The hadronic showers are therefore more complex and much larger than the electromagnetic showers. That also directly effects the energy resolution of the hadronic calorimeter, being worse than that of the EM calorimeter. For hadrons, the resolution is expected to be

$$\frac{\sigma(E)}{E} = \frac{100\%}{\sqrt{E}} \oplus 10\%. \tag{3.9}$$

Muon Spectrometer

The muon spectrometer is designed to measure muon momenta with high precision over a large range in pseudorapidity and azimuthal angle. The muon spectrometer is the outermost part of the ATLAS detector using a special magnet system along with high precision tracking and trigger chambers. The magnet system of the muon spectrometer consists of three air-core superconducting toroids generating large magnetic fields in order to bend the tracks of muons. Its aim is to measure a transverse momentum range of 10 GeV $< p_{\mathrm{T}} <$ 500 GeV with a resolution between 3% and 4% and p_{T} up to 1 TeV with a resolution better than 10%.

The barrel toroid spreads over 25 m length with a radius range between 4.7 m and 10 m. At the ends of the barrel toroid, there are two end-cap toroids with a radius range between 1.65 m to 10.7 m. Each toroid consists of eight superconducting coils which provide an average field of 0.5 T. The barrel toroid covering the range $|\eta| < 1.4$ provides 1.5 to 5.5 Tm of bending power while the end-cap toroids covering the range $1.6 < |\eta| < 2.7$ provide 1 to 7.5 Tm of bending power. In the transition regions, $1.4 < |\eta| < 1.6$, the bending power is less.

To achieve a high precision measurement of the muon trajectory, Monitored Drift Tubes (MDTs) are used. The MDT chambers contains three to eight layers of drift tubes which have a diameter of roughly 30 mm and are filled with 93% Ar together with 7% CO_2 gas at 3 bar. The total number of the tubes is between 16 and 72 per layer of a chamber. The electrons produced by ionisation of the gas are collected on a tungsten-rhenium wire at high voltage. The MDT aims to reach a resolution of 60 to 80 μm. They are installed into three layers in the barrel region. Resistive Plate Chambers (RPCs) inserted parallel to the electron-plate detectors filled with gas are also installed in the last two layers. The RPCs are for the muon trigger.

In the end-cap regions, there are one layer of Cathode Strip Chambers (CSCs) and two layers of MDTs. The first layer located closest to the beam pipe is equipped with CSCs instead of MDTs because the CSCs perform much better in this high background-rate region. The limit of MDT operation is 150 Hz cm^{-2} but the limit of CSC operation is up to 1000 Hz cm^{-2}. The CSCs are multi-wire proportional chambers with cathode planes. The resolution of a chamber is 40 μm in the bending plane and 5 mm in the transverse plane. Thin Gap Chambers (TGCs) are also installed into the first two layers. They operate with the same principle as the CSCs but with a high electric field and different characteristics leading to high precision of the time resolution. The TGCs are responsible for the muon trigger in this region.

As described above, the trigger system of the muon spectrometer consists of the RPC and TGC detectors. It covers a range of $|\eta| < 2.4$. It provides a high precision timing resolution of 1.5 ns to 4 ns to trigger on muons. Measurements of muon momenta can be done with the trigger system. Moreover, the

muon trigger detectors can identify a bunch-crossing and measure the muon coordinate in the direction orthogonal to that provided by the precision tracking system.

Trigger and Data Acquisition System

The ATLAS trigger system was made to reduce the event storage rate without losing many interesting physics events. Selection criteria and Regions-of-Interest (RoI) are strategies for this purpose. The trigger system consists of three levels of event selection: Level 1 (L1), Level 2 (L2) and Event Filter (EF). The L1 trigger is hardware-based, while the L2 and EF triggers are software-based.

The Level 1 (L1) trigger is responsible for accessing data from the calorimeter and muon spectrometer detectors. It searches for signatures of muons with high p_{T}, electrons, photons, hadronic decays of tau leptons, jets and events with large missing transverse energy as well as large transverse energy. The calorimeter trigger decision is based on the observed clusters' multiplicities and energy thresholds in the calorimeter sub-detectors, and the muon trigger decision on trajectory measurements in the muon trigger detectors. To operate at a maximum pass rate of 75 kHz, the L1 trigger must make a decision within 2 μs.

The Level 2 (L2) trigger uses the RoI information from the L1 trigger and then applies the L2 to refine remaining information. The L2 decisions are implemented in the form of selection algorithms, that allow the use of information from all detectors, including the inner detector, which is not available at L1. The L2 trigger can reduce the event rate to about 3.5 kHz and each event has the processing time of roughly 40 ms.

The events selected by the L2 trigger are passed on to the Event Filter (EF) trigger which is based on the complex offline analysis algorithms to fully reconstruct the events. The EF trigger reduces the event rate to roughly 100 Hz with the processing time of each event of roughly 4 s. That event rate corresponds to about 100 Mbyte per second. The events in this stage are classified into ATLAS physics data streams which are muons, electrons, photons, taus, jets and B-physics. However, one event can fulfil several data stream classification definitions; therefore, it can be recorded in more than one data stream.

Moreover, the ATLAS detectors also monitor and measure the relative and absolute luminosity. The luminosity can be independently determined using various detectors and algorithms. After the calibration using dedicated beam-separation scans, the luminosity is determined with an uncertainty of 3.7% [19] for the 2011 data.

Chapter 4

Signal and Background Processes

As described in Chapter 2, single top-quark production in the *Wt*-channel has not been measured so far. Measurements of single top-quark production in this channel might be done with the large amount of data from the proton-proton collisions at a center-of-mass energy of 7 TeV, since the *Wt*-channel production is the second largest source of the single top-quarks. In this thesis, the 2-jet topology of the *Wt*-channel production in the lepton+jets mode is mainly studied. The *Wt*-channel production process is described in detail in the first section of this chapter. All relevant background processes are addressed in the second section.

4.1 *Wt*-Channel Production in the Lepton+Jets Mode

In this analysis, the *Wt*-channel production in the lepton+jets mode is studied. The *Wt*-channel production was discussed in Section 2.3.2 and Section 2.3.3. The signature of this channel basically contains one charged lepton, missing transverse momentum, $E_{\mathrm{T}}^{\mathrm{miss}}$, and two light-quark jets together with one *b*-quark jet. The considered leptons are only electrons and muon in the final state, since they can be detected directly. τ-leptons are considered only if they decay leptonically, i.e. into one electron or muon and two neutrinos. The other modes of τ lepton decay are considered as background; therefore, the branching ratio of the signal is reduced to 34.1% of the total *Wt*-channel production. Figure 4.1 shows a Feynman diagram of *Wt*-channel production in the lepton+jets mode. The *W* boson from the top quark decays hadronically, while the other *W* boson decays leptonically. There is another diagram that also represents *Wt*-channel production in the lepton+jets mode. The *W* boson from the top quark decays leptonically, while the other *W* boson decays hadronically. This standard scheme is referred to as the 3-jet topology of the *Wt*-channel production. However, one of two light-quark jets might not be observed in the final state or additional light-quark jets can occur in some events, as well. Thus, there are two more topologies that are considered, called 2- and 4-jet topologies, which increase the fraction of the signal that can be observed.

The signal is distributed in the expected to be 2-, 3- and 4-jet topologies as 42:41:17 and the signal-to-background ratio of all three topologies as 2.4, 4.6 and 4.0%. All numbers are extracted from the analysis of *Wt*-channel production in the lepton+jets decay mode by the Bonn single top-quark group. After application of the event selection discussed in the full analysis [20], $t\bar{t}$ production and *W*+jets production are the two largest backgrounds in all three topologies. The role of each background is different in each topology. In the 2-jet topology *W*+jets production is the dominating background, while the contributions of $t\bar{t}$ production and *W*+jets production are rather similar in the 3-jet topology in size but not shape of their distributions. For the 4-jet topology $t\bar{t}$ production is the largest background because the appearance of an additional jet in the final state of *Wt*-channel production makes it look similar to the final state of $t\bar{t}$ production.

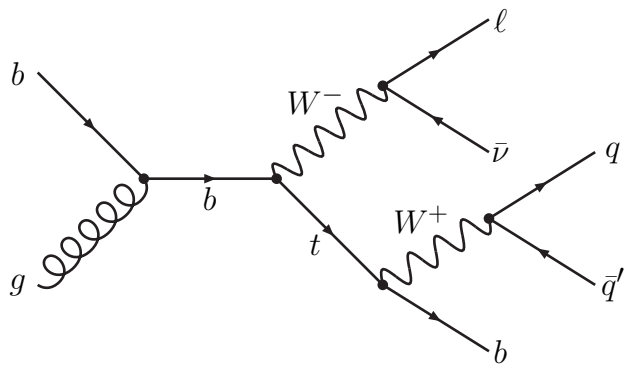

Figure 4.1: Feynman diagram of *Wt*-channel production in the lepton+jets mode. The *W* boson from the top quark decays hadronically, while the other *W* boson decays leptonically.

In this thesis, the *Wt*-channel production that is missing one of the two light-quark jets from the final state, "the 2-jet topology", is evaluated. This scenario might happen such that one of two light-quark jets could not pass the event selection discussed later in Section 5.4. Its event signature consists of exactly one charged lepton, missing transverse momentum, one light-quark jet and one *b*-quark jet. This topology is challenging, since it has the smallest signal-to-background ratio and $t\bar{t}$ production is always similar in shape with the signal. The complete analysis of *Wt*-channel production with all three topologies is discussed in [20].

4.2 Background Processes

There are several background processes which have the similar signatures to that of *Wt*-channel signal events. Figure 4.2 shows the production cross sections for physics processes involving *W* or *Z* bosons or top quarks occurring in the LHC. The cross section measurements of those processes are compared to their Standard Model prediction cross sections. All theoretical expectations were calculated at next-to-leading or higher order. Their measurements are grouped based on their components of *W* and *Z* bosons or top quarks. The background processes are the production of *W*+jets, *Z*+jets, diboson, the other two channels of single top-quark production and $t\bar{t}$ pairs. In addition, other processes that do not produce *W* or *Z* bosons or top quarks are classified as the so-called multijet background. Methods of estimating multijet background which are used in this thesis are detailed in Section 5.5. In the following, each background process is described in detail.

4.2.1 Single top-quark *t*-channel production

About 76% of single top-quark production is *t*-channel production. For a hadronically decaying top quark there are only jets in the final state of *t*-channel production; hence it is difficult to distinguish *t*-channel production from multijet background. In case of a leptonically decaying top quark, the final state consists of one lepton, missing transverse energy, and one *b*-quark jet together with one light-quark jet. One lepton and missing transverse energy are denoted as the main signatures of the signal. As a result, *t*-channel production can easily fake *Wt*-channel production, in particular with an additional light-quark jet, *t*-channel production fulfils the event selection of the 2-jet topology. However, its cross section is rather small compared to the overall background, so that the *t*-channel production is only a minor background.

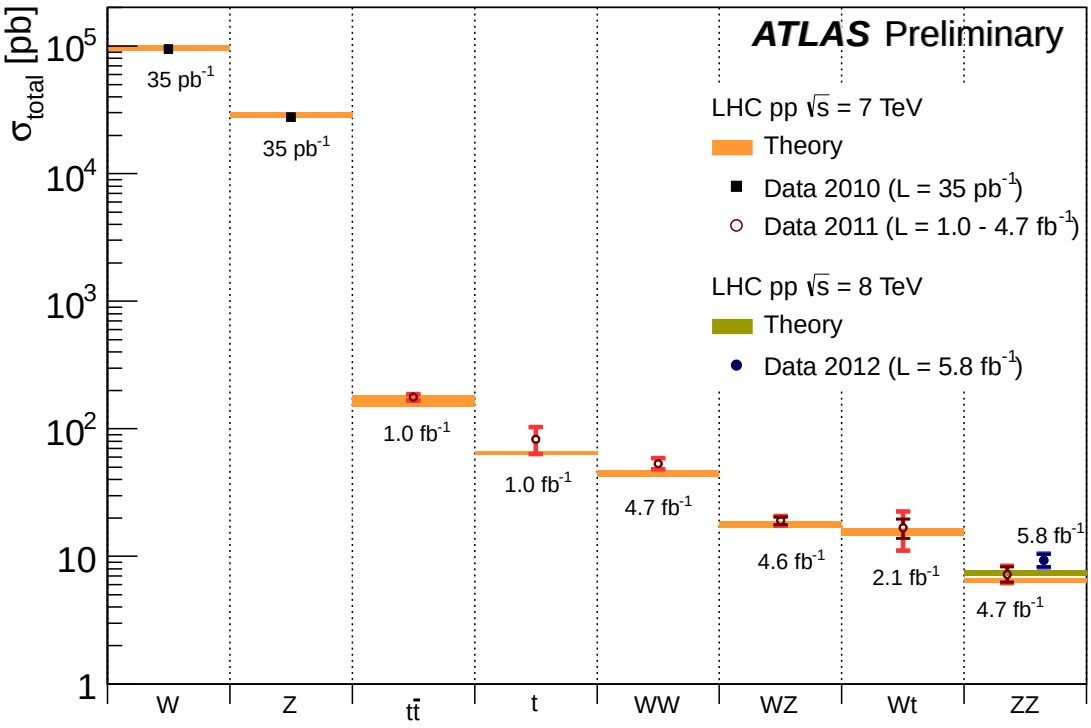

Figure 4.2: Measured cross sections for different physics channels compared to their corresponding expected cross sections. The 2010 dataset was used for measurements of the W and Z boson inclusive cross sections. The other measurements were made with the 2011 dataset or the 2012 dataset. The error bar in dark indicates only the statistical uncertainty. The error bar in red indicates the full uncertainty, i.e. statistical, systematics and luminosity uncertainties. All expected cross sections were calculated at next-to-leading or higher order.

4.2.2 Single top-quark *s*-channel production

Although single top-quark production in the *s*-channel contains the main signatures of *Wt*-channel production, if the top quark decays leptonically, nevertheless the cross section of *s*-channel production is only about 2/7 of the *Wt*-channel predicted cross section at $\sqrt{s}$ = 7 TeV. Furthermore, there are two *b*-quark jets in the final state of the *s*-channel production. The veto on a second *b*-quark jet, mainly performed to get rid of $t\bar{t}$ production, can also suppress *s*-channel production. ***s***-channel production is therefore the smallest background; however, it is still considered in the analysis.

4.2.3 Top-quark pair production

As discussed above, one of the two main backgrounds of the *Wt*-channel system in all three topologies is top quark pair ($t\bar{t}$) production. The final state of the $t\bar{t}$ system in the lepton+jets decay mode consists of one lepton, missing transverse energy, two light-quark jets and two *b*-quark jets as shown in figure 4.3 (left). Therefore, the final state of $t\bar{t}$ production is the same as the final state of *Wt*-channel production plus one additional *b*-quark jet as shown in figure 4.1. In collisions, one additional *b*-quark jet can simply in the final state of *Wt*-channel production. Furthermore, some final states of *Wt*-channel production at the next-to-leading order consist of exactly the final state of $t\bar{t}$ production as discussed in the theory

chapter. One possibility to suppress $t\bar{t}$ production is to veto the additional b-quark jet. However, the $t\bar{t}$ production cross section is about 10 times the Wt production cross section, which makes it the second largest background. To sum up, $t\bar{t}$ production can easily fake the signal, especially in the 4-jet topology, there are four quark jets in the final state. The fraction of the $t\bar{t}$ production is roughly 40% of the total background in the 4-jet topology.

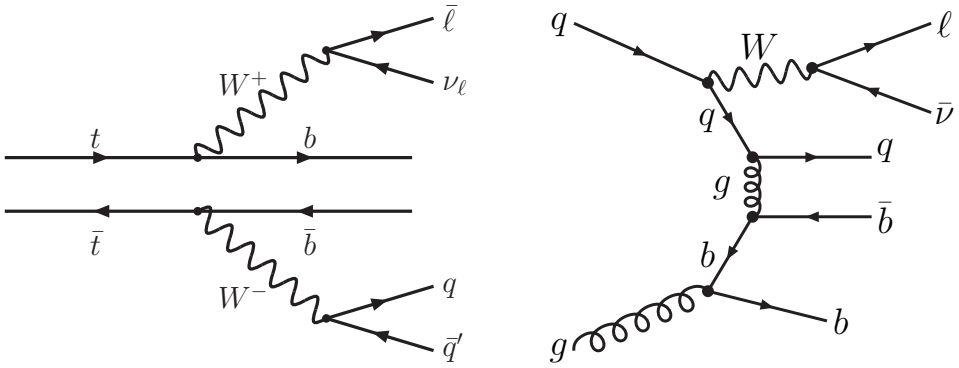

Figure 4.3: The two main backgrounds of Wt-channel production: an example Feynman diagram of $t\bar{t}$ production in the lepton+jets channel (left), W+jets production and its final state in the leptonically decaying W boson in association with a $b\bar{b}$ pair plus one light-quark jet (right).

4.2.4 W+jets production

The production of a single W boson is classified into two categories based on its decay products: hadronic and leptonic channels. It contains two jets in the hadronic decay mode. For the leptonically decaying W boson, it consists of one lepton and missing transverse energy, that leads to similar signatures to Wt-channel production, except two light-quark jets and one b-quark jet. It is also possible that the W boson is produced in association with one light-quark and two b-quark jets as shown in figure 4.3 (right). Therefore, the production of W+jets can fulfil all of the event selection as signatures of the signal therefore it is generally hard to get rid of. According to the so-called Behrends-Giele scaling factor [21], the W+jets events are suppressed exponentially with the higher jet bins; hence, it populates mostly in the 0- and 1-jet bins which are normally not populated by the signal. However, with its large cross section, W+jets production remains one of the two largest backgrounds of Wt production, especially for the 2-jet topology, where the W+jets production is about 57% of the overall background.

4.2.5 Z+jets production

The Z boson decays into either a pair of lepton-antilepton or quark-antiquark. Since it has no neutrino in the final state of the leptonic decay mode, it can easily be suppressed by the event selection. Also Z+jets production is suppressed by the Behrends-Giele scaling factor, it therefore prefers to accumulate in the two lowest jet bins. The production of a single Z boson is only a minor background, since it is easier to get rid of by the event selection than the W+jets production even though they have similar cross sections.

4.2.6 Diboson

There are three different processes of the diboson production: WW, WZ and ZZ production. The final state of the WW production looks like the final state of Wt-channel production excluding the b-quark jet. The WZ production can have the similar signatures to the production of W+jets. For the ZZ production, it may show the final state of Z+jets in pairs. However, diboson production has a rather small cross section; hence it is only a minor background.

4.2.7 Multijet background

The multijet background has the largest cross section among all backgrounds. The large amount of the multijet background is reduced in the first place by applying a good event selection; however, it cannot be totally suppressed. Furthermore data-driven methods are required for the multijet background estimation, since large Monte Carlo samples of multijet background is not available for the analysis.

Chapter 5

General Analysis Setup

This chapter presents the datasets used in this analysis. Section 5.1 explains the details of the collision dataset. The Monte Carlo samples of both signal and all backgrounds are described in section 5.2. Some details of Monte Carlo generators are also mentioned in that section. The criteria to discard a large amount of data without losing too much of the signal fraction is discussed in the second part. The definitions of physical objects are addressed in Section 5.3, e.g. what exactly is meant with an "electron" in this analysis. Then the event selection based on the Wt-channel production in the lepton+jets mode is given in Section 5.4. Section 5.5 will detail two methods of estimating the multijet background.

5.1 Data Samples

Collision data with a center-of-mass energy of $\sqrt{s} = 7$ TeV are used in this analysis. This dataset was taken during the 2011 run with an integrated luminosity of 4.7 fb^{-1}. There are two sub-datasets referred to as the electron channel whose data is from the Egamma stream requiring an electron trigger and the muon channel from the Muon stream requiring a muon trigger.

5.1.1 Luminosity blocks

The luminosity block is the unit for the luminosity measurement [22]. In general, the time period is every one to two minutes, but in some cases the value can be different due to other operational issues. The trigger system takes care of the timing information about the beginning and end time of each luminosity block. For each luminosity block, the average luminosity is measured. The collision data is taken corresponding to the luminosity blocks. Therefore, the sum of the luminosity blocks is easily interpreted as the integrated luminosity of the corresponding data. The data taking in small pieces of time allows one to select data based on the quality and performance of the sub-detectors together with beam conditions.

5.1.2 Good run lists

Data Quality (DQ) information of the collision data is studied in order to select only good data samples for any physics analysis. In ATLAS, DQ (status) flags are used to indicate the data quality. There are DQ flags for all components of the detector, for the trigger system and for the reconstruction of objects, since the data quality depends on how they all perform during data taking. Good run lists (GRLs) are formed using those DQ flags indicating only good data in the specified luminosity blocks. In general, the GRLs are provided by the ATLAS DQ group; however different DQ flags are required for different physics

analysis groups in order to achieve suitable good-data samples for particular analyses. For example, GRLs for this analysis and other top quark analyses are assessed by the top reconstruction group.

5.2 Monte Carlo Simulation

Monte Carlo samples are required in order to understand collision data. All relevant physics processes that pass the selection, expect for multijet are simulated using Monte Carlo generators. In the following, the general information corresponding to the Monte Carlo generators together with details of the simulated samples of both signal and relevant backgrounds are mentioned. All Monte Carlo samples used in the thesis stem from the 2011 Monte Carlo production round.

5.2.1 Event generators

The Monte Carlo generators AcerMC [23], MC@NLO [24, 25], PYTHIA [26], HERWIG [27] and ALPGEN [28] are general-purpose Monte Carlo generators. For parton showering and underlying event modelling [29], PYTHIA or HERWIG/JIMMY is used as the parton-level Monte Carlo generator. HERWIG models the parton showering and is complemented by JIMMY [30] for the underlying event modelling. The parton showers are generated via the QCD processes of gluon radiation and pair production. In addition to showering, there is another hadron production from the so-called underlying event. The underlying events are all multiple parton interactions including the effect of pileup in collisions [31, 32]. CTEQ6.6 [33] is the PDF set used for MC@NLO, the PDF set for ALPGEN is CTEQ6L and the LO* PDF set MRST [34] is used for AcerMC as well as for HERWIG. The default top quark mass is m_t = 172.5 GeV with Γ_t = 1.32 GeV. For the W and Z bosons, their default masses and decay widths are m_W = 80.4 GeV with Γ_W = 2.09 GeV and m_Z = 91.2 GeV with Γ_Z = 2.50 GeV.

Event Weights

For the next-to-leading order generators, for example the MC@NLO generator, events can have positive or negative event weights [35]. Furthermore, in the simulation the theoretical cross section is used as a fixed parameter for each process generated and the number of events is determined corresponding to this cross section. With the number of generated events and the theoretical cross section, the luminosity of each Monte Carlo sample is calculated. In real data, luminosity is the quantity normally used to express the amount of collected data. The Monte Carlo samples are scaled to the integrated luminosity of collision data.

Pileup reweighting

Pileup is a term used to present the information of more than one particle interaction in the collisions. Pileup becomes a big issue with the high instantaneous luminosity. In general, the Monte Carlo samples are generated before or during the data taking, only a best guess of the data pileup is therefore applied to the Monte Carlo. Therefore, the Monte Carlo pileup must be corrected to the pileup of collision data. Comparison of the distribution of the average number of pileup interactions of the data and that of the Monte Carlo allows one to compute this event weight.

Cross section

For higher orders of the perturbation theory, K-factors are used as the higher order QCD corrections of the cross section calculated at leading order. All Monte Carlo samples are therefore normalised to their

theoretical cross sections using K-factors.

5.2.2 Monte Carlo samples

Single top-quark production

The samples for the signal single top Wt-channel and the background single top s-channel samples are produced with the MC@NLO generator which is interfaced to HERWIG/JIMMY for parton showering. The leading order generator AcerMC is used for the single top t-channel production. PYTHIA is used as the parton shower generator for the AcerMC generator. Only the lepton+jets decay mode of the s- and t-channel production is simulated. For Wt-channel production, all decay channels are generated. The approximate number of the Wt-channel signal events is 900,000. The sample size is about 2,800,000 events for t-channel production and about 900,000 events for s-channel production.

Top quark pair production

The simulation of top quark pair production is done using the MC@NLO generator plus HERWIG/JIMMY for parton showering. The samples for the fully hadronic decay mode and for the non-fully hadronic decay mode are simulated separately. The simulated number is about 400,000 events for the fully hadronic samples and 14,600,000 events for the non-fully hadronic samples.

W+jets production

The WHF+jets and WLF+jets are used as abbreviations for the associated production of a W boson with heavy and light quarks, respectively. The WHF+jets sample can be broken down into $W+b\bar{b}$+jets; the production of a $b\bar{b}$ pair and the number of jets from zero to three in association with a W boson. $W+c$+jets is the production of a single c quark and the number of partons from zero to four. The last sub-sample of WHF+jets is $W+c\bar{c}$+jets which is the production of a $c\bar{c}$ pair with the number of partons from zero to three. The c quark is assumed to have no mass. The WLF+jets events contain only zero to five partons in association with the W boson. The W boson is allowed to decay only leptonically in both cases. The Monte Carlo generator ALPGEN interfaced to HERWIG/JIMMY is used to model both WHF+jets and WLF+jets events. For the WHF+jets events, heavy flavor overlap removal is applied using the HFOR tool [36] in order to avoid double counting. The WLF+jets samples are about 33,100,000 events in total. The simulated number is approximately 900,000 events for the $W+b\bar{b}$+jets, 9,200,000 events for the $W+c$+jets and 3,200,000 events for the $W+c\bar{c}$+jets.

Z+jets production

The Z+jets samples are also simulated using the ALPGEN generator interfaced to HERWIG/JIMMY. Only Z bosons which decay into a lepton-antilepton pair, $e\bar{e}$, $\mu\bar{\mu}$ or $\tau\bar{\tau}$ in the invariant mass range of 40 GeV $< m(l\bar{l}) <$ 2000 GeV are generated. In addition, the full Drell-Yan process from $\gamma^* \rightarrow l\bar{l}$ is included in the Z+jets samples since the Drell-Yan process allows an Z boson and virtual photon interference. The Z+jets production is separated corresponding to the number of partons (from zero to five) in association with the Z boson. The number of generated events is about 38,000,000.

Diboson production

The diboson WW, WZ and ZZ samples are produced with the leading order generator HERWIG. Only samples that have at least one lepton with $p_{\mathrm{T}} > 10$ GeV and $|\eta| < 2.8$ are selected. The number of

simulated WW events is about 2,500,000. The simulated number is approximately 250,000 events for WZ sample and 1,000,000 events for ZZ sample.

5.3 Object Definitions

The definitions of the physical objects which are used in the single top-quark analyses (together with other top-quark analyses) are based on the recommendations of the ATLAS top reconstruction group. In this analysis, electrons, muons, jets and missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}}$, are reconstructed following the recommendation of the ATLAS top reconstruction group for 2011 data [37].

5.3.1 Electrons

Electron reconstruction and identification

Information from the electromagnetic calorimeter and the inner detector is used to reconstruct electrons. ATLAS electron reconstruction is based on three algorithms: the standard electron algorithm, the soft electron algorithm and the forward electron algorithm. The standard electrons have high transverse momentum, p_{T}, and appear in the electromagnetic calorimeter as clusters together with momentum reconstruction of their tracks in the inner detector.

The standard electrons are used in this analysis. Other physical objects, e.g. hadrons, can be misidentified as electron candidates; therefore, electron identification is applied in order to identify real electrons. In the standard algorithm, electrons are reconstructed using the energy of cluster and the four-momentum of the track as well as the other two qualities: the η - ϕ distance between the cluster of 0.05 $\times$ 0.10 and the track denoting as the track–cluster matching, together with the ratio between the cluster energy and track momentum of $E/p < 10$. Electrons with a transverse energy of $E_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$ are selected as electron candidates.

According to the cut-based identification, three reference sets of cuts are defined: loose, medium and tight. **Loose cuts**: this set of cuts is based only on limited information of the calorimeters. It provides very high identification efficiency, but low background rejection. Cuts are performed only on the hadronic leakage and on shower shape variables from the middle layer of the electromagnetic calorimeter. **Medium cuts**: additional cuts are applied on the strips in the first layer of the electromagnetic calorimeter and on tracking variables. This set of cuts improves the jet rejection by a factor of 3-4 with respect to the loose cuts; however it reduces the identification efficiency by about 10%. **Tight cuts**: the tight (TRT) cuts optimised for isolated electrons include tighter cuts on the TRT information in order to further remove the background from the charged hadrons but without applying any energy isolation cut. This set of cuts provides the highest isolated electron identification and the highest jet rejection against jets.

Electron isolation

The electron isolation criteria may clean the signal by factors up to order of 10^3. There are two variables normally used to isolate electrons: the transverse energy deposited in a cone of $\Delta R = 0.\mathrm{X}$ around the electron cluster, $E_{\mathrm{T,cone}}(\mathrm{X})$ and the scalar sum of the transverse momentum of all tracks determined in a cone of $\Delta R = 0.\mathrm{X}$ around the electron cluster, $p_{\mathrm{T,cone}}(\mathrm{X})$.

Electron definition

Only the electrons passing the standard reconstruction algorithm together with the **tight++** definition are used in this study. The **tight++** definition is based on the tight cuts with more variables used and the cut values being varied in accord with a higher pileup environment [38]. Electrons are required to have a transverse energy, $E_{\mathrm{T}} = E_{\mathrm{cluster}} / \cosh(\eta_{\mathrm{track}}) > 25$ GeV in the central region of the detector, $|\eta_{\mathrm{cluster}}| < 2.47$. However electrons are discarded in the transition region between the barrel and end-caps of the electromagnetic calorimeter, $1.37 < |\eta_{\mathrm{cluster}}| < 1.52$, because the dead material in front of the electromagnetic calorimeter leads to large drops in the reconstruction efficiency and the energy resolution. The electron isolation criteria with 90% efficiency using the variables $E_{\mathrm{T,cone}}(20)$ and $p_{\mathrm{T,cone}}(30)$ are required for all single top-quark analyses. Electrons fulfilling the electron isolation criteria used as the standard electrons called **tight electrons** through this analysis, while electrons that do not fulfill the isolation criteria are used for multijet background estimation and are referred to as **loose electron**.

The overall reconstruction efficiency consists of the trigger, reconstruction and standard identification efficiencies. To correct the small Monte Carlo–data discrepancies in those efficiencies, scale factors are applied to all selected electrons. An energy scale factor is also applied to the measured electron energy in data in order to correct the cluster energy. The energy scale factor is determined from studies of Z, W and J/ψ decays. Smearing the measured cluster energy on all Monte Carlo samples is the method which is used to adjust the electron resolution between the Monte Carlo samples and collision data.

5.3.2 Muons

Muons are reconstructed using the track information of the inner detector and of the muon spectrometer. There are three types of the reconstructed muons: standalone, combined and tagged muons. Standalone muons are determined only from reconstructed tracks in the muon spectrometer by the MuID standalone algorithm. Combined muons are determined from an inner detector track matching to a muon spectrometer track by the MuID combined algorithm. There are two subsets of tagged muons: segment-tagged muons and calorimeter-tagged muons. The segment-tagged muons are created by the inner detector tracks which are associated to segments in the muon system by the MuGirl and MuTagIMO algorithms. The calorimeter-tagged muons are from the inner detector tracks which match signals in the calorimeter. Only the segment-tagged muons are referred to as tagged muons which are used in general analyses. Similar to the reference sets of cuts in the electron identification, there are three qualities of the muon identification: loose, medium and tight. For the muon isolation criteria, $E_{\mathrm{T,cone}}(\mathrm{X})$ and $p_{\mathrm{T,cone}}(\mathrm{X})$ are also used as the important isolation variables.

Muon definition

Only the MuID combined muons, which already fulfill the tight quality, are used in this analysis. The muon candidates are required to be in the central region with $|\eta| < 2.5$ and have a transverse momentum of $p_T > 25$ GeV. There are five additional hit requirements applied on the inner detector. In the b-layer of the Pixel Detector, the muon tracks have to show at least one hit. The sum of the number of hits and the number of crossed dead sensors for the Pixel Detector is at least two and for the SCT at least six. The sum of the number of holes in Pixel and SCT is less than three. Finally, N is defined as the sum of the number of hits and N_{outliers} as the number of outliers in the TRT. In the TRT barrel region with $\eta < 1.9$, N must be > 5 with $N_{\mathrm{outliers}}/N < 90$ % and in the end-cap region with $\eta \geq 1.9$, if $N > 5$ and N_{outliers}/N must be < 90 %. Muon isolation criteria, $E_{\mathrm{T,cone}}(30)/p_{\mathrm{T}} < 0.15$ and $p_{\mathrm{T,cone}}(30)/p_{\mathrm{T}} < 0.10$ are required. The muons fulfilling this criteria are used as the standard muons, named as **tight muons**; on the other hand, the muons falling the isolation criteria are called **loose muons** and used only for the multijet background

estimation. Moreover, an overlap removal between muons and jets is used in order to discard muons that have a distance of $\Delta R < 0.4$ to a jet with $p_{\mathrm{T}} > 20$ GeV and with a jet vertex fraction, JVF, of more than 0.75. The JVF quantifies the fraction of charged particle transverse momentum in each jet.

In the muon collision data, the trigger, reconstruction and matching-ID efficiencies as well as energy scale and resolution factors have to be applied similarly as for the electrons in order to reduce discrepancies between collision data and Monte Carlo samples. The scale factors are provided by the muon performance group.

5.3.3 Jets

In ATLAS, jet reconstruction is based on the information of the calorimeter system covering $|\eta| < 4.9$. Jets are reconstructed by the anti-k_T algorithm [39] with a size parameter of $R = 0.4$. In top-quark analyses, topological cell clusters are used as seeds. The jets are calibrated using the EM+JES scheme. The EM+JES calibration applies the electromagnetic (EM) scale which is the default energy scale for any calorimeter measurements and hereafter the JES correction is applied to each jet depending on its transverse momentum and pseudorapidity. Jets with a transverse momentum of $p_T > 25$ GeV in the central region with $|\eta| < 2.5$ are required for this analysis. Also, a jet with a distance of $\Delta R < 0.2$ to a selected electron is removed from the event. The jet-electron removal is applied after the muon-jet removal to each event.

5.3.4 *b*-quark jets

To identify the jets from the *b* quarks, *b*-tagging algorithms are employed. Since the lifetime of *b* quarks is significantly longer than that of light quarks, the *b*-quarks decay via the weak interaction at a longer distance from the primary vertex. This leads to a measurable secondary vertex used as the most important one of parameters to distinguish the *b* jets from the light jets. There are several *b*-taggers for example the JetFitterCOMBNN tagger and the MV1 tagger. The JetFitterCOMBNN tagger is the neural network based combination of the IP3D and JetFitterTagNN. The MV1 tagger is the multivariate combination of the three different taggers: IP3D, SV1 and JetFitterCOMBNN.

***b*-tagging working point.** The ATLAS flavor tagging working group provides different *b*-tagging working points for each tagger. A weight cut value, *b*-tagging efficiency, and a factor of the light quark rejection, RF, are provided for each working point. The probability of a *b* quark jet to be tagged by the *b*-tagging algorithm above the weight cut is referred to as *b*-tagging efficiency. The light quark rejection factor is the inverse of the mistagging probability (probability of a light quark jet with a tag weight above the weight cut).

5.3.5 Missing transverse energy

Missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}}$ is the measured energy of the escaping neutrino. The $E_{\mathrm{T}}^{\mathrm{miss}}$ is reconstructed from energy deposits in the calorimeters and muon tracks. It is calculated as the vector sum of all topological cell clusters in the event. Double counting of cells is taken into account such that only one type of object belongs to one cell.

5.4 Event Selection

Using the object definitions, the physical objects used through this analysis are selected in the first stage. The event selection is then applied in order to discard undesired data. The Monte Carlo samples and collision data should therefore be in good agreement after applying the event selection. The event selection is divided into two sections named as pretag selection and tag selection discussed in detail later. Before the event selection, event cleaning cuts are applied.

5.4.1 Event cleaning cuts

Only events which are in the good run list (GRL) up to period M are considered. Different triggers are needed for the electron and muon channels in the different run periods as shown in Table 5.1. Events with at least one "bad" jet are discarded according to prescriptions of the jet MET group [40]. The "bad" jet can imply a problem in the calorimeter in the event or the event as a non-collision event. To reject non-collision background, only events with a first primary vertex having at least four tracks are taken into account. Also, the dead front-end region in the liquid argon calorimeter leading to a large hole named as "the LAr hole" has to be considered such that all events with "larError" > 1 [37] have to be removed from the analysis.

Run period	$\mathcal{L}$ [fb^{-1}]	Electron channel		Muon channel
		Data	MC	Data & MC
B-I	1.46	EF_e20_medium	EF_e20_medium	EF_mu18
J	0.23	EF_e20_medium	EF_e20_medium	EF_mu18_medium
K	0.59	EF_e22_medium	EF_e22_medium	EF_mu18_medium
L-M	2.43	EF_e22vh_medium1	EF_e22_medium1	EF_mu18_medium

Table 5.1: Trigger requirement for electron and muon channels in the different run periods. Details of the triggers are discussed in [38].

5.4.2 Pretag selection

The pretag selection requires events which contain exactly one lepton with $P_T > 25$ GeV, that is the presence of exactly one electron and no muon in the electron channel, and exactly one muon and no electron in the muon channel. This lepton in each selected event has to match the lepton that also fired the trigger within a distance of $\Delta R = 0.15$. To take the neutrino from the W-boson decay into account, only events with a missing transverse energy of $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV in the electron channel and of $E_{\mathrm{T}}^{\mathrm{miss}} > 25$ GeV in the muon channel are considered. Furthermore, at least two jets have to be found in the events.

The transverse mass of the leptonically decaying W boson, $m_{\mathrm{T}}(W^{\mathrm{lep}})$, which is reconstructed from the four-momenta of lepton and neutrino is used to suppress multijet background events. In the electron channel, $m_{\mathrm{T}}(W^{\mathrm{lep}}) > 30$ GeV is required. A triangular cut of $m_{\mathrm{T}}(W^{\mathrm{lep}}) + E_{\mathrm{T}}^{\mathrm{miss}} > 60$ GeV is applied in the muon channel, since multijet events prefer to have low $m_{\mathrm{T}}(W^{\mathrm{lep}})$ and low $E_{\mathrm{T}}^{\mathrm{miss}}$. Further details of reconstructed variables are discussed in Section 6.1.

In this analysis, which focuses on the 2-jet topology, only events missing one light-quark jet are considered, i.e. exactly two jets must be found in the events.

5.4.3 Tag selection

After the pretag selection, the tag selection is applied, consequently the ratio of the signal to the background is significantly improved. The tag selection requires that exactly one of the jets has a *b*-tagged jet with a transverse momentum of $p_{\mathrm{T}} > 25$ GeV using the MV1 tagger at a working point efficiency of 70%. The requirement of exactly one *b*-quark jet aims at vetoing $t\bar{t}$ events, which is one of the two main backgrounds of the *Wt*-channel.

5.5 Multijet Background Estimation

The multijet background is problematic to generate directly by Monte Carlo generators, since the multijet background has the largest cross section. Hence very high statistics are used and large reweighting factors have to be applied to the Monte Carlo events. The lack of *W* and *Z* bosons (together with top quarks which decay to *Wb*) in the final state leads to only fake leptons which are selected in the events. Fortunately, these fake leptons are different from the real leptons due to their less isolation and smaller missing transverse energy. Since true *b*-quark production is also one subset of the multijet background, events containing the *b*-jets are then separated from the multijet background by the *b*-tagging requirement. Hence, a large amount of the multijet background can be suppressed from analyses using the lepton reconstruction with isolation criteria, the $E_{\mathrm{T}}^{\mathrm{miss}}$ kinematics and the *b*-tagging requirement. On the other hand, it cannot be fully suppressed due to its large cross section. Data-driven methods are normally used for the multijet background estimation.

In this analysis, for both the electron and muon channels the multijet background is estimated using the so-called "jet-electron model". Furthermore, for the muon channel the so-called "matrix method" is also implemented as a cross-check for the jet-electron method.

5.5.1 Jet-electron model

The jet-electron method models the shape of the multijet background by selecting events in which the electron is replaced by a jet showing a similar signature to the electron in the electromagnetic calorimeter. However they are not the real electron candidates a different trigger is used. Each of these jets is called as a jet-electron. Only events containing exactly one jet-electron and no electron according to the definition given above are used. The model uses a sample of jets with the Jet20 trigger selection as explained in [41]. If the energy deposit of the jet in the electromagnetic calorimeter is approximately 80-95% of its total energy, this jet is taken as candidate. Furthermore, this jet has to pass the same kinematic selection as electron candidates, i.e. only jets having a transverse momentum of $p_T > 25$ GeV in the central region with $|\eta| < 2.47$ and outside of the crack region are accepted as the signal electrons. Also, at least four tracks have to be found in the jet; hence, the contribution from converted photons is reduced. The triangular cut is applied to all jet-electron events as well. However the missing transverse momentum cut is excluded in the selection, since the shape of $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution is used to extract the number of multijet background events. In order to do so, a binned likelihood fit to the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution is performed. From the result of the fit, the normalisation of multijet events in the signal region with $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV is obtained and only this part is further used in the analysis.

For the muon channel, the fraction of multijet background events is also determined using the jet-electron model. But those jets are used as muons, thus the muon selection is used instead of the electron selection.

Agreement between Monte Carlo and data of variables with the use of jet-electron model is obtained as shown in figure 5.1. In particular the distribution of the $E_{\mathrm{T}}^{\mathrm{miss}}$, the shape of the multijet background

follows the shape of the $E_{\mathrm{T}}^{\mathrm{miss}}$. However some variables indicate that data is not well-descried by the Monte Carlo samples in the muon channel (also for electron channel). Figure 5.2 shows the two most problematic variables: the transverse momentum of the muon and η of the muon in the first row and two moderately problematic variables: the sum of all E_{T} and the invariant mass of the leptonically decaying top quark in the second row. One reason for the shape difference between Monte Carlo and data is the use of the jet-electron model which establishes the shape of the multijet background using the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution of the jet-electron events and it is initially designed for estimating multijet in electron channel.

5.5.2 Matrix method

Since the discrepancy between the Monte Carlo and data distributions mainly in the muon channel occurs when the multijet background is determined using the jet-electron model as described in Section 5.5.1, the matrix method for the multijet background estimation is implemented for the muon channel only. The matrix method is described in detail in [41] and [42]. In the matrix method, the multijet background events in the signal region are extracted from tight muons and loose muons as presented in Section 5.3.2, where the only difference between them is the loose selection does not require the muon isolation criteria. The number of events passing the loose and tight selections, labelled as N^{loose} and N^{tight}, respectively, can be expressed as linear combinations of the number of events with one real muon or one fake muon as

$$\begin{aligned} N^{\mathrm{loose}} &= N_{\mathrm{real}}^{\mathrm{loose}} + N_{\mathrm{fake}}^{\mathrm{loose}}, \\ N^{\mathrm{tight}} &= N_{\mathrm{real}}^{\mathrm{tight}} + N_{\mathrm{fake}}^{\mathrm{tight}} \\ &= \epsilon_{\mathrm{real}} N_{\mathrm{real}}^{\mathrm{loose}} + \epsilon_{\mathrm{fake}} N_{\mathrm{fake}}^{\mathrm{loose}}. \end{aligned} \tag{5.1}$$

$\epsilon_{\mathrm{real}} = \frac{N_{\mathrm{real}}^{\mathrm{tight}}}{N_{\mathrm{real}}^{\mathrm{loose}}}$ and $\epsilon_{\mathrm{fake}} = \frac{N_{\mathrm{fake}}^{\mathrm{tight}}}{N_{\mathrm{fake}}^{\mathrm{loose}}}$ are the fractions of real muons in the loose selection that also pass the tight selection and the fractions of fake muons in the loose selection that also pass the tight selection, respectively. After rearranging those equations, the number of events selected only by the tight selection with one fake muon is given as

$$N_{\mathrm{fake}}^{\mathrm{tight}} = \frac{\epsilon_{\mathrm{fake}}}{\epsilon_{\mathrm{real}} - \epsilon_{\mathrm{fake}}} (\epsilon_{\mathrm{real}} N^{\mathrm{loose}} + N^{\mathrm{tight}}). \tag{5.2}$$

Therefore, one can estimate the number of the multijet background events, $N_{\mathrm{fake}}^{\mathrm{tight}}$ from the event yields of the muon data after applying the loose selection and tight selection, if the real efficiency, ϵ_{real}, and the fake efficiency, ϵ_{fake}, are known. The matrix method takes care of only events that contain exactly one loose muon passing the tight selection. A problem is raised, since some events have one tight muon but more than one loose muons and such these events also pass the event selection. However they are only in a few percent of the total number of events.

The real efficiency, ϵ_{real}, is determined directly from the Monte Carlo samples of the physical processes that can produce the real isolated muons e.g. $Z \to \mu^+\mu^-$. The fake efficiency is estimated from the data sample by utilising the muon d0 significance using the assumption that the multijet background events in the muon channel are dominated by heavy flavor jets. The transverse impact parameter which is from the ID track with respect to the primary vertex. The d0 significance of a track is defined to measure the probability that the track generated from the primary vertex. The details of how to measure the

fake efficiency are described in another analysis of the *Wt*-channel production [43]. The fake efficiency is determined in a multijet background enriched control region, that is obtained by inverting the missing transverse energy and triangular cut requirements of the muon selection.

The implementation of the matrix method follows the documentation [44]. In this analysis, the loose and tight muons are chosen with the requirements of the missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}} > 25$ GeV and the triangular cut, $E_{\mathrm{T}}^{\mathrm{miss}} + M_{\mathrm{T}}(W) > 60$ GeV; hence, the fake efficiency is calculated using $E_{\mathrm{T}}^{\mathrm{miss}} < 25$ GeV and $E_{\mathrm{T}}^{\mathrm{miss}} + M_{\mathrm{T}}(W) < 60$ GeV. The multijet event weight for each selected event is obtained from the matrix method tool [43]. Then one can apply this multijet event weight back to the corresponding muon data to get the multijet event yield in the signal region. Table 5.2 shows the number of the multijet background events in 2, 3 and 4 jet bins.

Jet bin	Pretag events	Tagged events
2-Jet	15880 ± 7940	3190 ± 1600
3-Jet	4240 ± 2120	1180 ± 590
4-Jet	1050 ± 530	310 ± 160

Table 5.2: Multijet event yields using matrix method.

The uncertainty combines statistical and systematic uncertainties. The uncertainty of the multijet background estimated using the matrix method is taken to be 50% of the total event yield. The cross-check of the problematic distributions in the muon channel with the jet-electron model is shown in figure 5.3. The distribution of the transverse momentum of muon is significantly improved in the lower p_{T} region with the use of the matrix method, even if there is still a small discrepancy in 25 GeV $< p_{\mathrm{T}} <$ 30 GeV. The muon η, the sum of all E_{T} and the mass of the leptonically decaying top quark distributions are also improved. The distribution of other variables shows the similarity between the use of the matrix method and of the jet-electron model for the multijet background estimation in the muon channel. Nevertheless the consistency between the Monte Carlo and data in the $E_{\mathrm{T}}^{\mathrm{miss}}$ with the matrix method becomes worse in the low $E_{\mathrm{T}}^{\mathrm{miss}}$ region.

5.5.3 Comparison between matrix method and jet-electron model

Only the results in the muon channel can be used to compare between the jet-electron model and the matrix method. In the 2-jet bin, the event yield of the multijet background using the jet - electron model is 3450 that is larger than the multijet event yield using the matrix method, i.e. 3190 by about 8%; however one can interpret that the agreement is within the uncertainties. In the other bins, the consistency between those numbers within the uncertainties can also be seen. The first step of any analyses should show that the data can be (well) described by the Monte Carlo in (ideally) every variable. As discussed above, the matrix method shows the better consistency between data and Monte Carlo in some problematic variables. In particular, the multijet background estimation of the matrix method makes the muon p_{T} of data very well-described by the Monte Carlo. On the other hand, it makes the agreement on the $E_{\mathrm{T}}^{\mathrm{miss}}$ worse. The $E_{\mathrm{T}}^{\mathrm{miss}}$ of data is very well described by the Monte Carlo with the jet - electron model for the multijet estimation, nevertheless the muon p_{T} distribution in the lower p_{T} is not. Furthermore, the overall outcome of both methods is not much different; therefore, one can simply apply one's preferred method.

In the following, only the jet - electron model is used for both electron and muon channels, since the jet - electron model has been available for use since the beginning of the analysis, and gives promising

results in various other single top-quark analyses. Furthermore, from the above results, the discrepancy in the overall between the use of the jet - electron model and of the matrix method is rather small.

5.5.4 Event yields after event selection

Table 5.3. lists the event yields after event selection. The number of the multijet background events is determined using the jet-electron method for both the electron and muon samples. Except the 50% uncertainty assumed for the multijet events, all uncertainties are derived from their theoretical cross-section uncertainties. The expected event yields from Monte Carlo samples and the event yields of collision data are in good agreement after the selection.

	2-jet topology	
	Electron channel	Muon channel
Wt-channel	760 ± 50	1020 ± 70
t-channel	1740 ± 70	2340 ± 90
s-channel	127 ± 5	181 ± 7
$t\bar{t}$	4480 ± 450	5940 ± 590
W+jets	19180 ± 6520	28020 ± 9530
Diboson	370 ± 20	530 ± 30
Z+jets	720 ± 250	1020 ± 350
Multijet	2510 ± 1260	3450 ± 1720
Total Expected	29890 ± 6660	42500 ± 9710
Data	28960	43130

Table 5.3: Event yields for the electron and muon channels after event selection with the jet-electron model for multijet background estimation.

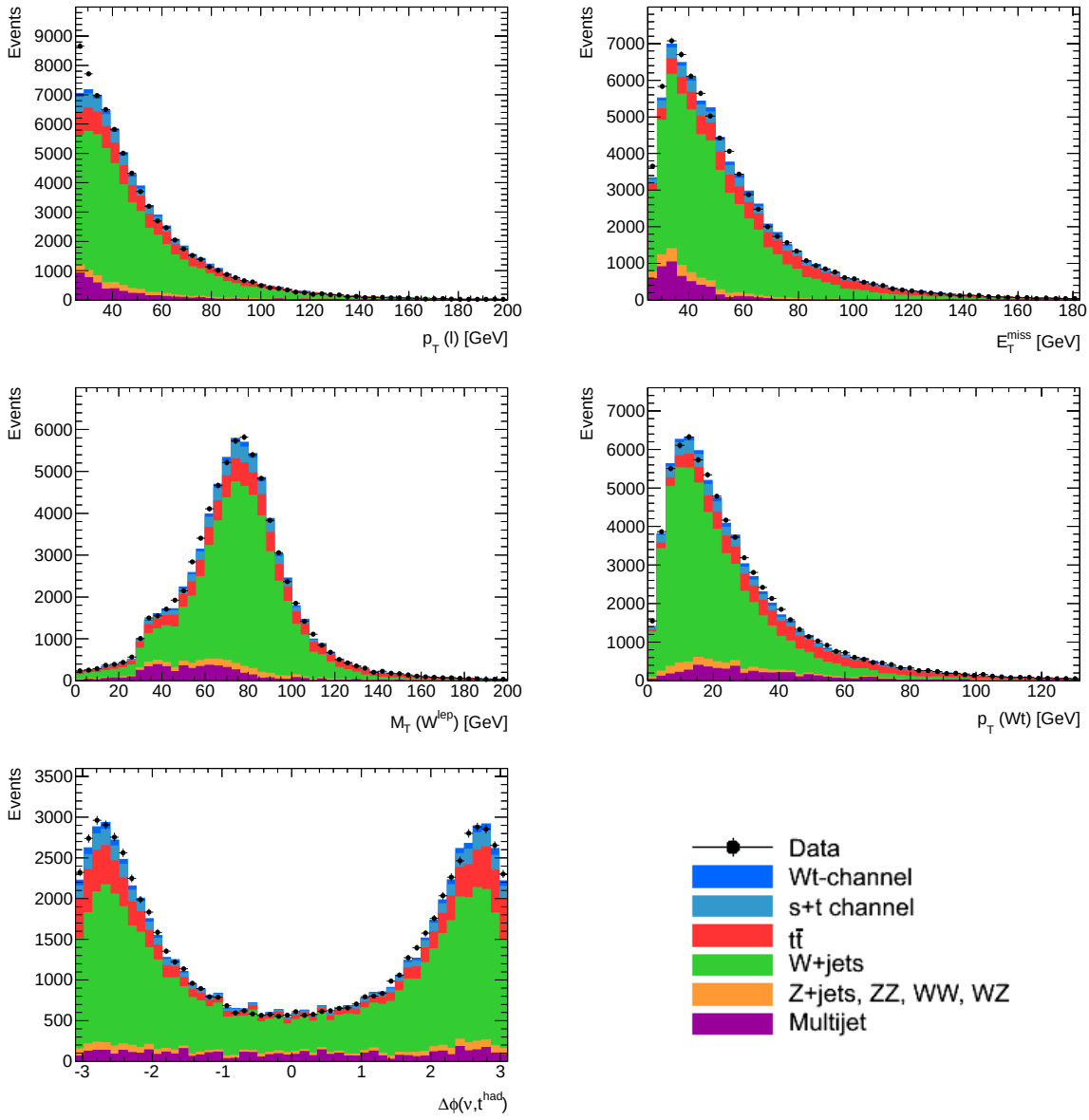

Figure 5.1: Control plots after the event selection using the jet-election method for the multijet background estimation. Electron and muon channels are summed up. The distributions of these plots show good agreement between data and Monte Carlo. In the first row: $p_{\mathrm{T}}(l)$ and $E_{\mathrm{T}}^{\mathrm{miss}}$. In the second row: $m_{\mathrm{T}}(W^{lep})$ and $p_{\mathrm{T}}(Wt)$. In the third row: $\Delta\phi(\nu, t^{\mathrm{had}})$. All reconstructed variables are detailed in Section 6.1

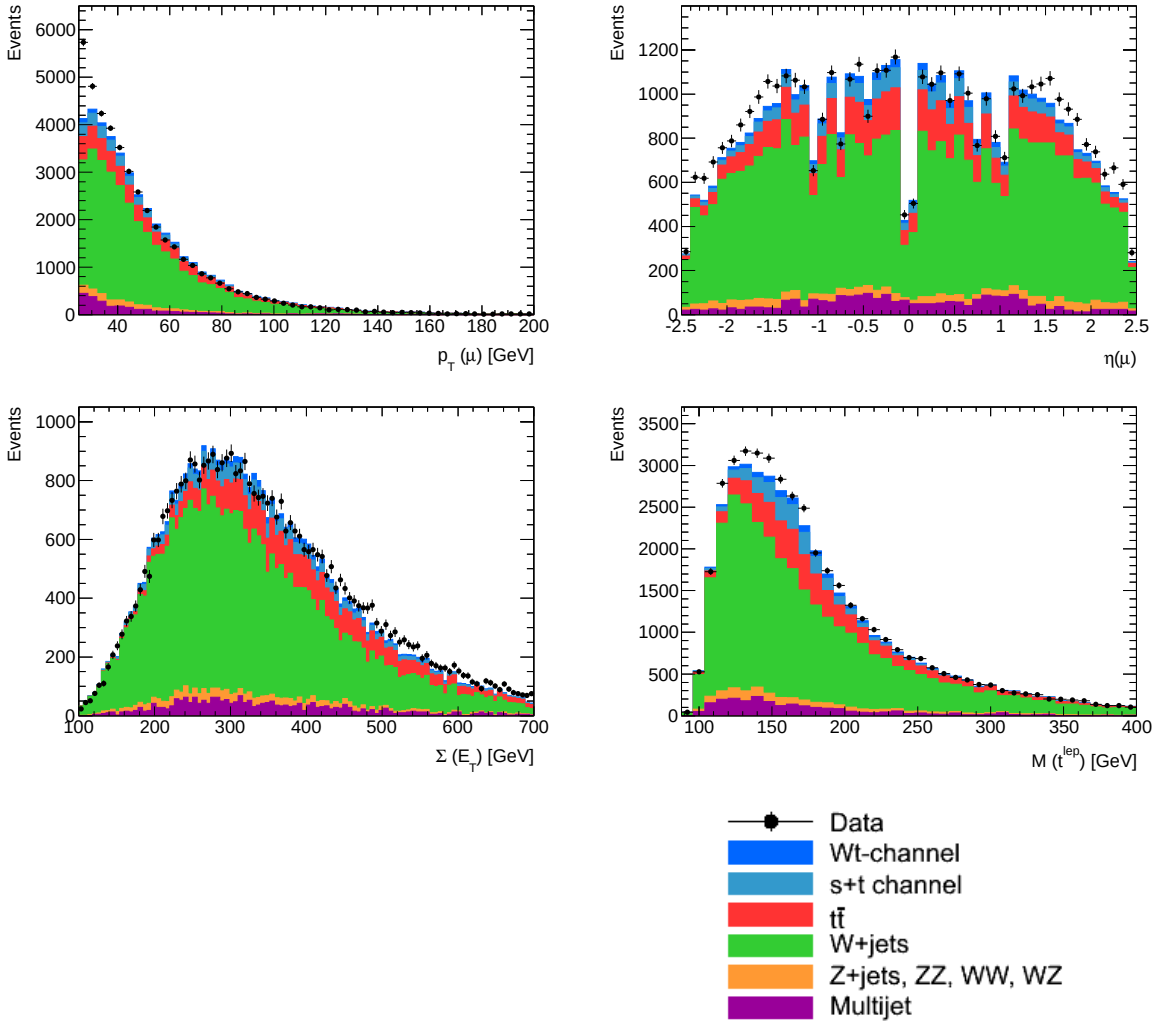

Figure 5.2: Problematic plots using the jet-electron model in the muon channel. In the first row: two most problematic variables, $p_{\mathrm{T}}(\mu)$ and $\eta(\mu)$. In the second row: two moderately problematic variables, $\Sigma(E_{\mathrm{T}})$ and $M(t^{lep})$.

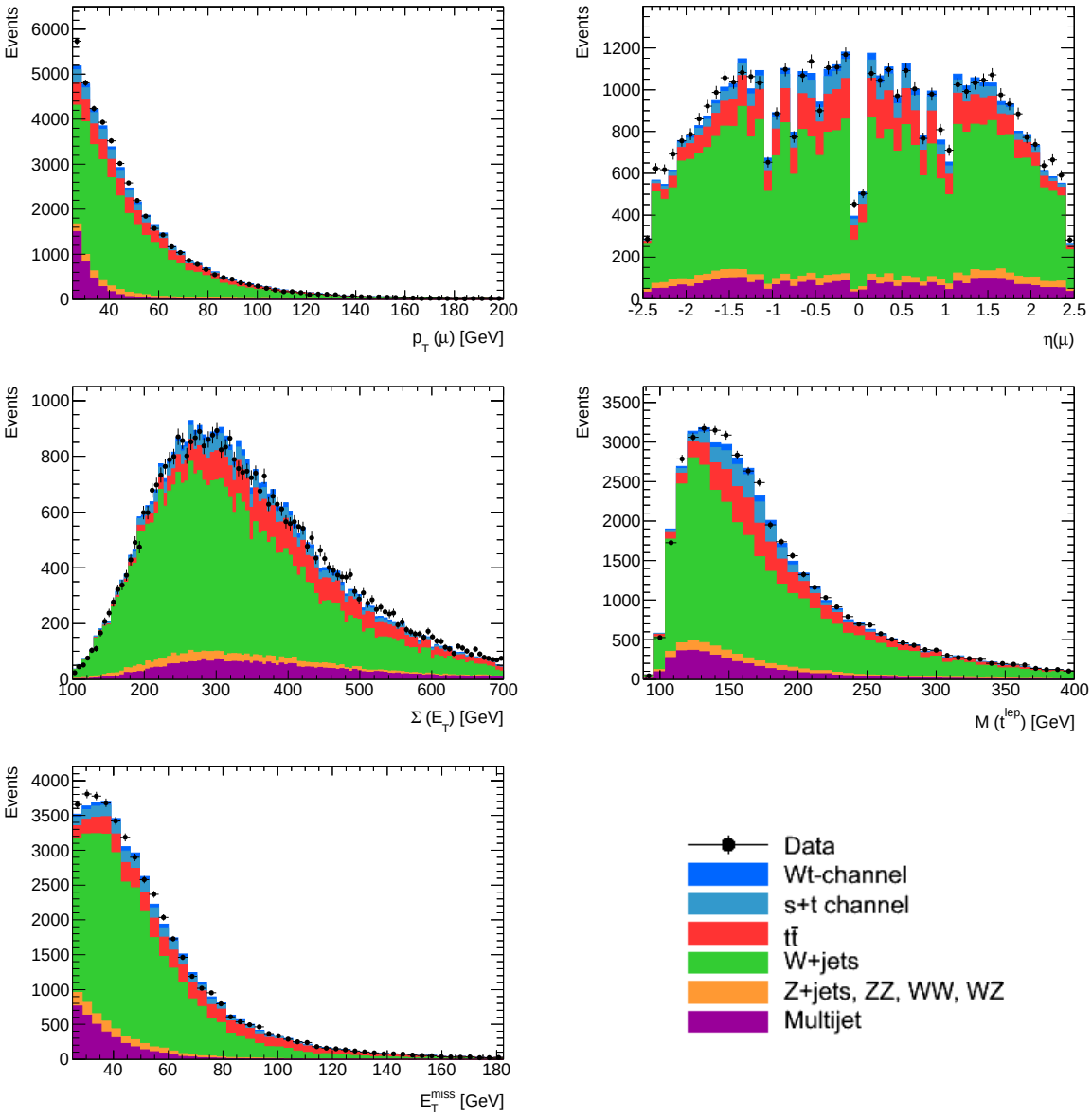

Figure 5.3: Control plots in the muon channel with multijet from the matrix method. In the first row: $p_T(\mu)$ and $\eta(\mu)$. In the second row: $\Sigma(E_T)$ and $M(t^{lep})$. In the third row: E_T^{miss}.

Chapter 6

Neural Network Analysis

Using the event selection a large fraction of background can be discarded without losing a large fraction of Wt-channel signal. Nevertheless, because the cross sections of these backgrounds, especially W+jets production and $t\bar{t}$ production, are very large compared to the Wt-channel signal cross section, it is virtually impossible to obtain a pure signal sample. In the 2-jet topology, the dominating background is W+jets production; however, $t\bar{t}$ production remains the main problematic background since it overlaps with the signal in the phase space. This analysis uses an artificial neural network tool called NeuroBayes [45] in order to separate the signal from the backgrounds. Several strategies aiming at a good discrimination of the signal from the backgrounds, especially the two main backgrounds, will be discussed in detail in this chapter. The Monte Carlo samples are used for network training and the neural network output is then calculated for each Monte Carlo and collision data event to allow a comparison between Monte Carlo and data. In the first half of this chapter, an overview of discriminating variables is first given and then the general concept of neural networks is briefly described. The main part of this analysis, searching for a performant method to separate the signal from particularly the two main backgrounds, is given in the second half of the chapter.

Moreover the truth information of the Monte Carlo Wt-channel sample is used to investigate the neural network efficiency in each decay mode of the signal. In general the truth information contains details on decay-channel types of any event. This study is performed in order to get a feeling which decay mode of the Wt-channel gives the best separation power. Details on the truth information of the Wt-channel sample and its results from the neural network trainings are described in Appendix B.

6.1 Discriminating Variables

To discriminate the Wt-channel signal from the backgrounds, potential discriminating variables are required. The variables should have different shapes for the signal compared with each background, especially the two main contributions: W+jets production and $t\bar{t}$ production. The discriminating variables that are used in this analysis are considered mainly based on the kinematics of the full Wt-channel system; even though in this case one light-quark jet is missing in its final state. Furthermore, event shape variables are taken into account as well. The following information describes these variables in detail.

6.1.1 Kinematics variables of the Wt-channel system

The final state of the standard Wt-channel system consists of exactly one charged lepton, a neutrino and three quark jets including one jet which is tagged as a b-quark jet, "$b$jet" — hereafter the other quark jets are referred to as light-quark jets, "*light jet*". The charged lepton is denoted as "l". Missing transverse energy, "$E_{\mathrm{T}}^{\mathrm{miss}}$", is interpreted to as a neutrino signature. The basic kinematic variables are created based

on four momenta of particles in the *Wt* system. One important difference between the 2-jet topology and the 3-jet topology is that one light-quark jet is missing in the final state of the 2-jet topology; hence, a hadronically decaying *W* boson ,"W^{had}", that is generally reconstructed from the two light-quark jets cannot be reconstructed. A hadronically decaying top quark ,"t^{had}", is allowed, but with an incomplete reconstruction. It is reconstructed from the four-momentum of only the *b*-quark jet and a light-quark jet. On the other hand, a leptonically decaying *W* boson, "W^{lep}", is reconstructed from the four-momentum of the lepton and the $E_{\text{T}}^{\text{miss}}$, and a leptonically decaying top quark, "t^{lep}", from those variables together with the four-momentum of the *b*-quark jet. Furthermore, the reconstruction of the *Wt* system of the 2-jet topology is also allowed, but only the information of the unobserved second light-quark jet is missing from the reconstruction.

Distances between those objects of the Wt-channel system in η and ϕ together with the distance, ΔR, are also considered. The absolute values of the masses of the *W* boson and the top quark from their PDG masses are considered as well. For the final training approach described later in Section 6.6, the absolute values of the distance variables together with the mass of the light-quark jet and the *b*-quark jet are additionally taken into account.

6.1.2 Event shape variables

In order to increase the discrimination power, combinations of some variables from the kinematics of the *Wt*-channel system are considered. Some variables which are not calculated directly from the *Wt*-system kinematics are also used. These variables are classified as event shape variables. The six event shape variables that are fed into the network are listed below:

- The aplanarity of the event, *A*, defined as a measure of the transverse momentum component out of the event plane [46];
- the sphericity of the event, *S* , defined as a measure of the summed transverse momentum squared with respect to the event axis;
- the centrality of the event, *C*, defined as the sum of the p_{T} of all objects considered, over the sum of the p of the same objects;
- the missing transverse energy in the event, $E_{\text{T}}^{\text{miss}}$;
- the sum of all transverse energy in the event, $\Sigma(E_{\text{T}})$;
- the scalar sum of the transverse momenta of all jets and lepton in the event [47], H_{T}.

As described in Section 5.4, an important point of the variables which are allowed to go into the network training in order to discriminate the signal from the background have to show good agreement between Monte Carlo and data. Therefore, all discriminating variables have to be checked in the first place. A Monte Carlo to data comparison of some variables picked by the preprocessing step of NeuroBayes described later in Section 6.2.2 are shown in figure 5.1 and in Appendix A.

6.2 Neural Network Technique

6.2.1 Neural network

Since the total number of discriminating variables considered based on the *Wt* system is more than 100, and a large amount of data is available at ATLAS, an application of multivariate techniques is

necessary to use to separate this system from its backgrounds. In order to choose only most significant variables and combine such variables into one final discriminant, an artificial neural network (in the following simply denoted neural network) serves as a tool for this purpose in this thesis. The concept of the neural network technique is to replicate how the human brain works. In the neural network framework, there are nodes which are similar to neurons of the brain and the connections between these nodes. All initial information is passed to the nodes via the incoming connections and then the output value of each node is calculated. All output values are weighted and transferred to an output node via its outgoing connection in order to combine them into one final output. The weights are determined automatically by the network. In this analysis, the neural network package NeuroBayes is utilised. It is a two-layered feed-forward neural network. "Two-layered" denotes that it consists of one input layer and one intermediate layer (or hidden layer), together with one output layer. The term "feed-forward" means that the nodes are arranged in serial layers and each node obtains input from all nodes in the previous layer and transfers its output to all nodes in the next layer. In the following, a brief description of NeuroBayes based on [45, 48, 49] is presented.

6.2.2 NeuroBayes

NeuroBayes consists of a neural network and a complex robust preprocessing. It uses Bayesian statistics; therefore, it reaches a very good and fast performance, and avoids an overtraining problem of network training. There are three neural network layers; input layer, hidden layer and output layer as shown in figure 6.1. According to the NeuroBayes infrastructure, the first (input) layer contains one input node for each input variable and one bias node. These input variables are the discriminating variables described in the previous section. Users can adjust the number of hidden nodes in order to achieve good performance of network training. Since one would like to have one final result, one node in the output layer is required.

In order to obtain one final result, all variables are ranked and chosen by the preprocessing process and only the picked variables are used in the network training to separate signal from background. The distribution of the network output is in a range of [-1,1], generally -1 representing pure background and 1 is attributed to only signal. It is important that the fractions of the signal and background have to be normalised during the training process. In this analysis, the ratio of the signal to the total background is always chosen to be 50:50.

Preprocessing

All variables that might help to achieve separation of signal from background are fed into the network system and only the most significant input variables are picked by the preprocessing step. The number of variables is reduced since only the picked variables are allowed to go into the network training. In the following a brief summary of how to select those variables is presented. Firstly, each input variable is transformed non-linearly to be distributed over a range of [-1,1] to reduce the influence of outliers. In the next step, this distribution is transformed into a Gaussian distribution with standard deviation one, as a result all variables are normalised. Then all variables are ranked according to the significance of their correlation to the target which is determined automatically during this point of the preprocessing. The correlation matrix of all variables and the total correlation of the input variable set to the target together with the total significance of this correlation are calculated. In general, the total correlation to the target given in % and the total significance of this correlation given in sigma, σ, are used to judge the discriminating performance. Then one variable is removed from the input set in order to compute the loss of the total correlation to the target. The variables are sorted according to this loss of information

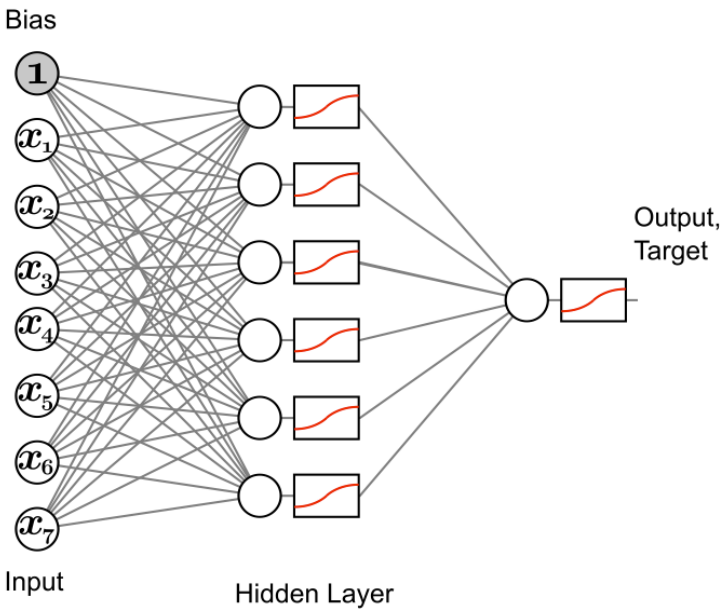

Figure 6.1: An overview of the neural network with three layer: input layer, hidden layer and output layer. There is one additional node used as a bias node in the input layer. The number of hidden nodes is chosen by users and the output layer always contains only one node.

and the variable with the smallest loss of this total correlation is discarded since it is interpreted as the least significant variable for the network training. The same procedure from calculating the correlation matrix to discarding the least important variable is repeated with the (n-1) remaining input variables until only the most significant variable remains. The significance of a variable is defined as the loss of the total correlation to the target caused by its removal at the relevant point in the serial procedure multiplied by the square root of the sample size.

In the preprocessing step, the importance of a variable is indicated by four different quantities: additional significance, single significance, significance loss and global correlation.

- **Additional significance.** The significance of a variable with the successive method described above is called the additional significance or simply the significance. This quantity is used as the indicator for the ranking of the input variables. To pick only the most significant variables, a cut on this significance is defined by users. Furthermore, the total significance of the input variable set is equal to the quadratic sum of the additional significances of all input variables.

- **Single significance.** The single significance is the total significance computed from one variable only. In case of the most significant variable, the single significance is identical to its additional significance.

- **Significance loss.** The significance loss when the variable is excluded from the input variable set. In case of the least significant variable, the significance loss is identical to its additional significance.

- **Global correlation.** The global correlation is the total correlation of the variable to all others. This quantity is calculated with the complete $n \times n$ matrix.

By definition, the additional significance, the significance loss and the global correlation depend on each other. In general, only variables having an additional significance above a threshold cut are selected; nevertheless, in some cases there are variables with lower additional significance picked by the preprocessor. This can happen since an additional significance of a particular variable depends on other variables. The higher-ranked variable with the lower additional significance shows up since if this variable is ranked out from the list, its effect is that the lower-ranked with the higher additional significance is also ranked out from the list. Therefore, the overall separation becomes worse. To sum up, in the ranking step the important point is that the last-ranked variable has to have an additional significance above a threshold cut.

Neural network training

Only variables kept by the preprocessor are used in the neural network training. Ideally between these variables should show that signal and background differ. The network learns to characterise differences signal and background during training; hence, the discrimination of the signal from the background is achieved. A weighted sum of all values of the input layer is applied to each node of the hidden layer. The bias node in the input layer is needed to activate the other input nodes for this computation. For the output node, a weight is calculated from the outputs of the hidden layer. This regularisation technique is referred to as the Bayesian regularisation scheme. The neural network uses the Bayesian regularisation in order to avoid overtraining and improve generalisation performance during the training process. Furthermore, some connections which are insignificant are pruned away since they simply cause noise during the training. Further details can be found in [50].

Performance

The overall performance of the neural network is estimated by the purity and the efficiency. Plots of these two variables as well as the total result of the training are printed out after the network training process.

Purity. The signal purity is defined as the ratio of the number of signal events to the sum of signal and background events.

Efficiency. In NeuroBayes, there are two important efficiencies: the signal efficiency and the total efficiency. The signal efficiency is defined in the normal way as the ratio of selected signal events to all signal events and the total efficiency is also defined as the ratio of selected events to all events. The so-called Lorenz-curve is the plot of the signal efficiency against the total efficiency. The efficiency curve obtained after the training normally lies above the unity line. One important parameter that is extracted directly from the Lorenz-curve is called the Gini index. The Gini index is given as the ratio of the area between the efficiency curve and the unity line to the area below the unity line. In general one can interpret the overall discrimination power achieved by the network training using the value of the Gini index. The higher the Gini index, the better separation power the training has. Also the maximum possible value of the Gini index is limited. Since the signal-to-background ratio is chosen as 50:50, the maximum possible value is 50%.

6.3 *b*-Tagging Working Point Study

There are many supported operating points of the different *b*-tagging algorithms. For the dataset taken from the end of March to the end of June 2011 with an integrated luminosity of 1035 pb^{-1}, the recommended *b*-tagger was JetFitterCombNN with three working points at 60%, 70% and 80% efficiencies corresponding to weight cuts at 2.00, 0.35 and -1.25. In order to find the best choice of the operating point for the analysis of the *Wt*-channel production in lepton+jets mode, one needs to study the performance of each *b*-tagging working point. The study of different *b*-tagging working points is carried out here. Monte Carlo samples in the 2010 Monte Carlo production round are used with the 3σ cut on significance of input variables for this purpose. These datasets are used only for this study and the optimisation study of the neural network NeuroBayes described later in Section 6.4. For only these two studies, the 3-jet topology of the *Wt*-channel production is investigated with the use of 178 variables that are fed into the preprocessing process of the neural network trainings. The signal is trained against a mixture of all backgrounds (referred to as "standard training" from here on) for all networks.

After trainings, the overall separation powers of all three cases are shown in figure 6.2 together with their corresponding used variables in Table 6.1 using the optimum hidden nodes. The details of these used variables and optimum hidden nodes will be described in the next section. The operating point with 80% efficiency gives the best separation power of 26.2%. It is very high compared to the other two values; however, one has to have a closer look on the behaviour of the distribution of the two main backgrounds to the distribution of the signal since the two main backgrounds are mainly considered. The other backgrounds are summed up into one combined sample set called as "others". For this purpose, each neural network output is explored for signal and the two main backgrounds independently in order to compare them.

Figure 6.3 shows all three outputs of signal and the two main backgrounds for electron and muon channels individually. The signal is more shifted to the right with the higher efficiency. Also the *W*+jets background is more moved to the background region with the higher efficiency. The separation of the signal from *W*+jets production is therefore achieved effectively for the network of 80% efficiency (in the bottom plots). The $t\bar{t}$ background behaves interestingly. It populates mostly in the central region and shows an overlapping part to the signal for 60% and 70% efficiencies. For the network of 80% efficiency, the $t\bar{t}$ background always overlaps with the signal tightly even in the signal region. One reason for that is that *W*+jets production is much larger than $t\bar{t}$ production for 80% efficiency as expressed in Table 6.2; consequently the network prefers to train hard to discriminate against *W*+jets production. This table presents the event yields of the signal and all backgrounds except the multijet background after the event selection for all three operating points. The higher the efficiency, the lower the ratio of the signal to all backgrounds it has. Also the sum of *W*+jets and $t\bar{t}$ events increased dramatically while the number of *Wt*-channel events dropped in case of 80% efficiency. One can simply avoid using the 80% efficiency; even though it provides best separation power from the overall background. Nevertheless, it is not clear yet to decide which *b*-tagging working point, either 60% efficiency or 70% efficiency, is appropriate for the analysis. 70% efficiency seems to be promising since it provides the better separation power from the overall background and increases the signal number; but on the other hand, the overall background is much less for 60% efficiency.

To make a conclusion of the appropriate operating point of the *b*-tagger, the limit that is determined by fitting Monte Carlo data to Monte Carlo data with an assumption of a zero signal cross section [51]. This limit can be used to compare such that the closer the limit is to 0, the better separation is achieved. In this case, only the two main backgrounds are considered. Table 6.3 presents the value of the cross section limit together with the summarised information for this study. The limit for 70% efficiency is smallest; therefore, one can interpret that the signal is most discriminated from these two

60% WP		70% WP		80% WP	
Variable	Sig.[σ]	Variable	Sig.[σ]	Variable	Sig.[σ]
$\lvert M(W^{\text{had}}) - M(W^{\text{had}}_{\text{PDG}})\rvert$	20.5	C	21.8	C	25.6
C	15.2	$\lvert M(W^{\text{had}}) - M(W^{\text{had}}_{\text{PDG}})\rvert$	17.7	$\lvert M(W^{\text{had}}) - M(W^{\text{had}}_{\text{PDG}})\rvert$	17.0
$p_{\text{T}}(W^{\text{lep}})$	7.1	$p_T(W^{\text{lep}})$	8.5	$p_{\text{T}}(W^{\text{lep}})$	10.1
$\lvert M(W^{\text{had}}) - M(W^{\text{had}}_{\text{PDG}})\rvert$	5.9	$\lvert M(t^{\text{lep}}) - M(t^{\text{lep}}_{\text{PDG}})\rvert$	7.5	$\lvert M(t^{\text{lep}}) - M(t^{\text{lep}}_{\text{PDG}})\rvert$	10.9
$E(\nu)$	5.4	$\lvert M(t^{\text{had}}) - M(t^{\text{had}}_{\text{PDG}})\rvert$	6.6	p_{T}(jet3)	8.9
$\lvert M(t^{\text{lep}}) - M(t^{\text{lep}}_{\text{PDG}})\rvert$	3.2	p_{T}(jet2)	6.6	$\lvert M(t^{\text{had}}) - M(t^{\text{had}}_{\text{PDG}})\rvert$	9.0
p_{T}(jet2)	3.2	$E(\nu)$	5.0	$\Delta\phi(W^{\text{had}}$,jet2)	1.9
η(jet1)	3.1	$E^{\text{miss}}_{\text{T}}$	5.5	$\Delta\phi$(jet2,jet1)	8.9
ΔR(jet1,ν)	2.8	$\Delta R(Wt, t^{\text{had}})$	2.3	$E^{\text{miss}}_{\text{T}}$	6.2
$\phi(b$Jet)	2.9	$p_{\text{T}}(Wt)$	3.7	$E(\nu)$	7.7
$\Delta R(Wt, b$Jet)	3.2	η(jet1)	3.5	$p_{\text{T}}(b$Jet)	3.9
		ΔR(jet1,bJet)	3.2	E(jet2)	5.1
		p_{T}(jet3)	2.9	$\Delta R(W^{\text{had}}$,jet1)	4.9
		$M(t^{\text{had}})$	2.5	$\Delta R(Wt, t^{\text{had}})$	4.4
		ΔR(jet2,jet1)	2.7	η(jet1)	3.6
		$\Delta\phi$(jet2,jet1)	3.0	p_{T}(jet1)	3.4
		$p_{\text{T}}(b$Jet)	2.6	$M(t^{\text{had}})$	3.1
		p_{T}(jet1)	3.0	p_{T}(jet2)	3.3
		A	3.1	ΔR(jet2,bJet)	3.2
		$\Delta\phi(Wt, b$Jet)	2.2		
		$\Delta\phi(Wt, W^{\text{had}})$	2.2		
		$\Delta\eta(Wt, t^{\text{had}})$	3.1		

Table 6.1: The most significant input variables in the 3-jet topology for the three different b-tagging working points. They are ranked according to their (additional) significance. The results are only for the 3-jet topology with a significance cut of 3σ. In the 3-jet topology, there are 2 jets denoted as jet1 and jet2 together with W^{had} in this table. Jet1 is the hardest light-quark jet and the jet2 is the second-hardest light-quark jet. W^{had} denotes the hadronically decaying W boson.

main backgrounds compared to the other working points. Together with the reasons discussed above, the most appropriate solution is the 70% working point.

Therefore, the 70% working point is decided to be the appropriate choice for the analyses of Wt-channel production; although the MV1 tagger was recommended to be used instead of the JetFitterCombNN tagger for the 2011 data. Both MV1 and JetFitterCombNN taggers are suitable for multivariate analyses and they are the combination of JetFitterTagNN and IP3D taggers (plus SV1 tagger for the MV1 tagger). The b-tagging working point at 70% efficiency of the MV1 tagger is performant to be used.

6.4 Optimisation of the Neural Network NeuroBayes

Some basic studies of NeuroBayes are discussed in this section. They are the relation between the used variables and the Gini index and the hidden node optimisation. Results of these studies were used in the b-tagging working point study above and all concepts will be exploited for all further studies as well.

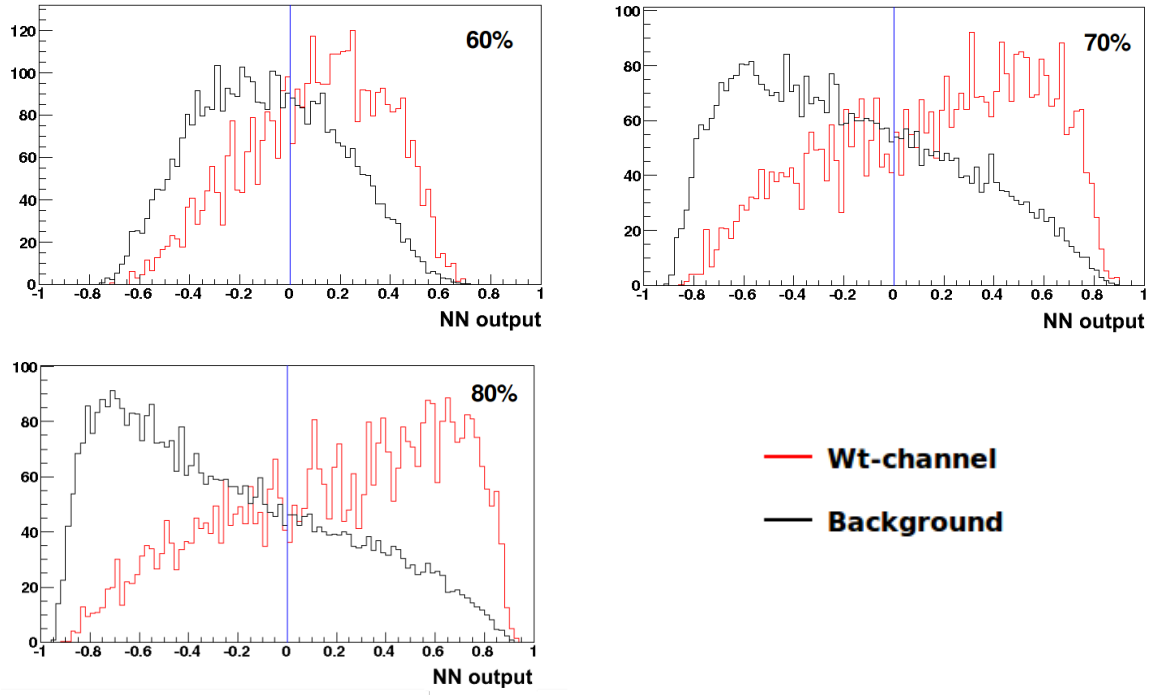

Figure 6.2: Neural network outputs in **the 3-jet topology** for all three *b*-tagging working points using the optimised number of hidden nodes for each case. All distributions are expressed in arbitrary units and the *Wt*-channel signal and all backgrounds are normalised to be 50:50. In the first row: the neural network output for 60% efficiency with a Gini index of 19.6% using 11 variables and 16 hidden layers (left). The neural network output for 70% efficiency with a Gini index of 22.4% using 22 variables and 24 hidden layers (right). In the second row: the neural network output for 80% efficiency with a Gini index of 26.2% using 19 variables and 25 hidden layers.

6.4.1 Used variables and the Gini index

In the preprocessing step, the variables are ranked according to their (additional) significances as described in Section 6.2.2. One important task of this analysis is to find the minimum number of used variables without losing too much separation power. The more variables are used in the network training, the more the output discriminant may suffer from systematic uncertainty. Even if systematic uncertainty is outside of the scope of this thesis, one should always consider this issue. Furthermore, all used variables have to show agreement between the distributions of Monte Carlo and data samples. Basically the significance of 3.0σ is used to indicate an evidence of any signal; hence, 3.0σ is used as the threshold significance cut for those variables. One important benchmark for the discrimination power of signal from background is the Gini index. Therefore, this study is performed by investigating the value of the Gini index after changing the significance cuts since the number of used variables depends upon the σ-cut which can be adjusted manually. Technically, the σ-cut, can be chosen in steps of 0.5σ.

After the preprocessing step, the number of used variables is 11, 22 and 19 with a Gini index of 19.6%, 22.4% and 26.2% for the working points of 60%, 70% and 80%, respectively. Table 6.1 shows those used variables and their significances with the significance cut of 3σ. One remarkable point is that some variables with a lower additional significance are also chosen. As discussed above in Section 6.2.2, the (additional) significance of one variable is related to the other variables. If those lower-significant variables are discarded, the overall separation power will be decreased. That is why they are picked into

	b-tagging efficiency		
	60%	70%	80%
Wt-channel	400	440	440
Other single tops	510	540	510
$t\bar{t}$ production	3450	3410	2950
W+jets (all)	2870	4880	9780
Diboson	60	90	170
Z+jets	130	280	780
Sum of background	7020	9200	14190
Sum of MC	7420	9640	14630
Data	7760	9930	15040

Table 6.2: Event yields of the 3-jet topology excluding any uncertainty of the signal and its backgrounds for the b-tagging working point study. All backgrounds are shown except multijet since the main point of this comparison is the two main backgrounds: $t\bar{t}$ production and W+jets production. Therefore, the result shows a discrepancy between the sum of all Monte Carlo samples and data.

	b-tagging efficiency		
	60%	70%	80%
Gini index [%]	19.6	22.4	26.2
Limit [pb]	10.8	10.2	11.4
$s/b_{\text{(all)}}$ [%]	5.6	4.8	3.1
$s/b_{(t\bar{t}\ \&\ W\text{+jets})}$ [%]	6.3	5.3	3.4

Table 6.3: Summary of the b-tagging working point study. The information includes the Gini index, cross section limit, signal-to-all backgrounds ratio and signal-to-two main backgrounds ratio. Ideally the best performance of the b-tagging working point has to achieve a high value of Gini index, a low value of the cross section limit and a high value of both ratios. The 70% working point therefore gives the most appropriate choice for the analysis.

the network training. After a harder sigma cut of 3.5σ is applied to all three working points. Only the first 5 variables in the table of 60% and the first 11 variables of 70% are kept for the training process but the Gini indices drop to 16.0% and 20.3%, respectively. For 80% the first 15 variables are kept with a Gini index of 24.7%. For 70%, the number of the kept variables is 8 with the Gini indix of 19.5% using 4.0σ-cut. For 80%, the number of the kept variables is 14 with the Gini indix of 24.4% using 4.0σ-cut. After a sigma cut of 6.0σ is applied, the number of kept variables is 3 with a Gini index of 11.7% for 60%. All of these Gini indices are provided by training the networks using the optimised number of hidden nodes in the hidden layer described later in Section 6.4.2.

As a result of this study, one keeps the appropriate number of used variables which indicates the less decreasing of the Gini index. Hence, the variables kept for the network training in these three cases should be 11, 11 and 14 with the sigma cut of 3.0σ, 3.5σ and 4.0σ, respectively.

6.4.2 Hidden node optimisation

To investigate the effect of the number of hidden nodes on the discrimination power between the signal and the overall background, the neural networks of the three different b-tagging working points are

retrained with the number of the hidden nodes ranging from 1 to 40. After the preprocessing step, only the most significant input variables with at least 3σ are kept for each neural network training. A number of 11, 22 and 19 variables are chosen as the most significant input variables for the efficiencies of 60%, 70% and 80%, respectively as described in the previous study. The consequences of the network trainings for this study are that the Gini index is found in a range of 18.9-19.6% for 60% efficiency, in a range of 21.8-22.4% for 70% efficiency and in a range of 25.5-26.2% for 80% efficiency. The interesting point is the number of hidden nodes which gives the highest value of the Gini index is always found to be a bit larger than the number of used variables. Increasing the number of hidden nodes also makes the network training process longer. To reduce the time consumption of scanning hidden-node space in order to obtain the largest value of the Gini index, only the hidden nodes near the number of the used variables are scanned. And, the best choice for choosing the number of hidden nodes is the one in which gives the first maximum of the Gini index. In these three cases, the optimised solution of the number of the hidden nodes is 16, 24 and 25, respectively. Figure 6.2 shows the neural network outputs of the signal and the overall background for these three *b*-tagging working points using the 3-sigma cut along with the optimised number of the hidden nodes for each efficiency. These neural network outputs combine electron and muon channels together; hence, one can get a first impression of how well the separation of the signal from all backgrounds works using these outputs.

6.5 Separate-Training Technique

Hereafter the dataset with an integrated luminosity of 4.7 fb^{-1} and the Monte Carlo samples of the 2011 Monte Carlo production round are used. Only the analysis of the 2-jet topology is presented. 122 variables are used in the preprocessing step. The network training of the "standard approach" does not achieve a good separation of the signal from the $t\bar{t}$ background (see the overlay plots of figure 6.4). 13 used variables are presented in the first column of Table 6.4 using 14 hidden nodes. The separation of the signal from the overall backgrounds is high. The network achieves a Gini index of 25.1%. However, considering each background separately, $t\bar{t}$ production always overlaps with the signal while discrimination against W+jets production and "others" is achieved. Moreover $t\bar{t}$ production is more shifted to the signal region than the signal itself for both the electron and muon channels. These two issues obviously bring the analysis of the 2-jet topology to a critical point; therefore, new approaches are required.

A Monte Carlo to data comparison of the neural network output for each channel is presented in the right column of figure 6.4. For the electron channel (top plot), the overall distribution of the collision data is well described by the Monte Carlo samples; even though some small variations can also be seen. For the muon channel, a larger discrepancy appears, especially in the range -0.8 to -0.5. One reason for that is the use of the jet-electron model for the multijet estimation. The four largest discrepancy bins are definitely in the enriched region of the multijet background. But considering the whole distributions, one can see good agreement between Monte Carlo and data.

The standard approach cannot provide good separation from all backgrounds at the same time since the network mostly separates the signal from the largest background. That is one reason why separating the signal from the $t\bar{t}$ background is difficult. Also the $t\bar{t}$ background and the signal might accumulate in the similar region; therefore, "others" is also shifted to the background region. For the first point, one can try out by training the network against each background separately then combine all results into one final discriminant. This approach is named as "separate-training approach", aiming for better separation against each background.

Four independent networks in the 2-jet topology are trained for this study. For three networks the

Wt signal is trained against each of three backgrounds, for one network $t\bar{t}$ production is trained against W+jets production. The neural network outputs of these four networks are further provided to a second network training as the discriminating variables. Only the absolute value of the neural network output of the $t\bar{t}$ and W+jets network is used. Even though, neither $t\bar{t}$ production nor W+jets production is the signal, using such a network could help to separate the signal from the $t\bar{t}$ background. The absolute value shows that the signal populates the central region and $t\bar{t}$ production and W+jets production move to the right region. Moreover only passing the first three independent network outputs into the second training does not achieve the better separation as described in detail later.

Table 6.4 shows the used variables and their significances compared to the target for all four primary networks. The network of the signal and W+jets production uses 12 variables with 13 hidden nodes and provides a Gini index of 31.1%. The network of the signal and $t\bar{t}$ production uses 17 variables with 17 hidden nodes but provides a Gini index of only 16.8%. The network of the signal and "others" uses 14 variables with 14 hidden nodes and provides a Gini index of 27.2%. Furthermore, the network of $t\bar{t}$ and W+jets production uses 15 variables with 15 hidden nodes and provides a Gini index of 36.3%. As one would expect, the networks of the signal trained against W+jets production and against "others" show large separation. On the other hand, the network of signal trained against the $t\bar{t}$ background again achieves the lowest separation power even though the number of discriminating variables which are used in the network training is highest. For the $t\bar{t}$ and W+jets network, the best separation power among all four networks is achieved. One remark is that the p_{T} of the Wt system always appears as the best variable for all networks, except for the signal and "others" network, where it still shows up as the eighth-best variable.

Each neural network output is then calculated for both signal and all three backgrounds in order to compare their shape distributions closely. Figure 6.5 shows the overlay plots of the neural network outputs for each network. The electron and muon channels are plotted separately. Only the network of the signal trained against $t\bar{t}$ production shown in the 1$^{\mathrm{st}}$ row achieves a good separation of the signal from the $t\bar{t}$ background. Furthermore, the overlap between the signal and $t\bar{t}$ production is observed with the use of the network trained against W+jets production and "other". For the $t\bar{t}$ against W+jets network, the absolute value of the neural network output is used since the shape distribution of the signal differs from the shape distribution of $t\bar{t}$ production. All four networks show that the separation between the signal and W+jets background is most effective. Also it is possible to discriminate "others" from the signal.

In order to combine those four network outputs into one final discriminant, the second training of the network using those four outputs as the input variables is performed. As a result, all of these variables are picked by the training process with a significance greater than 3σ as shown in the first column of Table 6.5. This network provides a Gini index of 25.0% using 4 hidden nodes in the hidden layer. Even though the overall separation power is comparable to the result of the standard training, this network provides a small improvement of the separation between the signal and the $t\bar{t}$ background as shown in the left column of figure 6.6. The problem that $t\bar{t}$ production is shifted to the signal region is improved by this approach in both electron and muon channels. Unfortunately, it is still hard to separate the signal from $t\bar{t}$ production and the total number of original discriminating variables used in this approach is high, 35. The significance of the analysis can be destroyed easily when systematic uncertainties are considered. The comparison between the Monte Carlo and collision data in electron and muon channels for this network output are shown in the right column. A discrepancy between the Monte Carlo and data is visible in both channels. Fortunately combining both channels shows that the distribution of collision data can be described by Monte Carlo samples.

As stated above the second training using three independent network outputs of signal trained against each of three backgrounds does not help the separation from $t\bar{t}$ production. Here the result of the second

Wt-channel vs. all backgrounds		Wt-channel vs. W+jets		Wt-channel vs. $t\bar{t}$		Wt-channel vs. others		W+jets vs. $t\bar{t}$	
Variable	Sig. [σ]	Variable	Sig. [σ]	Variable	Sig. [σ]	Variable	Sig. [σ]	Variable	Sig. [σ]
$p_{\mathrm{T}}(Wt)$	47.1	$p_{\mathrm{T}}(Wt)$	61.0	$p_{\mathrm{T}}(Wt)$	59.5	$M(t^{\mathrm{had}})$	56.9	$p_{\mathrm{T}}(Wt)$	146.5
$\phi(\nu)$	48.5	$p_{\mathrm{T}}(b\mathrm{Jet})$	54.6	$M(t^{\mathrm{had}})$	26.5	$E_{\mathrm{T}}^{\mathrm{miss}}$	49.2	$E_{\mathrm{T}}^{\mathrm{miss}}$	79.2
$p_{\mathrm{T}}(b\mathrm{Jet})$	44.9	$\phi(\nu)$	49.2	$\eta(b\mathrm{Jet})$	21.5	$\eta(\mathrm{lightJet})$	42.3	$p_{\mathrm{T}}(b\mathrm{Jet})$	67.7
$M(t^{\mathrm{had}})$	35.9	$E_{\mathrm{T}}^{\mathrm{miss}}$	32.8	$E_{\mathrm{T}}^{\mathrm{miss}}$	20.6	$\mathrm{H_T}$	37.8	$\eta(l)$	45.8
$\Delta\eta(b\mathrm{Jet,lightJet})$	29.2	$M(t^{\mathrm{had}})$	33.7	$p_{\mathrm{T}}(W^{\mathrm{lep}})$	6.6	$M_{\mathrm{T}}(W^{\mathrm{lep}})$	32.2	$p_{\mathrm{T}}(\mathrm{lightJet})$	43.6
$M_{\mathrm{T}}(W^{\mathrm{lep}})$	28.1	$\Delta\eta(b\mathrm{Jet,Jet})$	29.5	$\Delta\phi(\nu, l)$	16.9	$\Delta\phi(\nu, b\mathrm{Jet})$	21.7	$\Delta R(b\mathrm{Jet,lightJet})$	40.2
$\Delta\phi(b\mathrm{Jet,lightJet})$	22.2	$\mathrm{H_T}$	23.3	$\mathrm{H_T}$	11.5	$\phi(\nu)$	18.2	$\phi(\nu)$	28.1
$p_{\mathrm{T}}(\mathrm{lightJet})$	22.9	$M(W^{\mathrm{lep}})$	22.6	$\phi(\nu)$	12.8	$p_{\mathrm{T}}(Wt)$	18.6	$\Delta\eta(b\mathrm{Jet,l})$	24.4
$E_{\mathrm{T}}^{\mathrm{miss}}$	18.0	$p_{\mathrm{T}}(\mathrm{lightJet})$	22.1	$\Delta\eta(Wt, \nu)$	8.1	$Q(l)$	18.6	$M_{\mathrm{T}}(W^{\mathrm{lep}})$	22.4
$\lvert M(t^{\mathrm{lep}}) - M(t^{\mathrm{lep}}_{\mathrm{PDG}})\rvert$	16.4	$\Delta\phi(b\mathrm{Jet,Jet})$	21.2	$\Delta R(Wt, \nu)$	12.2	$\Delta\eta(b\mathrm{Jet,Jet})$	16.5	$\lvert M(t^{\mathrm{lep}}) - M(t^{\mathrm{lep}}_{\mathrm{PDG}})\rvert$	20.3
$\eta(\mathrm{lightJet})$	15.4	$\Delta\eta(l, t^{\mathrm{had}})$	13.6	$\eta(\mathrm{lightJet})$	9.6	$p_{\mathrm{T}}(b\mathrm{Jet})$	10.8	$M(t^{\mathrm{had}})$	19.7
$\Delta\eta(l, t^{\mathrm{had}})$	11.6	$p_{\mathrm{T}}(\nu)$	10.1	$\Delta R(Wt, b\mathrm{Jet})$	8.4	$E(\mathrm{lightJet})$	8.9	$\Delta\phi(Wt,\mathrm{lightJet})$	10.7
$Q(l)$	10.1			$\eta(Wt)$	6.1	$p_{\mathrm{T}}(l)$	8.3	$\Delta\phi(Wt, b\mathrm{Jet})$	14.0
				$\Delta\phi(Wt, \nu)$	6.9	$\Delta R(t^{\mathrm{had}}, b\mathrm{Jet})$	10.9	$p_{\mathrm{T}}(W^{\mathrm{lep}})$	12.8
				$p_{\mathrm{T}}(\mathrm{lightJet})$	6.8			$p_{\mathrm{T}}(t^{\mathrm{had}})$	17.1
				$\Delta R(t^{\mathrm{lep}}, \nu)$	6.2				
				$p_{\mathrm{T}}(l)$	6.8				

Table 6.4: Discriminating variables used in the network training for the separate-training technique and their (additional) significances. The variables in the 1st column are from the training of signal against all backgrounds. The variables in the 3rd column are from the training of signal against W+jets production. The variables in the 5th column are from the training of signal against $t\bar{t}$ production. The variables in the 7th column are from the training of signal against "others". The variables in the the 9th column are from the training of W+jets production against $t\bar{t}$ production.

Combined training with 4 input variables		Combined training with 3 input variables	
Variable	Significance [σ]	Variable	Significance [σ]
NN output Wt vs. W+jets	105.7	NN output Wt vs. W+jets	105.7
NN output Wt vs. others	20.8	NN output Wt vs. others	20.8
NN output Wt vs. $t\bar{t}$	3.2	NN output Wt vs. $t\bar{t}$	3.2
NN output W+jets vs. $t\bar{t}$	4.7		

Table 6.5: Discriminating variables used in the combined training of the separate-training approach and their (additional) significances. Used variables for the combined training with four input variables are presented in the first column and used variables for the combined training with three input variables in the second column.

training with these three networks is summarised briefly. Figure 6.7 shows the 2$^{\text{nd}}$ neural network output for the electron and muon channels separately. The network provides a Gini index of 24.9% with 5 hidden nodes. The used variables of this second training network are presented in the second column of Table 6.5. Separating the signal from $t\bar{t}$ production is difficult and the $t\bar{t}$ distribution is more signal-like than Wt production. In total, 28 variables from the discriminating variable set are used in this case.

In summary, the final result from the second layer of training using all four separate network outputs shows that the $t\bar{t}$ background still overlaps with the signal. One reason for that is that the cross section of the $t\bar{t}$ background in the 2-jet topology is small compared to the cross section of the W+jets background so that the network training performs less to separate the $t\bar{t}$ background from the signal. Also the signal and $t\bar{t}$ production might stay in a similar region. The total number of used variables is rather high, that might cause the analysis to suffer from systematic uncertainty. This approach is not good enough to further use since the $t\bar{t}$ background is still a problematic issue. Therefore, other approaches are needed.

6.6 Two-Dimensional Technique

To discriminate the signal from $t\bar{t}$ production and still provide a good discrimination against W+jets production, another approach has to be used. A two-dimensional distribution is made from the neural network outputs of the signal trained against $t\bar{t}$ production and against W+jets production since they are the largest backgrounds. A histogram with 10×10 bins is filled with the neural network outputs of these two networks for both the signal and each of backgrounds to produce the two-dimensional distributions. An advantage of this technique is that information of the two main background networks are used in both signal and all backgrounds at the same time. From these distributions the occupied region of the signal and each background is directly visible.

In this stage the mass of each jet and the absolute value of the distances between two objects are also included in the training process. In total 141 variables are fed into the preprocessing step. The variables used in the network training of the signal and W+jets network and of the signal and $t\bar{t}$ production network are presented in Table 6.6. For the signal and W+jets network, a Gini index improved to 32.1% with 11 used variables and 12 hidden nodes. For the network of the signal and $t\bar{t}$ production, a Gini index increased to 17.5% and only 13 variables are used. In this case 15 hidden nodes are performed. One can see that this set of variables can provide better separation from W+jets production and $t\bar{t}$ production, compared to the values of the Gini indices of the previous training in Section 6.5. The variables of $M(b\text{Jet})$ and $M(\text{lightJet})$ together with the absolute values of $\Delta\eta(\text{lightJet}, b\text{Jet})$ and $\Delta\eta(Wt, \nu)$ are picked by the preprocessing step. Moreover the p_{T} of the Wt system is still the first variable that is picked into the network training.

Wt-channel vs. W+jets		Wt-channel vs. $t\bar{t}$	
Variable	Significance [σ]	Variable	Significance [σ]
$p_{\mathrm{T}}(Wt)$	55.0	$p_{\mathrm{T}}(Wt)$	66.1
$M(t^{\mathrm{had}})$	24.3	$M(t^{\mathrm{lep}})$	39.3
$E_{\mathrm{T}}^{\mathrm{miss}}$	19.1	$\eta(l)$	38.1
$\lvert\Delta\eta(Wt,\nu)\rvert$	18.0	$M(t^{\mathrm{had}})$	35.7
$\Delta R(Wt,\nu)$	18.4	$E_{\mathrm{T}}^{\mathrm{miss}}$	32.2
$\eta(l)$	16.4	$\lvert\Delta\eta(\mathrm{lightJet}, b\mathrm{Jet})\rvert$	29.6
$p_{\mathrm{T}}(l)$	14.9	$M(b\mathrm{Jet})$	23.9
$M(\mathrm{lightJet})$	13.4	$p_{\mathrm{T}}(\mathrm{lightJet})$	20.0
$\eta(b\mathrm{Jet})$	10.6	$\Delta\eta(b\mathrm{Jet},W^{\mathrm{lep}})$	18.7
$\eta(\mathrm{lightJet})$	6.6	$M_{\mathrm{T}}(W^{\mathrm{lep}})$	13.8
$p_{\mathrm{T}}(\mathrm{lightJet})$	8.0	$\Delta\eta(l,t^{\mathrm{had}})$	10.9
		$\eta(Wt)$	10.8
		$M(\mathrm{lightJet})$	10.1

Table 6.6: Discriminating variables used in the network training for the 2D distribution strategy and their (additional) significances. All used variables in the first column are from training the signal against W+jets production. All used variables in the second column are from training the signal against $t\bar{t}$ production.

The two-dimensional distributions of the neural network outputs of the signal against W+jets production and of the signal against $t\bar{t}$ production for the signal and all backgrounds are presented in figure 6.8. Electron and muon channels are plotted separately. The left-column plots show the distributions for the electron channel and the right-column plots for the muon channel. In the first two plots, the signal region is located at higher values of the discriminant for both channels. The similarity between the $t\bar{t}$ background and the signal is visible in the second two plots. The $t\bar{t}$ background accumulates in the signal region. The other two backgrounds are indeed background-like. They are shifted to lower value of the discriminants; therefore, separating W+jets production and "others" from the signal is again more effective.

In order to use the information of these two-dimensional distributions as one discriminant for separating the Wt-channel signal from the backgrounds, the one-dimensional distributions from unrolling these two-dimensional distributions are presented. Each two-dimensional histogram with 10×10 bins is filled with the two network outputs; hence, it is unrolled row by row from the bottom to the top to create its corresponding one-dimensional histogram from the first bin to the tenth bin. The one-dimensional distributions are more appropriate for creating overlay plots to get an impression of how well the separation of the signal from each background is and making a comparison between Monte Carlo and data. The one-dimensional distributions are technically used as templates to extract the signal which is described later in Chapter 7. Since the combined data of the electron and muon channels is used to extract the signal, only the one-dimensional distributions of the combined data are presented in figure 6.9.

The overlay of the signal and the two main backgrounds is shown in the upper plot of figure 6.9. Each distribution is normalised individually to unity. One can see directly that the W+jets background is mostly filled in the first three bins and then drops rapidly in the signal region. The Wt-channel signal populates mostly the seventh to ninth bin. The signal is again separated effectively from W+jets production. On the other hand, $t\bar{t}$ production populates the higher bins, but one can see that better separation between the signal and $t\bar{t}$ production shows in every single bin; the signal is shifted to the right while the $t\bar{t}$ background is moved more to the left. Furthermore, the total number of used discriminating variables

is only 18, compared to 35 for the separate-training approach. Therefore, this approach provides the best result of separation the signal from the $t\bar{t}$ background as well as from the W+jets background. A Monte Carlo to data comparison is presented in the bottom plot. Overall the data is well described by the Monte Carlo. A large discrepancy between these two datasets shows up only in the first bin.

In summary, a discrimination of signal from $t\bar{t}$ production is achieved using the two-dimensional approach. This approach still provides very good separation power against the W+jets background and the sum of other backgrounds. Therefore, its discriminant is further used to extract the Wt signal as detailed in Chapter 7.

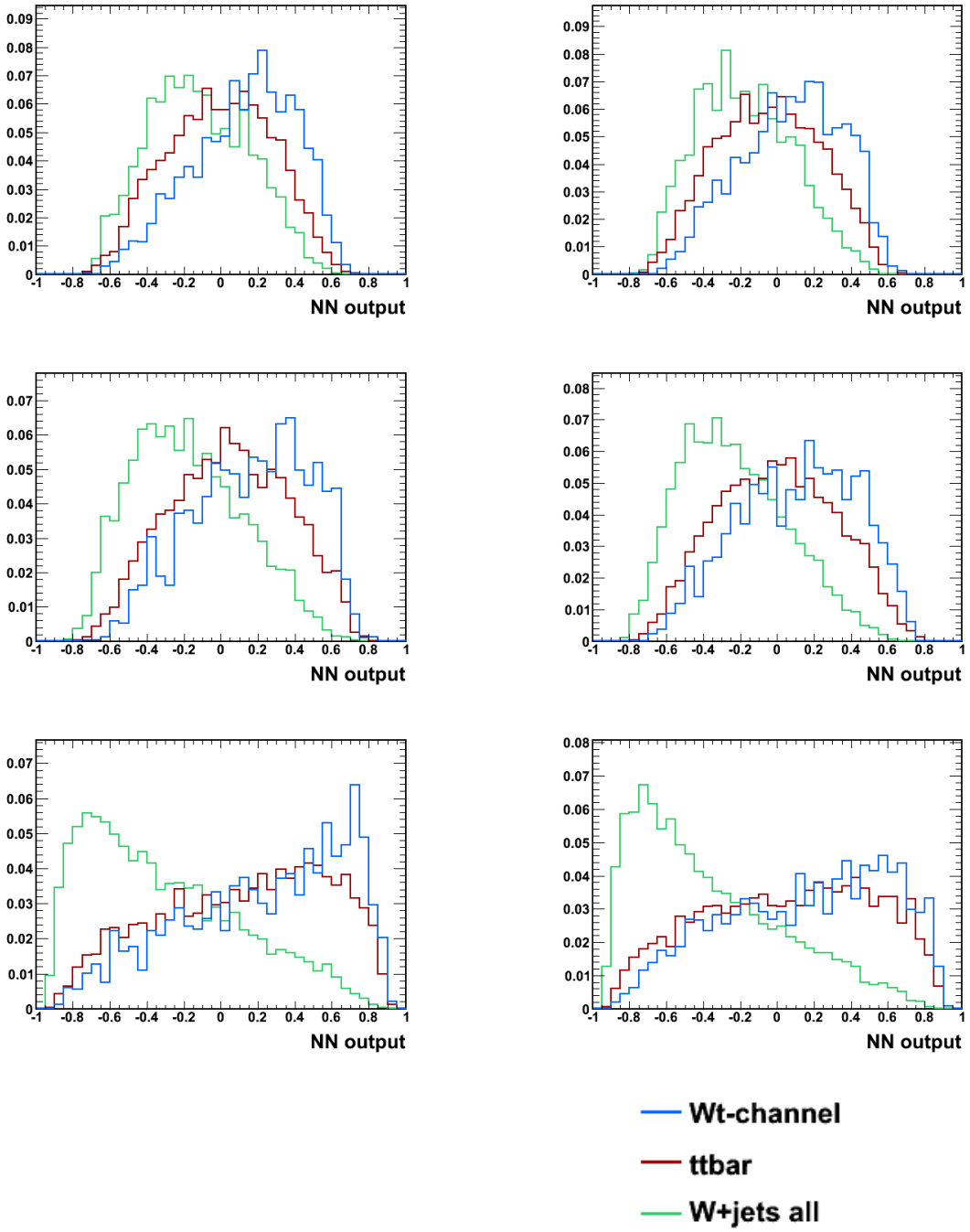

Figure 6.3: Neural network outputs in **the 3-jet topology** for b-tagging working point study. All distributions are expressed in arbitrary units. The two main backgrounds: $t\bar{t}$ production and W+jets production are considered separately. The network outputs of the overall background are shown in figure 6.2. The left column shows the network outputs for the electron channel and the right column is for the muon channel. In the first row: the neural network outputs for 60% efficiency with a Gini index of 19.6% and a cross section limit of 10.8 pb. In the second row: the neural network outputs for 70% efficiency with a Gini index of 22.4% and a cross section limit of 10.2 pb. In the third row: the neural network outputs for 80% efficiency with a Gini index of 26.2% and a cross section limit of 11.4 pb.

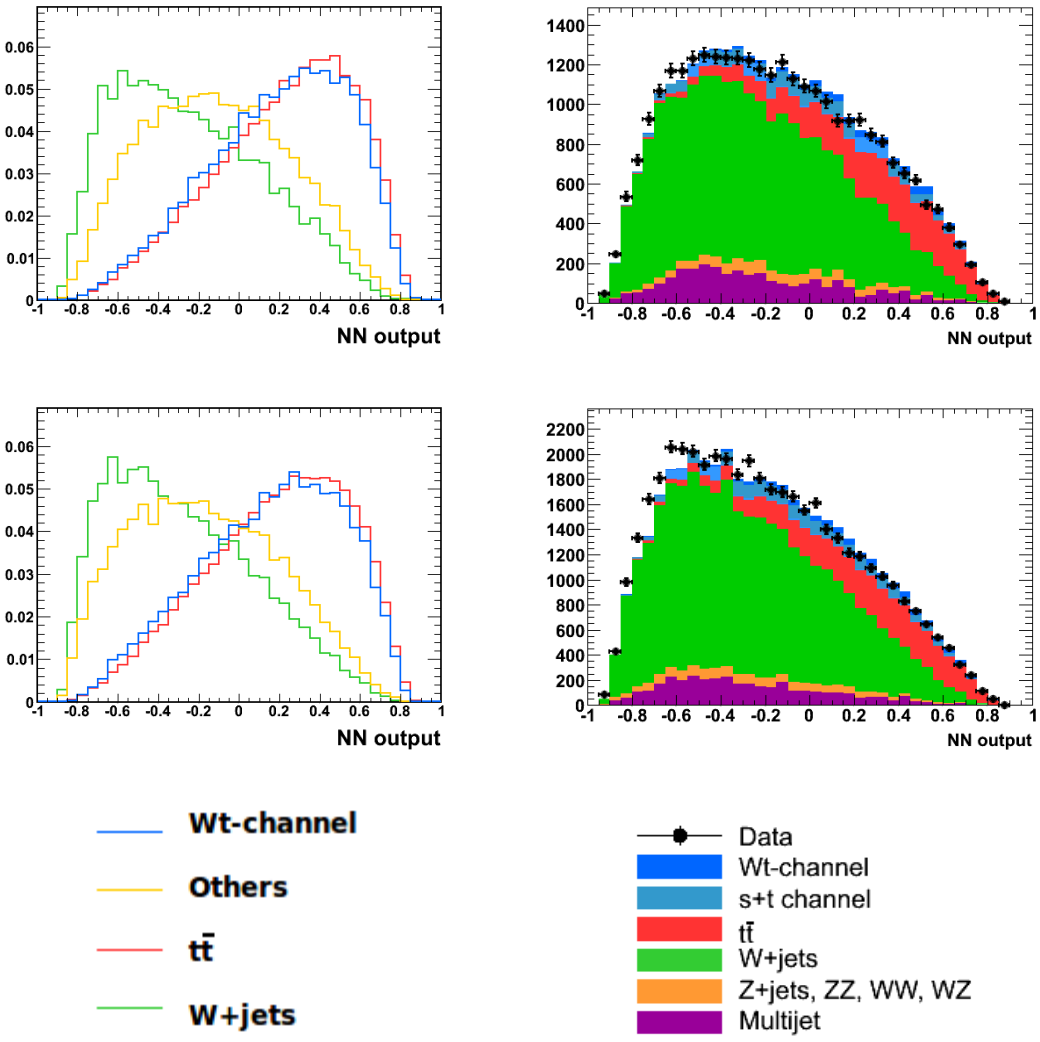

Figure 6.4: The neural network outputs in **the 2-jet topology** for the standard approach. In the first row: the result for the electron channel. In the second row: the result for the muon channel. The normalised overlay plots of the signal and three kinds of the backgrounds that are W+jets production, $t\bar{t}$ production and the rest of the backgrounds referred to as "others" are shown in the left column. The stack plot of each channel is in the right column.

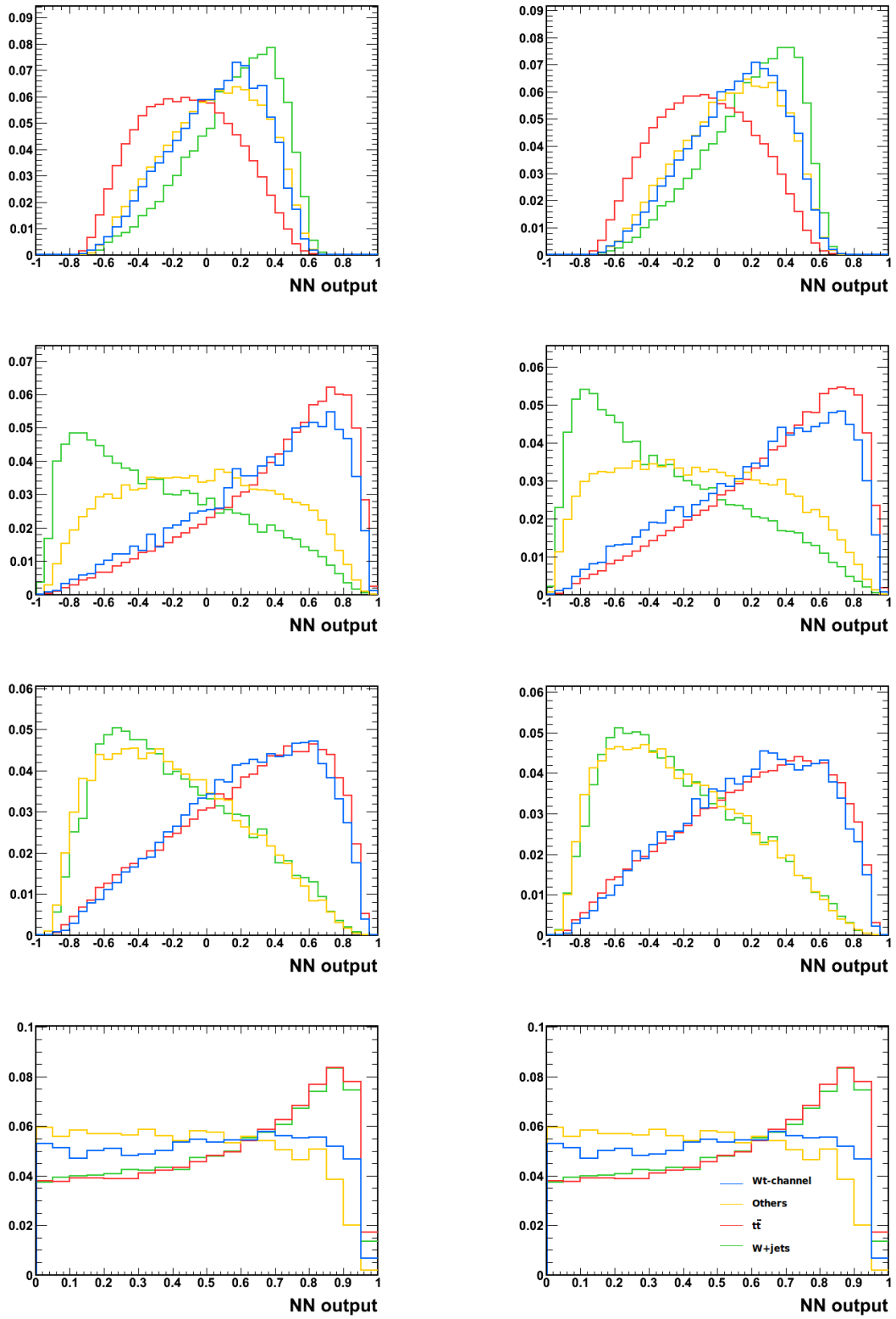

Figure 6.5: The neural network (NN) outputs for the separate-training approach. All backgrounds are presented separately. The left column is for the electron channel and the right column is for the muon channel. The 1st row is the NN output from training the signal against $t\bar{t}$ production. The 2nd row is the NN output from training the signal against W+jets production. The 3rd row is the NN output from training the signal against "others". The 4th row is the absolute value of the NN output from training W+jets against $t\bar{t}$ production.

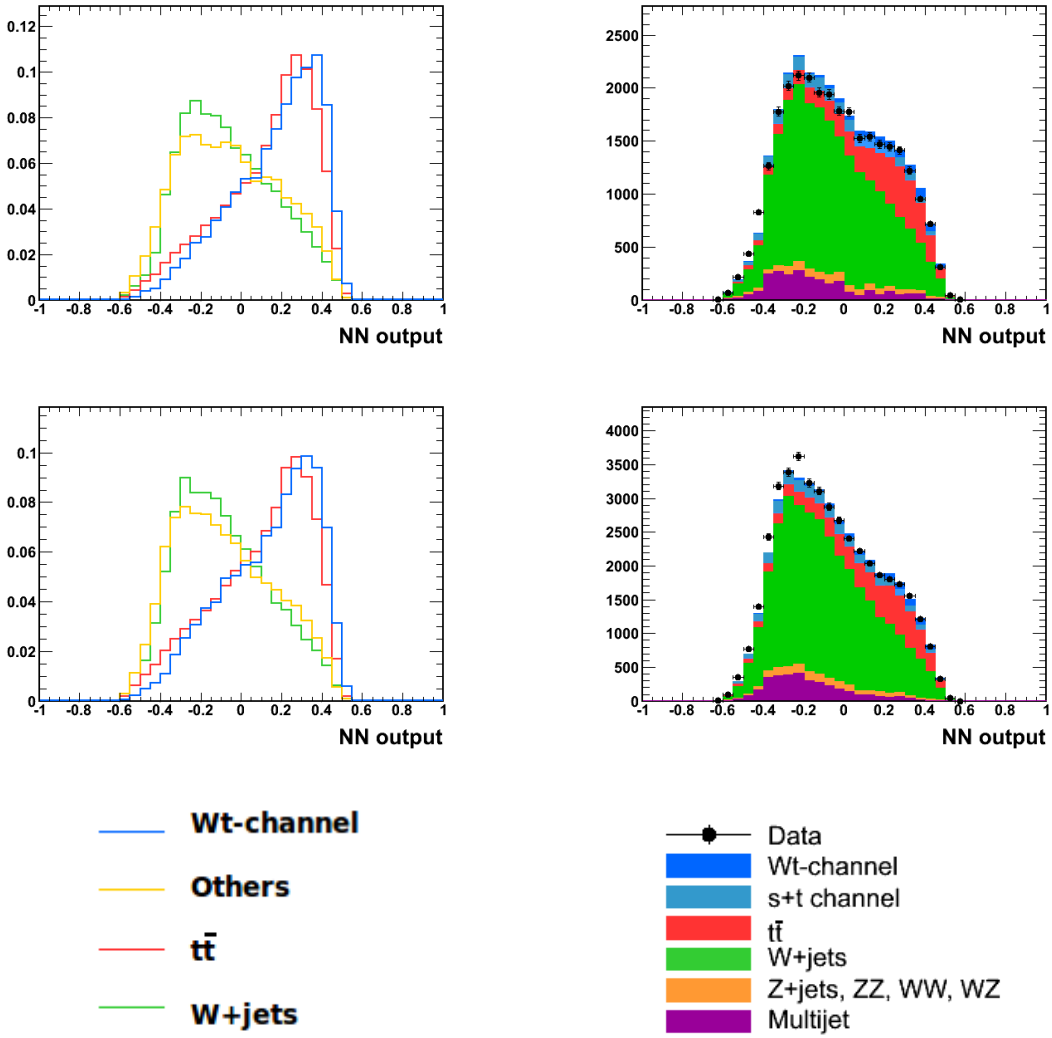

Figure 6.6: The combined outputs using all four network outputs as the input variables. In the first row: the result for the electron channel. In the second row: the result for the muon channel. The normalised overlay plots of the signal and three kinds of the backgrounds that are W+jets production, $t\bar{t}$ production and the rest of the backgrounds referred to as "others" are shown in the left column. The stack plot of each channel is in the right column.

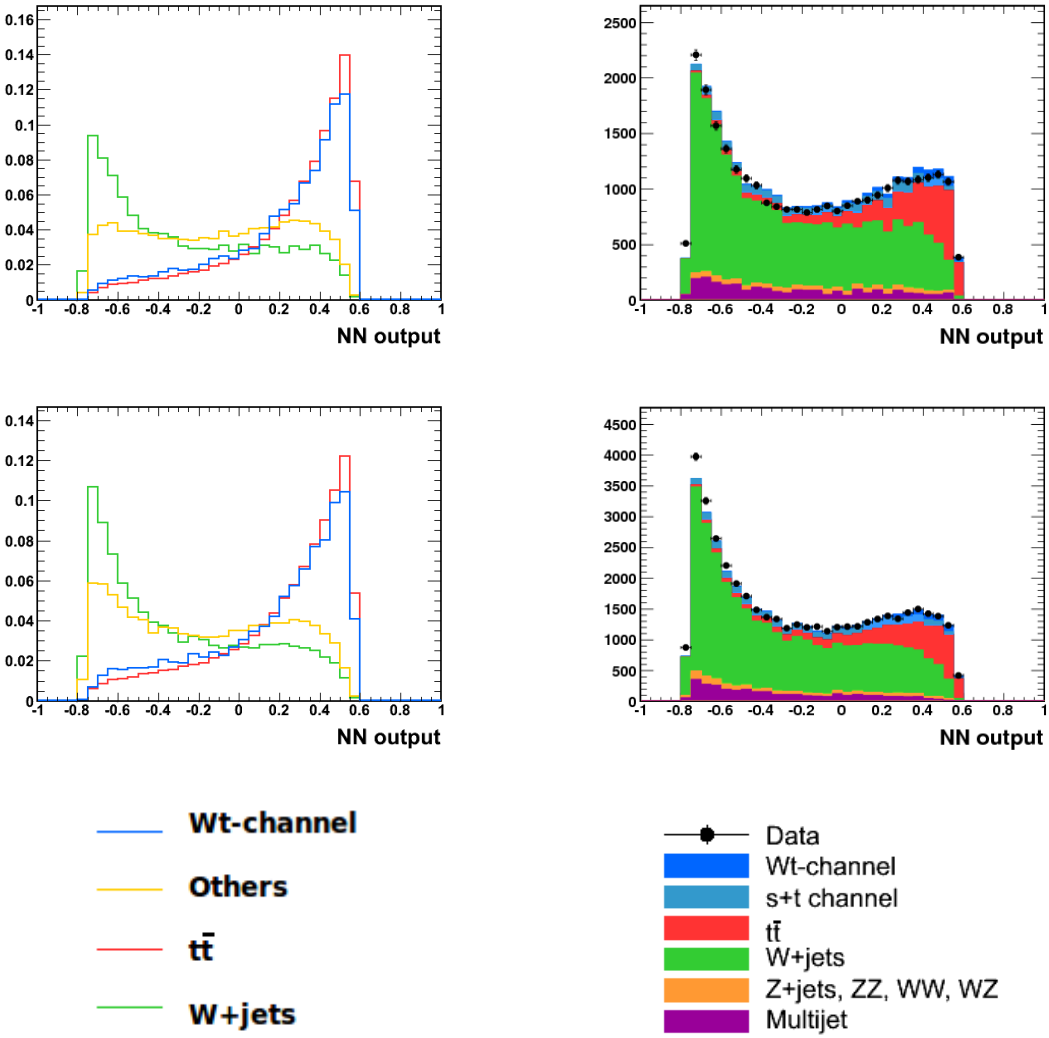

Figure 6.7: The combined outputs using the three network outputs of training the signal against each of three backgrounds as the input variables. In the first row: the result for the electron channel. In the second row: the result for the muon channel. The normalised overlay plots of the signal and three kinds of backgrounds that are W+jets production, $t\bar{t}$ production and the rest of the backgrounds referred to as "others" are shown in the left column. The stack plot of each channel is in the right column.

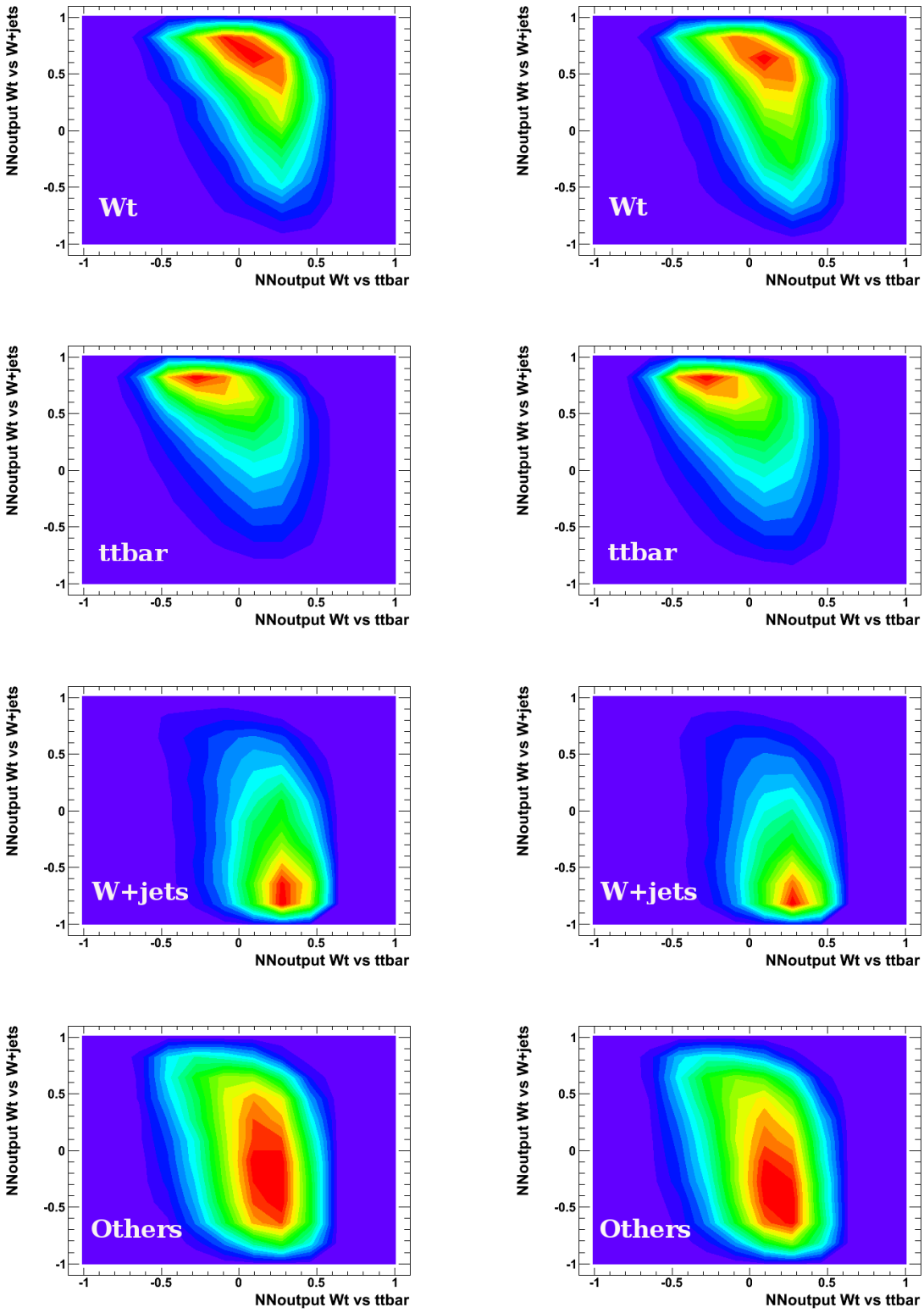

Figure 6.8: Two-dimensional distributions of the neural network outputs of the signal against W+jets production and the signal against $t\bar{t}$ production for the signal and the backgrounds: $t\bar{t}$ production, W+jets production and "others". Plots in the left column are for the electron channel and plots in the right column for the muon channel. The most concentrated region is shown in red.

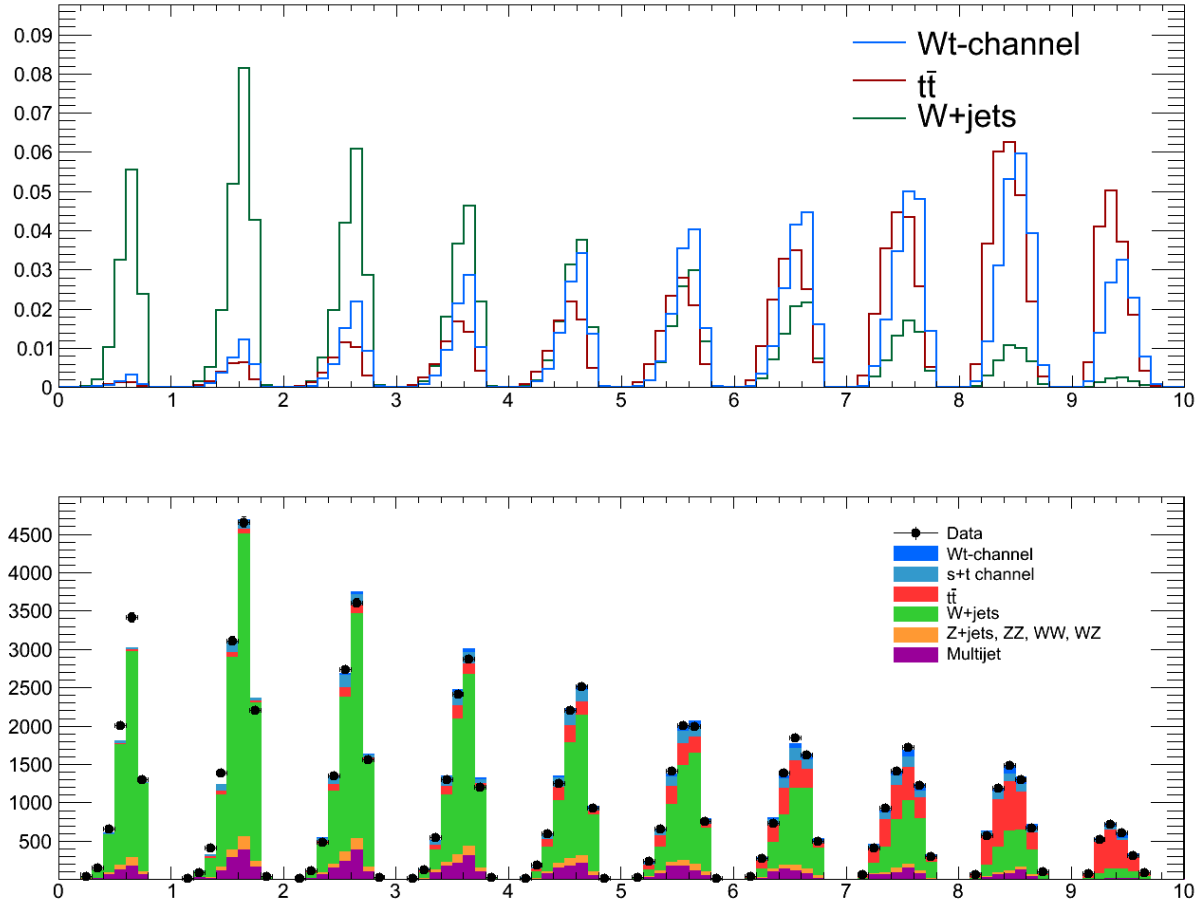

Figure 6.9: One-dimensional plots from unrolling the two-dimensional distributions. The unrolling direction of the two-dimensional distributions is from the bottom to the top corresponding to the 1st bin to the 10th bin. Electron and muon channels are summed. The top plot is the normalized overlay of the signal and two main background distributions. The bottom plot shows a Monte Carlo to data comparison.

Chapter 7

Signal Extraction

In the previous chapter the two-dimensional distribution of the neural network outputs of the signal against $t\bar{t}$ production and the signal against W+jets production was found as the most promising approach to isolate the Wt-channel signal from its backgrounds. The next important task is the extraction of the cross section and the computation of the significance of the observed Wt-channel signal. This is done with the fitting package Bill. In the first part of this chapter, a brief overview of the mathematical formalisms used by Bill and a measurement of the cross section are given. How the significance is calculated is described in detail in the second part of this chapter. Full details of the tool Bill can be found [13, 52].

7.1 Bill Tool

The fitting package Bill (BInned Log Likelihood) was developed by the ATLAS group at the University of Wuppertal and has already been used in other single top-quark analyses [13]. In this analysis, the entire range of the one-dimensional distribution from unrolling the two-dimensional distributions of the two neural network outputs is used in the binned maximum likelihood fit, in order to extract the cross section of the observed Wt-channel signal and to compute the significance of the measurement. Fitting all bins of the distribution is beneficial since all signal events are involved in the procedure. The probability density distributions of the discriminants are called templates for the fit. There are five templates from the Monte Carlo samples: the Wt-channel signal, the other two single top-quark channels, $t\bar{t}$ production, W+jets production, diboson and Z+jets, together with one template of multijet and one template of the combined data of the electron and muon channels are provided to Bill to perform the fit.

The likelihood function used by Bill is given as the product of the Poisson likelihoods of each template histogram bin multiplied with a Gaussian prior to include the background rates. The likelihood function is defined as:

$$L(\beta^s;\beta^b_j) = \prod_{k=1}^{M} \frac{e^{-\mu_k} \times \mu_k}{n_k!} \times \prod_{j=1}^{B} G(\beta^b_j; 1.0, \Delta_j) \quad \text{with} \tag{7.1}$$

$$\mu_k = \mu^s_k + \sum_{j=1}^{B} \mu^b_{jk}, \quad \mu^s_k = \beta^s \tilde{v}^s \alpha^s_k \quad \text{and} \quad \mu^b_{jk} = \beta^b_j \tilde{v}_j \alpha_{jk}. \tag{7.2}$$

The scale factors of signal and background samples are represented by the fit parameters β^s and β^b_j. The first product runs over the index k of bins up to the maximum bin M and includes the Poisson function with the expected number of events μ and the number of observed events n. $\tilde{v}^s$ denotes the predicted

number of signal events. The indices s and b indicate signal and background samples. The term α_{jk} is the fraction of the background templates in each bin k, for each source j, and can be normalised such that $\Sigma_k\, \alpha_{jk} = 1$.

The second product is the Gaussian prior, G, multiplied over all background samples. The Gaussian is defined as:

$$G(\beta_j^b; 1.0, \Delta_j) = \frac{1}{\sqrt{2\pi\Delta_j^2}} \cdot \exp\left(-\frac{1}{2}\left(\frac{\beta_j^b - 1}{\Delta_j}\right)^2\right) \tag{7.3}$$

The purpose of the Gaussian prior is to constrain the background rates. For each background, the function has a mean of one and a width Δ_j. The widths Δ_j represent the relative rate uncertainty of the background processes. Table 7.1 shows the relative rate uncertainties according to their theoretical cross section uncertainties for the 2-jet topology. The relative uncertainty of the multijet background is set close to zero to stabilise the fit since the number of multijet events was already determined by a fraction fit. Its uncertainty will be included in systematic uncertainties only. The expected rates are from the event yields of Monte Carlo samples shown in Table 5.3.

Process	Other single tops	$t\bar{t}$ production	W+jets	Diboson & Z+jets	Multijet
Δ_j	0.04	0.10	0.33	0.15	0.00

Table 7.1: Relative rate uncertainties from their theoretical cross section uncertainties for the 2-jet topology. These relative uncertainties are applied as constrains to the fit of the scale factors for all backgrounds.

Since the fit parameters determined directly by maximising the likelihood function are numerically unstable, the negative logarithm of the likelihood function is instead exploited using Minuit [52]. From the minimum of the negative log likelihood, the maximum of the likelihood is calculated. The likelihood function can be parametrised with one function of β_1 by minimising the negative log likelihood with respect to all other free parameters. This is known as the reduced likelihood function $L_{\text{red}}(\beta_1)$.

A measurement of the cross section is one of the results of the likelihood fit. The scale factor for the fit, $\hat{\beta}^s$, given is related to the theoretical cross section by:

$$\hat{\sigma} = \hat{\beta}^s \cdot \sigma_{\text{theo}}. \tag{7.4}$$

The uncertainties on the cross section measurement are estimated using pseudo-experiments that are created by a Monte Carlo method. In general, an ensemble of pseudo-experiments refers to as a set of independent replications of any interesting physics experiment. Systematic uncertainties can also be included in this procedure. However systematic uncertainties are not considered in this thesis; therefore this will not be detailed here. A full description can be found in [20]. The result of the template fit to the signal and backgrounds using 1 million pseudo-experiments is shown in Table 7.2.

As discussed above, the cross section of the measurement is related directly to the fit parameter β. For this analysis, the cross section of the Wt-channel signal from combining the electron and muon channels is determined to be

$$\sigma/\sigma_{SM} = 0.50 \pm 0.22,$$
$$\sigma = 7.9 \pm 3.5 \text{ pb}.$$

In this case, σ_{SM} is the signal cross section predicted at a center-of-mass energy of 7 TeV.

process	β
Wt-channel	0.50 ± 0.22
Other single tops	0.99 ± 0.04
$t\bar{t}$ production	0.98 ± 0.03
W+jets	1.04 ± 0.01
Diboson & Z+jets	0.94 ± 0.14
Multijet	1.00

Table 7.2: Result of the template fit. All results are extracted using the combined data from the electron and muon channels. The scale factors, β, represent the measured cross sections. Only statistical uncertainties are presented. The uncertainty of the multijet background is set to zero in the fit.

7.2 Significance

One application of Bill is the calculation of the significance of the signal. In order to do so, 1 million pseudo-experiments are used. The hypotheses tested are with the Wt-channel signal at the Standard Model prediction, H_1, and without the Wt-channel signal, H_0. The Q-value of the signal+background hypothesis and the Q-value of the background-only hypothesis are estimated. The Q-value is defined as:

$$Q = -2[\ln(L_{\text{red}}(\beta_1 = 1)) - \ln(L_{\text{red}}(\beta_1 = 0))]. \tag{7.5}$$

$L_{\text{red}}(\beta_1 = 1)$ is the value of the reduced likelihood function for a cross section of the signal as predicted by the Standard Model, while $L_{\text{red}}(\beta_1 = 0)$ is the value of the reduced likelihood function for a signal cross section of zero. The Q-value distributions are presented in figure 7.1. The expected Q-value is shown as the solid line. The observed Q-value shown as the dashed line is obtained by fitting to collision data. In order to determine how well the observed value, Q_0, agrees with the null hypothesis, the so-called p-value is derived as:

$$p(Q_0) = \frac{1}{I_q} \cdot \int_{-\infty}^{Q_0} q_0(Q')\, dQ' \quad \text{with} \tag{7.6}$$

$$I_q = \int_{-\infty}^{+\infty} q_0(Q')\, dQ', \tag{7.7}$$

where q_0 is the Q-value distribution for the null hypothesis. Also, the expected p-value can be computed from the median of the Q-value distribution for the hypothesis H_1. In practice, the expected/observed p-value is determined from the overlapping area of the distribution from the hypothesis with and without signal left of the expected/observed Q-value. The expected/observed significance of the measurement is then calculated from the expected/observed p-value. All these values are reported in Table 7.3.

The observed significance of this analysis excluding any systematic uncertainty is only 2.4σ while the expected significance is 4.4σ. It is unlikely that the value of observed significance is smaller than the expected value. In order to cross-check the results another tool for signal extraction might be used.

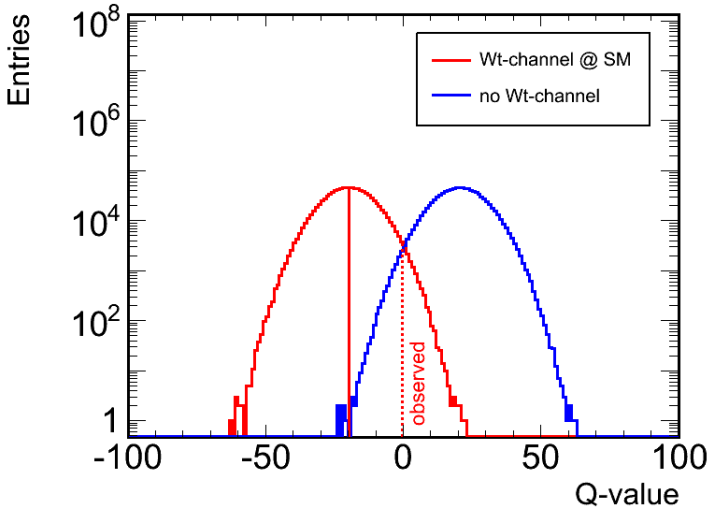

Figure 7.1: Distribution of Q-values for the hypothesis with and without the signal derived from pseudo-experiments. The expected Q-value is denoted by the solid line. The observed Q-value is denoted by the dashed line. The value of the expected/observed p-value is estimated from a ratio of area between these two distributions left of the Q-value.

	Expected	Observed
Q-value	-19.84	-0.18
p-value	5.0×10^{-6}	9.2×10^{-3}
Significance [σ]	4.4	2.4

Table 7.3: Expected and observed Q-values, p-values and significances of the measurement for the 2-jet topology. All these values are evaluated using 1 million pseudo experiments. Systematic uncertainties are not taken into account. The expected/observed significance is determined from the expected/observed p-value.

Chapter 8

Summary

A search for the Wt-channel single top-quark production in the lepton+jets channel with exactly one electron or muon, two quark jets, one of which is a b-tagged jet and missing transverse energy has been presented in this thesis. In this channel, one of the two light-quark jets does not pass the event selection; therefore, it can imply such that this light-quark is missing from the final state of Wt-channel system. The signal fraction of the 2-jets topology is about 42%; however it has very large backgrounds. Therefore, the study of this topology is challenging. Proton-proton collision data collected by the ATLAS detector in 2011 at $\sqrt{s} = 7$ TeV with an integrated luminosity of 4.7 fb^{-1} are used. Approaches to discriminate the Wt signal from backgrounds, in particular top-quark pair production and W+jets production, were examined.

In order to obtain a good estimation of the multijet background, the matrix method and jet-electron model, which are recommended to use for estimating multijet in the muon channel for single top-quark analyses, have been compared to each other. The implementation of the multijet background estimation in the muon channel using the matrix method has been presented in detail. The matrix method was found to provide the better muon transverse momentum distribution while the jet-electron provides the better distribution of the missing transverse energy. Apart from that, the results obtained with each of the two methods were similar. The jet-electron model was applied to both the electron and the muon channel in this thesis.

Three different b-tagging working points have been studied using an artificial neural network trained to separate signal from a mixture of all backgrounds. As a result the working point of 70% efficiency was found to perform best.

For the analysis, three different approaches of signal extraction were compared with respect to their final discrimination power of signal-to-background. Signal trained against a mixture of all backgrounds showed that the overall discrimination power was achieved. While the discrimination against W+jets production was found to be good, top-quark pair production distribution is very similar in shape with the signal distribution. Two new approaches that aim to discriminate the signal from top-quark pair production and still provide a good discrimination against W+jets production were introduced.

The first alternative strategy combined the neural network outputs of three independent networks of the signal trained against each background and the absolute value of the neural network output of the W+jets and top-quark pair production network as inputs for another neural network to obtain one final output discriminator. This approach did not provide a better separation against top-quark pair production and the overall separation power of the combined neural network became worse compared to the standard approach. The problematic issue is that the result of the network trained against top-quark pair production did not play an important role in the combined neural network.

As a second alternative approach, the use of the two-dimensional distribution of the outputs of two neural networks, each trained against one of the two main backgrounds was investigated. Shape dif-

ferences between the signal and top-quark pair production are clearly visible. A discrimination from W+jets production was achieved. This approach gave the best results.

The fit of the combined data performed as binned maximum likelihood fit with the Bill package, provided a cross section of Wt signal:

$$\sigma = 7.9 \pm 3.5 \text{ pb}.$$

An significance of 4.4σ was expected, a significance of 2.4σ was observed excluding systematics.

Appendix A

Kinematic Variables

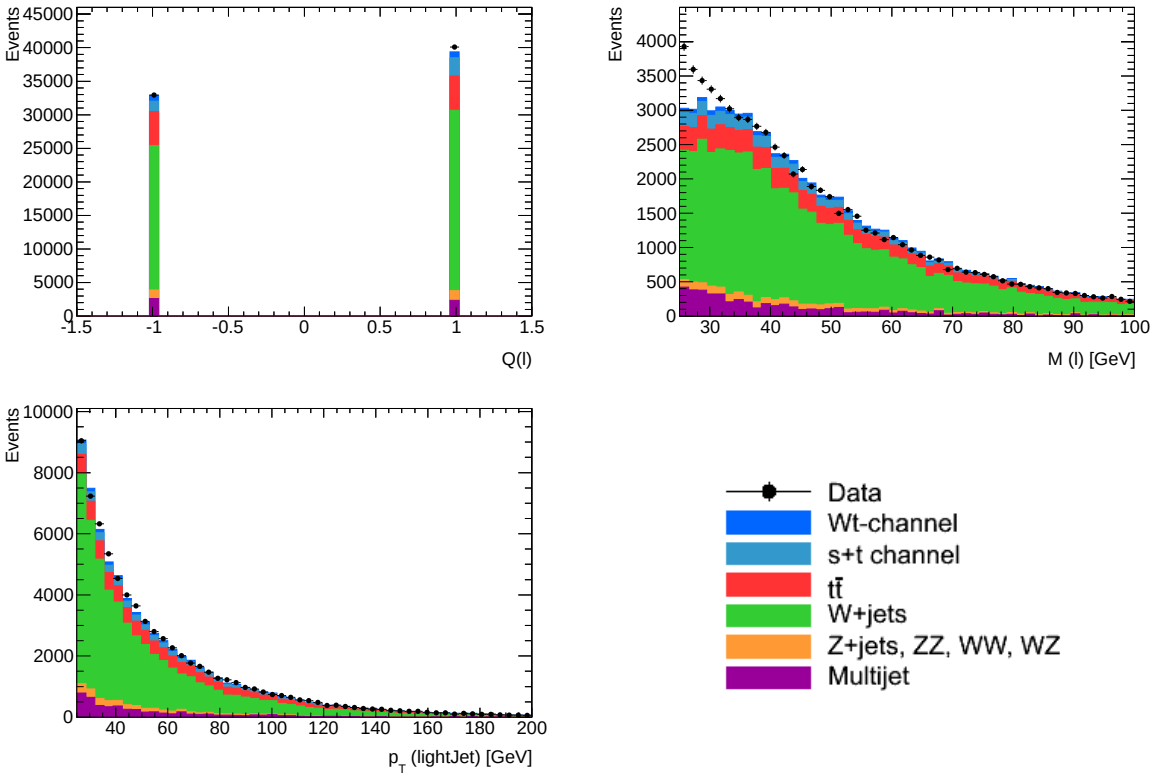

Figure A.1: Control plots of the kinematic variables which are picked in the preprocessing step and then used in the network training. Electron and muon channels are summed up. In the first row: $Q(l)$ and $M(l)$. In the second row: p_T(lightJet).

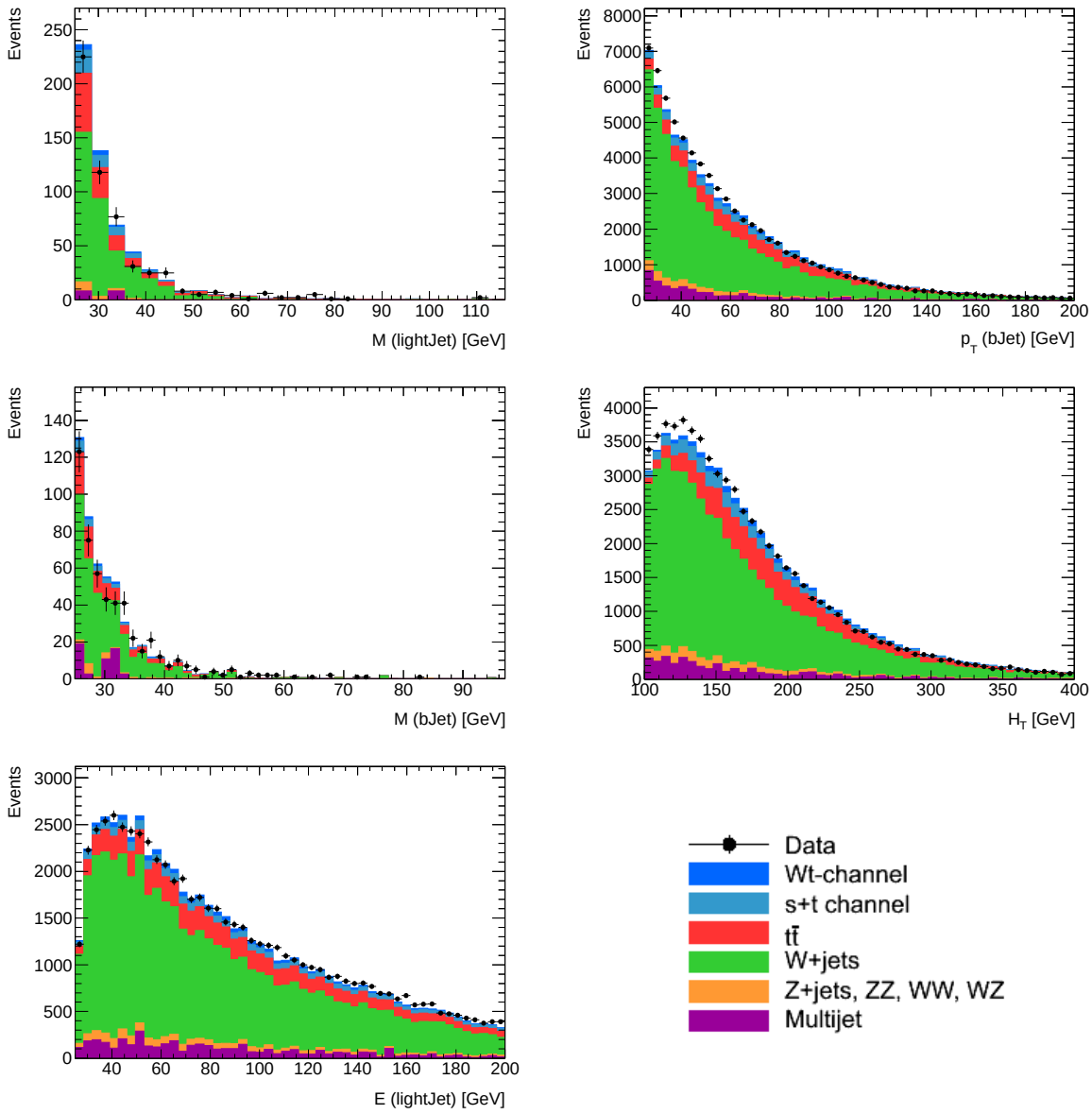

Figure A.2: Control plots of the kinematic variables which are picked in the preprocessing step and then used in the network training. Electron and muon channels are summed up. In the first row: M(lightJet) and p_{T}(bJet). In the second row: M(bJet) and H_{T}. In the third row: E(lightJet).

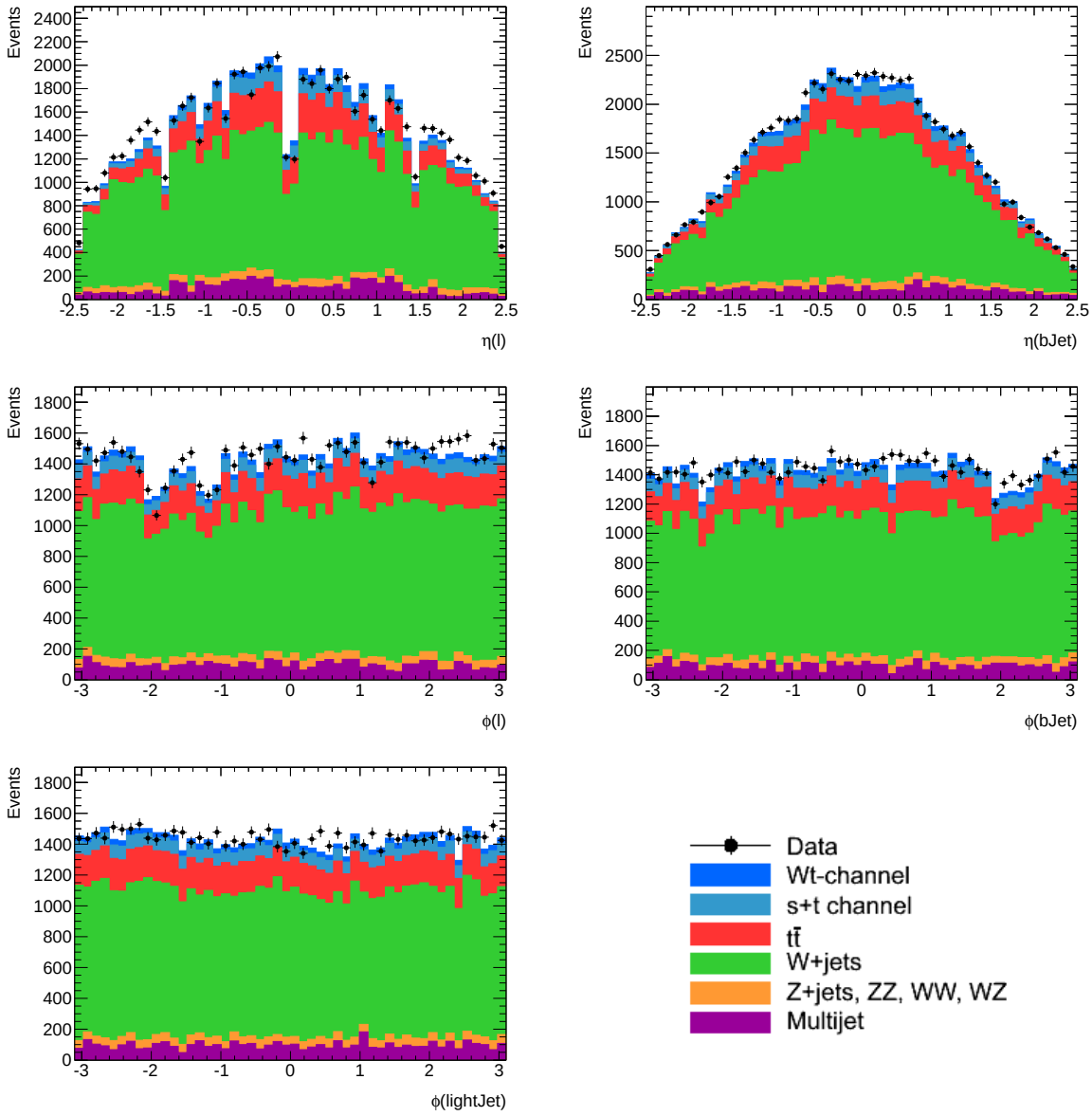

Figure A.3: Control plots of the kinematic variables which are picked in the preprocessing step and then used in the network training. Electron and muon channels are summed up. In the first row: $\eta(l)$ and $\eta(b\text{Jet})$. In the second row: $\phi(l)$ and $\phi(b\text{Jet})$. In the third row: $\phi(\text{lightJet})$.

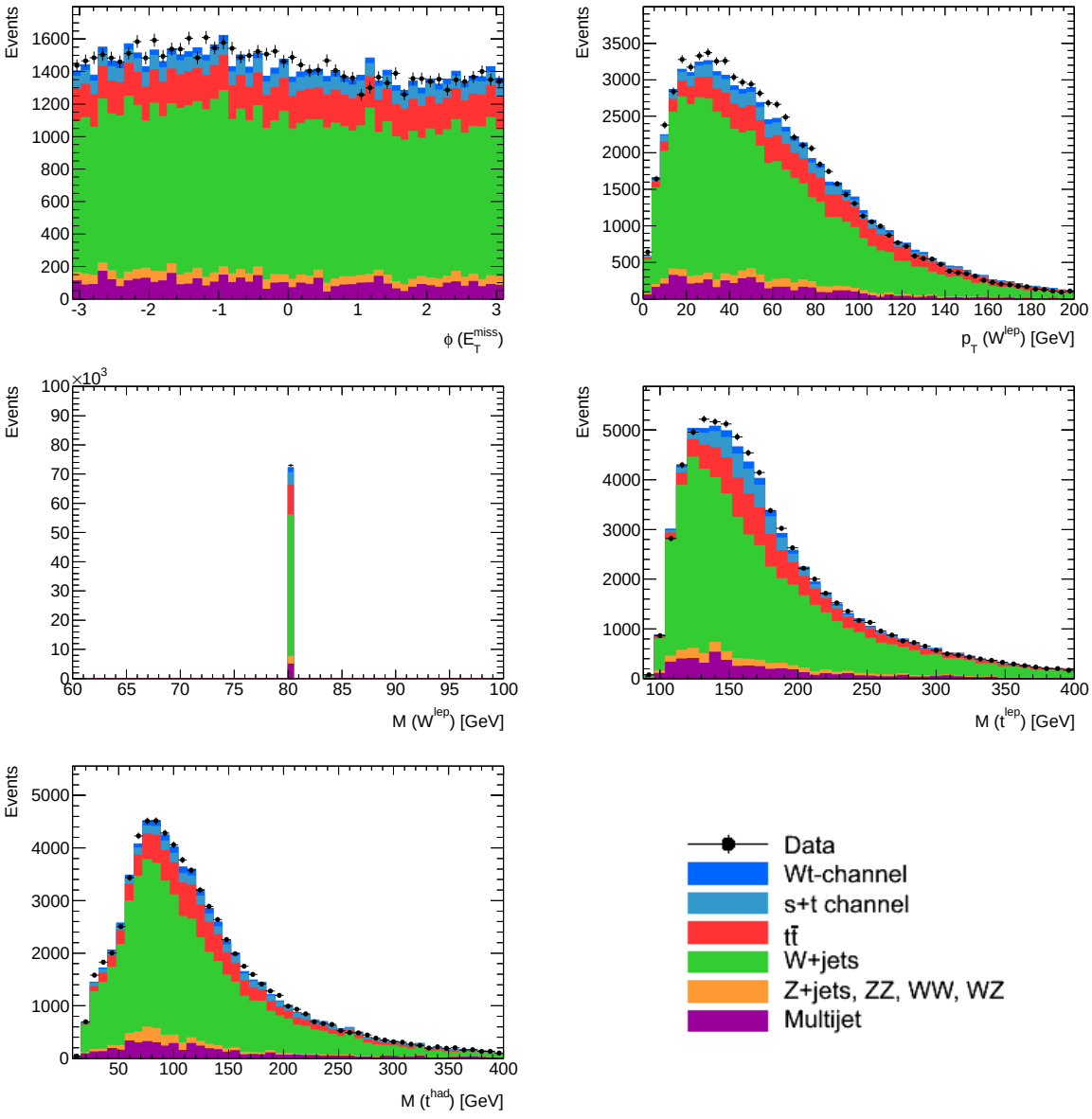

Figure A.4: Control plots of the kinematic variables which are picked in the preprocessing step and then used in the network training. Electron and muon channels are summed up. In the first row: $\phi(E_{\mathrm{T}}^{miss})$ and $p_{\mathrm{T}}(W^{\mathrm{lep}})$. In the second row: $M(W^{\mathrm{lep}})$ and $M(t^{\mathrm{lep}})$. In the third row: $M(t^{\mathrm{had}})$.

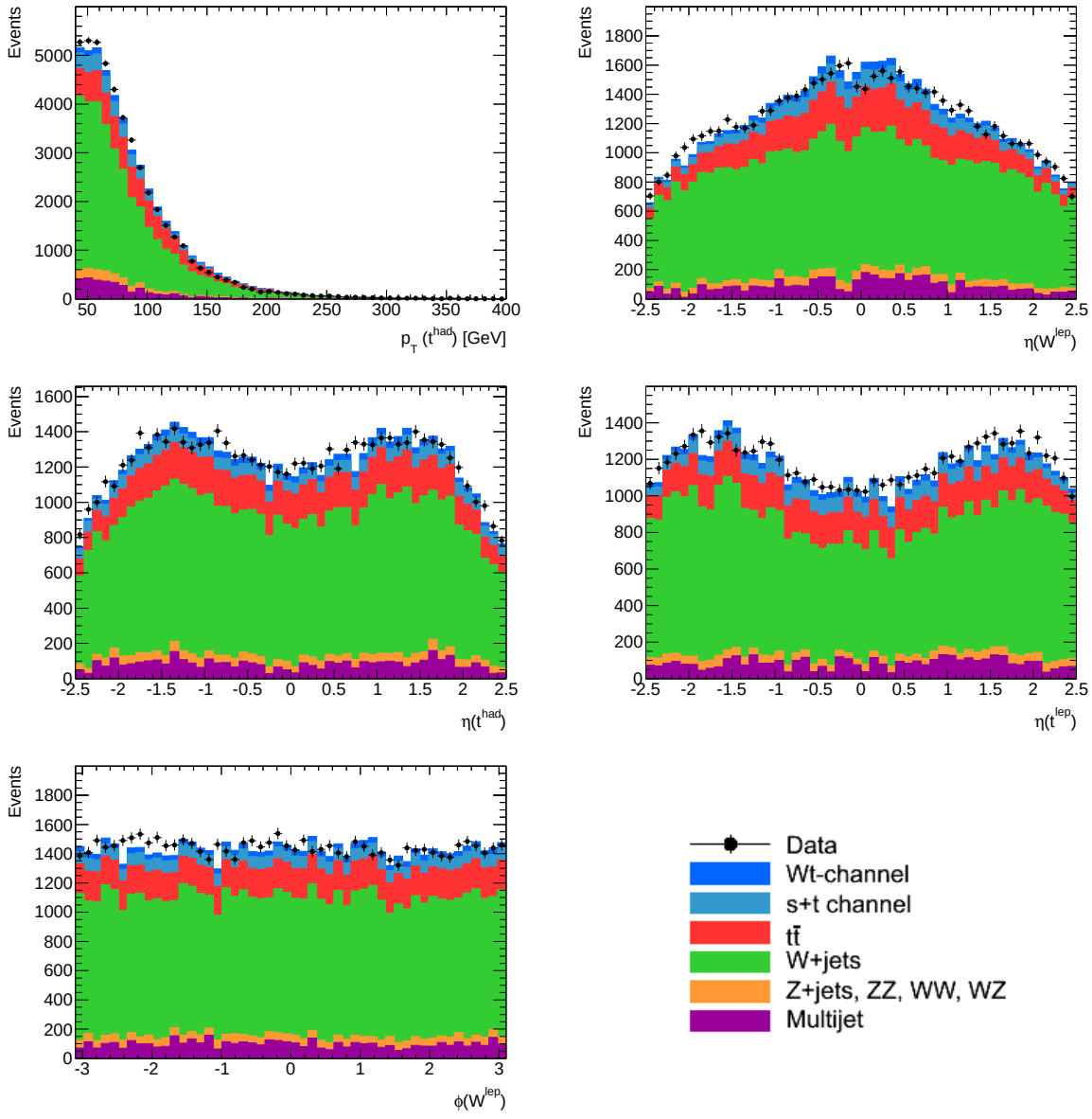

Figure A.5: Control plots of the kinematic variables which are picked in the preprocessing step and then used in the network training. Electron and muon channels are summed up. In the first row: $p_{\mathrm{T}}(t^{\mathrm{had}})$ and $\eta(W^{\mathrm{lep}})$. In the second row: $\eta(t^{\mathrm{had}})$ and $\eta(t^{\mathrm{lep}})$. In the third row: $\phi(W^{\mathrm{lep}})$.

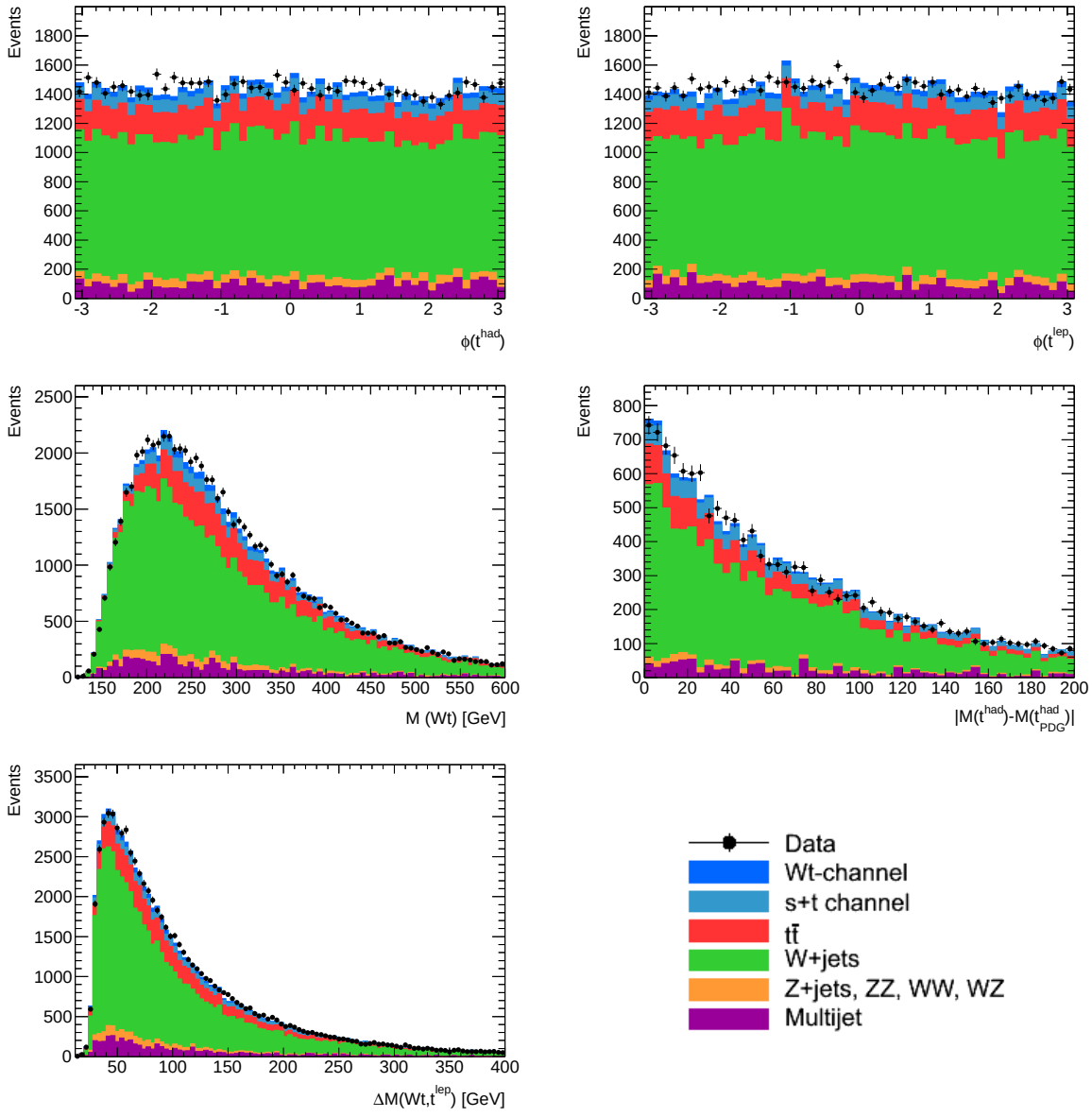

Figure A.6: Control plots of the kinematic variables which are picked in the preprocessing step and then used in the network training. Electron and muon channels are summed up. In the first row: $\phi(t^{\text{had}})$ and $\phi(t^{\text{lep}})$. In the second row: $M(Wt)$ and $|M(t^{\text{had}}) - M(t^{\text{had}}_{\text{PDG}})|$. In the third row: $\Delta M(Wt, t^{\text{lep}})$.

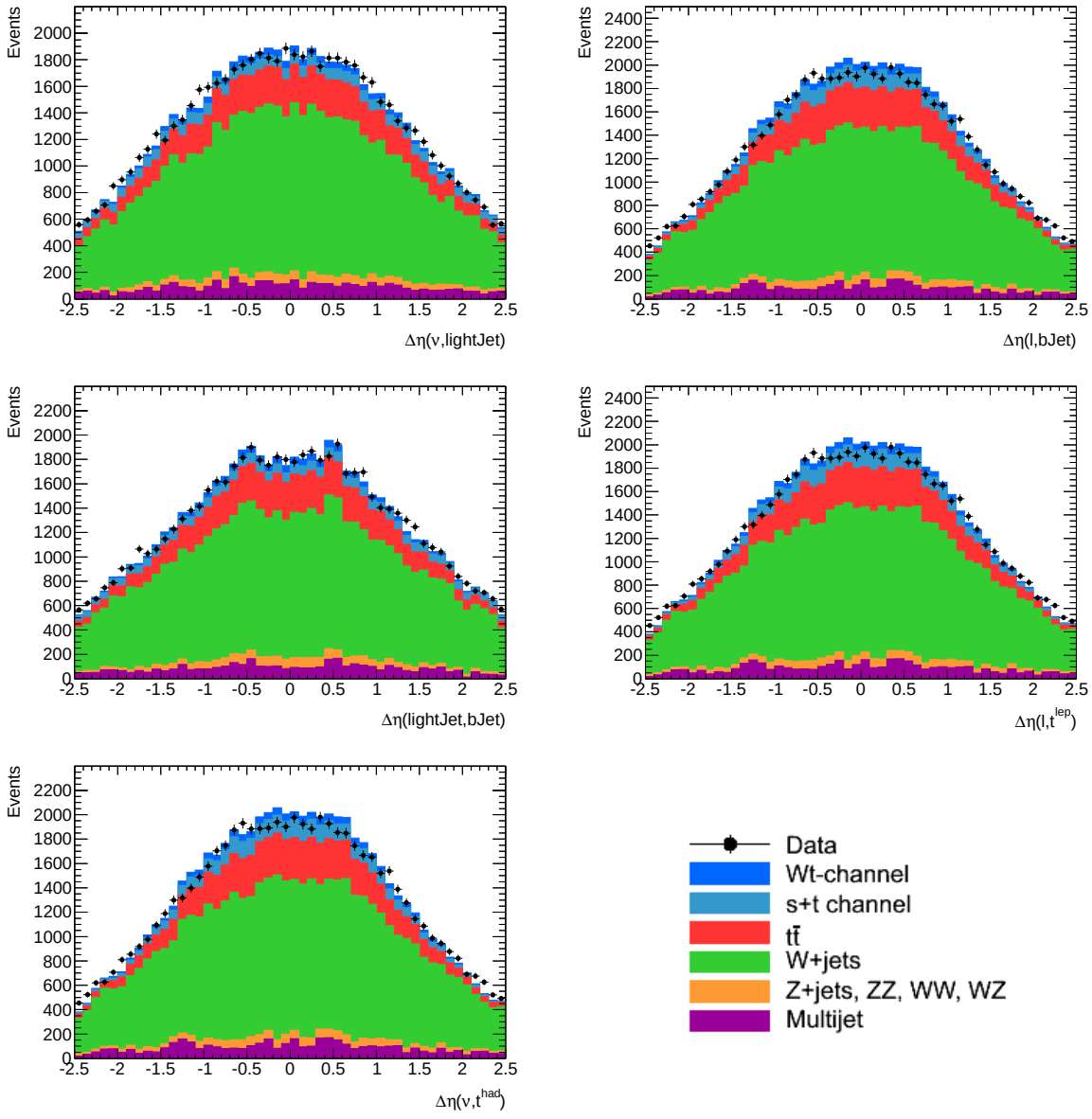

Figure A.7: Control plots of the kinematic variables which are picked in the preprocessing step and then used in the network training. Electron and muon channels are summed up. In the first row: $\Delta\eta(\nu,\text{lightJet})$ and $\Delta\eta(l,b\text{Jet})$. In the second row: $\Delta\eta(\text{lightJet},b\text{Jet})$ and $\Delta\eta(l,t^{\text{lep}})$. In the third row: $\Delta\eta(\nu,t^{\text{had}})$.

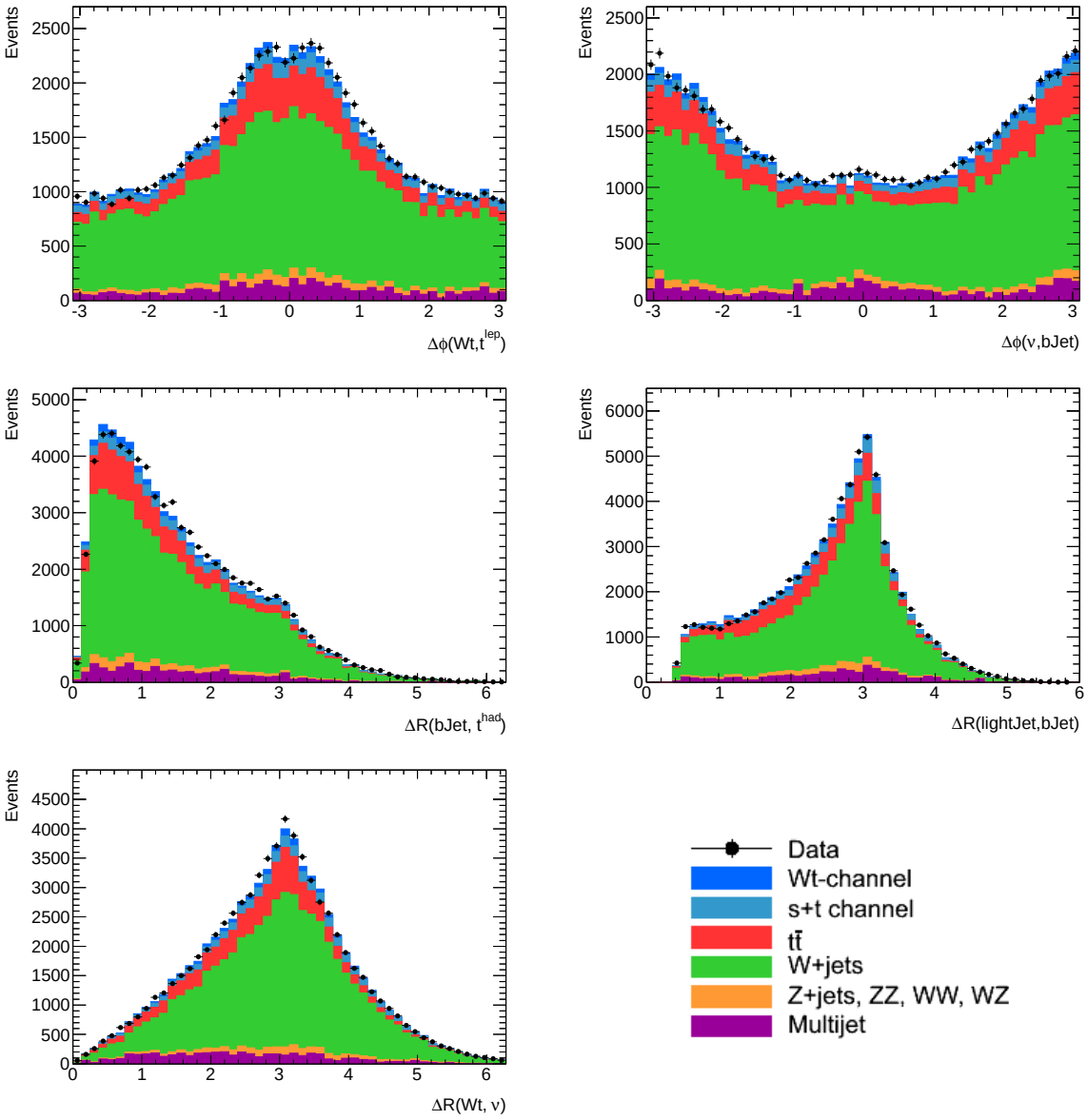

Figure A.8: Control plots of the kinematic variables which are picked in the preprocessing step and then used in the network training. Electron and muon channels are summed up. In the first row: $\Delta\phi(Wt, t^{\text{lep}})$ and $\Delta\phi(\nu, b\text{Jet})$. In the second row: $\Delta R(b\text{Jet},t^{\text{had}})$ and $\Delta R(\text{lightJet},b\text{Jet})$. In the third row: $\Delta R(Wt, \nu)$.

Appendix B

Truth Information

Using the truth information of the Monte Carlo *Wt*-channel sample, one is able to study the decay modes of the *Wt*-channel sample. A description of the truth study can be found in detail in [48]. Table B.1 presents all decay modes in the signal sample both after pretag and tag selections of the events that have exactly 2 jets. For the signal sample in the electron and muon channels, most of the events are from electron+jets and muon+jets decays, respectively.

Considering the combined channel all-hadronic decays can be neglected since they add up less than 1% of the overall distribution. Dilepton events contribute approximately 26% in the pretag selection and 31% in the tag selection. Another contribution of the *Wt*-channel system is from τ+jets decays providing about 9% in the pretag selection and 6% in the tag selection. Furthermore, a minor fraction of the *Wt*-channel sample shows that they could not fulfill any classification, that is referred to as "not-classified".

Decay channel type	Pretag selection		Tag selection	
	Muon channel	Electron channel	Muon channel	Electron channel
Not-classified	21	18	9	8
all-hadronic	3	1	0	0
e+jets	1	9750	0	6770
μ+jets	22230	2	9480	1
τ+jets	1470	1300	584	286
ee	0	2420	0	1170
$e\mu$	2720	1040	1460	484
$e\tau$	164	3690	87	1700
$\mu\mu$	1140	0	593	0
$\mu\tau$	4040	52	2100	19
$\tau\tau$	235	235	115	88

Table B.1: Configuration of the true decay channels in the *Wt*-channel sample for only events with the number of jets equal to two both after pretag and tag selections. An event that could not be classified by any of the classifications is named as "not-classified".

Only the events from the electron+jets and muon+jets decays are referred to as the lepton+jets decay sample of the *Wt*-channel, and they are used to investigate the efficiency of the neural network training to each sub-decay mode of the lepton+jets decay channel. For this lepton+jets decay mode, the *Wt*-channel events are separated depending upon the decay modes of the top quark (leptonic, t^{lep} or hadronic, t^{had})

in the *Wt* final state. The other decay types: are grouped into the category named as "other". "Other" signifies that an event could not be classified by the criteria of the lepton+jets decay mode. Furthermore, the samples are sub-divided according to whether a 1:1 relation between the light quarks and the light-quark jets and the *b* quark and the *b*-quark jet can be established. Ten individual networks of the *Wt*-channel signal according to the true decay modes as described above are trained for this purpose.

For this comparison, all discriminating variables are fed into each network with a significance cut of 8σ and the Monte Carlo samples after the tag selection are used. Each of these networks is trained against all backgrounds. The cut of 8σ is used since the reasonable number of picked variables is obtained for all network trainings. Also, the training of each network is done with the optimum number of hidden nodes in the hidden layer. Table B.2 shows the training results including their event fractions. As discussed above, the fraction of all lepton+jet decay events is highest and most of these events (51.6% of the overall) fulfill the "matching" criteria. In addition, "t^{had} matching" contributes a larger amount than "t^{lep} matching" does. Among all networks, the "t^{had} matching" network provides the highest value of the total correlation to target that leads to the largest Gini index. Therefore this network provides the best separation power among the others from all backgrounds. One can see that network with a high total significance is not necessary to provide a good separation power from the backgrounds. Furthermore the network of the signal with "t^{lep} matching" gives the second best separation power. Interestingly, the next best separation power is provided by the network called "other" instead of "lepton+jets mode matching" that is the combination between "t^{had} matching" and "t^{lep} matching". One possible reason for that would be the "t^{had} matching" events strongly differ from the "t^{lep} matching" events in phase space, which leads to a non-homogeneous sample in the combination. Such an inhomogeneous sample may be more difficult to separate for the neural network. Furthermore, the events in the "other" sample may be closer together in phase space.

The result looks really promising since separating samples of the signal increases the separation power significantly; although all sample-sets of the non-matching classification give worse separation powers compared to the network using the full *Wt*-channel sample. Actually, that is what one would expect to have here "the networks of separated samples with the right matching are trained more effectively against the backgrounds". Further investigation is needed on this topic (this was just a brief proof-of-principle-like study anyway), in particular one still has to find a way how to exploit these promising results for the analysis.

Sample	Event fraction after tag sel. [%]	Total correlation to target [%]	Total significance [σ]	Gini index [%]
Full *Wt*-channel sample	100	44.6	103.7	25.8
After truth selection				
Lepton+jets mode	65.1	46.9	92.6	27.0
Lepton+jets mode matching	51.6	50.9	91.7	29.7
Lepton+jets mode not matching	13.5	45.2	44.2	23.7
t^{had}	34.5	55.6	84.0	32.2
t^{had} matching	27.4	61.7	84.2	36.0
t^{had} not matching	7.1	50.4	36.1	24.5
t^{lep}	30.6	50.9	72.8	29.2
t^{lep} matching	24.2	56.7	73.2	33.0
t^{lep} not matching	6.4	48.1	32.9	20.6
Other	34.9	56.0	84.9	32.7

Table B.2: Summary of the results of training the eleven networks using the truth information of the *Wt*-channel signal for only events with the number of jets equal to two. An event fraction of each true decay mode after tag selection is also given. "Other" signifies that an event could not be classified by the criteria of the lepton+jets decay mode. The lepton+jets decay mode is further separated into two cases based on two possibilities of the top quark decay either hadronic mode or leptonic mode. "Matching" denotes that one light-quark gets matched to one light-jet and one *b*-quark to one *b*-jet. All network trainings are cut with a significance of 8σ and using the combined data.

Bibliography

[1] J. Beringer and et al. (Particle Data Group). ‘Review of particle physics 2012’. In: *Phys. Rev. D*86, 010001 (2012).

[2] D. Griffiths. Introduction to elementary paricles. *WILEY-VCH*, 2008.

[3] F. P. Schilling. ‘Top Quark Physics at the LHC: a Review of the First Two Years’. In: *International Journal of Modern Physics A* 27, 1230016 (July 2012), p. 30016. DOI: 10.1142/S0217751X12300165. eprint: 1206.4484.

[4] V. M. Abazov et al. ‘Improved determination of the width of the top quark’. In: *Phys. Rev. D* 85.9, 091104 (May 2012), p. 091104. DOI: 10.1103/PhysRevD.85.091104. eprint: 1201.4156.

[5] CMS Collaboration. ‘Measurement of the mass difference between top and antitop quarks’. In: *ArXiv e-prints* (Apr. 2012). eprint: 1204.2807.

[6] J. Pumplin et al. ‘New generation of parton distributions with uncertainties from global QCD analysis’. In: *Journal of High Energy Physics* 7, 012 (July 2002), p. 12. DOI: 10.1088/1126-6708/2002/07/012. eprint: ArXiv:hep-ph/0201195.

[7] S. Moch and P. Uwer. ‘Theoretical status and prospects for top-quark pair production at hadron colliders’. In: *Phys. Rev. D* 78.3, 034003 (Aug. 2008), p. 034003. DOI: 10.1103/PhysRevD.78.034003. eprint: 0804.1476.

[8] H. Fritzsch and J. Sola. ‘Matter Non-conservation in the Universe and Dynamical Dark Energy’. In: *ArXiv e-prints* (Feb. 2012). eprint: 1202.5097.

[9] A. Szczurek and V. Uleshchenko. ‘The range of validity of the QCD-improved parton model’. In: *Physics Letters B* 475 (Feb. 2000), pp. 120–126. DOI: 10.1016/S0370-2693(00)00066-6. eprint: ArXiv:hep-ph/9911467.

[10] J. R. Incandela et al. ‘Status and prospects of top-quark physics’. In: *Progress in Particle and Nuclear Physics* 63 (Oct. 2009), pp. 239–292. DOI: 10.1016/j.ppnp.2009.08.001. eprint: 0904.2499.

[11] N. Kidonakis. ‘Next-to-next-to-leading-order collinear and soft gluon corrections for *t*-channel single top quark production’. In: *Phys. Rev. D* 83.9, 091503 (May 2011), p. 091503. DOI: 10.1103/PhysRevD.83.091503. eprint: 1103.2792.

[12] Tevatron Electroweak Working Group, for the CDF Collaboration and the D0 Collaboration. ‘Combination of CDF and D0 measurements of the single top production cross section’. In: *ArXiv e-prints* (Aug. 2009). eprint: 0908.2171.

[13] ATLAS Collaboration. ‘Measurement of the *t*-channel single top-quark production cross section in *pp* collisions at center-of-mass energy of 7 TeV with the ATLAS detector’. In: *ATL-COM-PHYS-2012-1338* (May 2012).

[14] A. P. Heinson, f. t. D0 and CDF Collaborations. ‘Observation of Single Top Quark Production at the Tevatron’. In: *ArXiv e-prints* (Sept. 2009). eprint: 0909.4518.

[15] CERN accelerator complex (online). Available from: URL:http://public.web.cern.ch/. public/en/research/AccelComplex-en.html. (Accessed: 15/08/2012).

[16] ATLAS photos (online). Available from: URL:http://atlas.ch/photos/index.html. (Accessed: 15/08/2012).

[17] ATLAS Collaboration. 'Expected performance of the ATLAS experiment - detector, trigger and physics'. In: *ArXiv e-prints* (Dec. 2009). eprint: 0901.0512.

[18] S. Laplace and ATLAS liquid argon calorimeter group. 'Commissioning of the ATLAS liquid argon calorimeter'. In: *Nuclear Instruments and Methods in Physics Research A* 617 (May 2010), pp. 30–34. DOI: 10.1016/j.nima.2009.09.108. eprint: 1005.2935.

[19] C. Gabaldon. Measurement of the luminosity by the ATLAS experiment. Tech. rep. *ATL-LUM-PROC*-2011-004. Geneva: CERN, 2011.

[20] I. C. Brock et al. Measurement of the *Wt* production in the lepton+jets channel using neural networks. Tech. rep. *ATL-COM-PHYS*-2011-1706. Geneva: CERN, 2011.

[21] G. Brandt et al. 'Recommendations for measurements of vector boson production in association with jets'.

[22] P. Onyisi on behalf of the ATLAS Collaboration. 'Operation of the ATLAS detector with first collisions at 7 TeV at the LHC'. In: *ArXiv e-prints* (Dec. 2010). eprint: 1012.1036.

[23] B. P. Kersevan and E. Richter-Was. 'The Monte Carlo Event Generator AcerMC 2.0 with Interfaces to PYTHIA 6.2 and HERWIG 6.5'. In: *ArXiv High Energy Physics - Phenomenology e-prints* (May 2004). eprint: ArXiv:hep-ph/0405247.

[24] L. Bellagamba et al. Validation of the POWHEG MC Generator for the ATLAS Experiment. Tech. rep. *ATL-COM-PHYS-2010-699*. Geneva: CERN, 2010.

[25] S. Frixione and B. R. Webber. 'Matching NLO QCD computations and parton shower simulations'. In: *Journal of High Energy Physics* 6, 029 (June 2002), p. 29. DOI: 10.1088/1126-6708/2002/06/029. eprint: ArXiv:hep-ph/0204244.

[26] A. Buckley et al. Summary of ATLAS Pythia 8 tunes. Tech. rep. *ATL-COM-PHYS-2012-738*. Geneva: CERN, 2012.

[27] G. Corcella et al. 'HERWIG 6: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)'. In: *Journal of High Energy Physics* 1, 010 (Jan. 2001), p. 10. DOI: 10.1088/1126-6708/2001/01/010. eprint: ArXiv:hep-ph/0011363.

[28] M. L. Mangano et al. 'ALPGEN, a generator for hard multiparton processes in hadronic collisions'. In: *Journal of High Energy Physics* 7, 001 (July 2003), p. 1. DOI: 10.1088/1126-6708/2003/07/001. eprint: ArXiv:hep-ph/0206293.

[29] D. Kar. ATLAS MPI tunes with various PDFs. Tech. rep. *ATL-COM-PHYS-2012-073*. Geneva: CERN, 2012.

[30] N. A. Styles. New ATLAS Event Generator Tunes to 2010 pp Collision Data at 7 TeV Centre-of-Mass Energy. Tech. rep. *ATL-COM-PHYS-2011-1235*. Geneva: CERN, 2011.

[31] R. Field. 'The Underlying Event in Hadronic Collisions'. In: *Annual Review of Nuclear and Particle Science* 62.1 (2012), null. DOI: 10.1146/annurev-nucl-102711-095030.

[32] ATLAS Collaboration. 'Measurement of underlying event characteristics using charged particles in pp collisions at $\sqrt{s}$ = 900 GeV and 7 TeV with the ATLAS detector'. In: *Phys. Rev. D* 83 (11 2011), p. 112001. DOI: 10.1103/PhysRevD.83.112001.

[33] P. M. Nadolsky et al. 'Implications of CTEQ global analysis for collider observables'. In: *Phys. Rev. D* 78.1, 013004 (July 2008), p. 013004. DOI: 10.1103/PhysRevD.78.013004. eprint: 0802.0007.

[34] R. S. Thorne et al. 'Status of MRST/MSTW PDF sets'. In: *ArXiv e-prints* (July 2009). eprint: 0907.2387.

[35] S. Frixione and B. R. Webber. 'Matching NLO QCD computations and parton shower simulations'. In: *Journal of High Energy Physics* 6, 029 (June 2002), p. 29. DOI: 10.1088/1126-6708/2002/06/029. eprint: ArXiv:hep-ph/0204244.

[36] B. Alvarez et al. 'Measurement of Single Top-Quark Production in the Lepton+Jets Channel in pp Collisions at $\sqrt{s}$ = 7 TeV'. Tech. rep. *ATL-PHYS-INT-2011-049*. Geneva: CERN, 2011.

[37] ATLAS top group. TopCommon Objects (online). Available from: URL: https://twiki.cern.ch/.twiki/bin/viewauth/AtlasProtected/TopCommonObjects. (Accessed: 25/08/2012).

[38] F. Monticelli, J. Lorenz and A. Tricoli. Performance of the ATLAS Electron and Photon Trigger in pp Collisions at $\sqrt{s}$ = 7 TeV in 2011. Tech. rep. *ATLAS-COM-CONF-2012-071*. Geneva: CERN, 2012.

[39] M. Cacciari, G. P. Salam and G. Soyez. 'The anti-k_t jet clustering algorithm'. In: *Journal of High Energy Physics* 4, 063 (Apr. 2008), p. 63. DOI: 10.1088/1126-6708/2008/04/063. eprint: 0802.1189.

[40] ATLAS Collaboration. Data-Quality Requirements and Event Cleaning for Jets and Missing Transverse Energy Reconstruction with the ATLAS Detector in Proton-Proton Collisions at a Center-of-Mass Energy of $\sqrt{s}$ =7 TeV. Tech. rep. *ATLAS-CONF-2010-038*. Geneva: CERN, 2010.

[41] ATLAS Collaboration. 'Mis-identified lepton backgrounds to top quark pair production'. In: *ATL-COM-PHYS-2010-849* (Oct. 2010).

[42] ATLAS Collaboration. 'Measurement of the top quark-pair production cross section with ATLAS in pp collisions at center-of-mass energy of 7 TeV'. In: *European Physical Journal C* 71 (Mar. 2011), p. 1577. DOI: 10.1140/epjc/s10052-011-1577-6. eprint: 1012.1792.

[43] S. Gadatsch et al. Measurement of the *Wt* production in the lepton+jets channel using a likelihood method. Tech. rep. *ATL-COM-PHYS*-2011-1704. Geneva: CERN, 2011.

[44] ATLAS top group. TopFake muon matrix-method (online). Available from: URL: https://.twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/TopFakesMorind2012MmMujetsB. (Accessed: 11/04/2012).

[45] M. Feindt and U. Kerzel. 'The NeuroBayes neural network package'. In: *Nucl. Instrum. Meth. A559* (2006), pp. 190–194.

[46] T. Sjöstrand, S. Mrenna and P. Skands. 'PYTHIA 6.4 physics and manual'. In: *Journal of High Energy Physics* 5, 026 (May 2006), p. 26. DOI: 10.1088/1126-6708/2006/05/026. eprint: ArXiv:hep-ph/0603175.

[47] G. Hallenbeck et al. 'Model discrimination at the LHC: A case study'. In: *Phys. Rev. D* 79.7, 075024 (Apr. 2009), p. 075024. DOI: 10.1103/PhysRevD.79.075024. eprint: 0812.3135.

[48] T. Loddenkoetter. 'Implementation of a kinematic fit of single top-quark production in association with a W boson and its application in a neural-network-based analysis in ATLAS'. PhD thesis. Physikalisches Institut, Universität Bonn, 2012.

[49] I. Schall. ‘Development and application of a software framework for the single top-quark analysis using neural network with the CDF II experiment’. MA thesis. Physikalisches Institut, Universität Karlsuhe, 2008.

[50] M. Feindt. ‘A Neural Bayesian Estimator for Conditional Probability Densities’. In: *ArXiv Physics e-prints* (Feb. 2004). eprint: `ArXiv:physics/0402093`.

[51] S. Jain, H. B. Prosper and R. Schwienhorst. ‘Statistical methods implemented in the package: top statistics’. In: *D0 Note 5817* (2011).

[52] ATLAS top group. Bill tool (online). Available from: URL: https://twiki.cern.ch/twiki/bin/.viewauth/AtlasProtected/BillTool. (Accessed: 08/08/2012).

List of Figures

List of Tables

Acknowledgements

My sincere gratitude and respect go to my advisor, Prof. Dr. Ian C. Brock, for all his unlimited support and valuable recommendations during two years of my master's studies in Bonn, Germany. I would like to acknowledge him for giving me the great opportunity to work in his research group and for his help to make this thesis successful in time.

My big thanks are extended to Jan Stillings for supporting me since day one. He advised me on the *Wt* analysis framework in the beginning, and encouraged me time to time to learn all the needed skills for the thesis. Apart from work, we also discussed on economics, cultures, religions and especially Bonn weather. I wish to thank Thomas Velz (Loddenkötter) for his computer programming support and for the constructive suggestions on this thesis. Many thanks to Sebastian Mergelmeyer for the great discussions during the writing process and for helping me with programming problems as an expert. I should also thank you for all the love, understanding and support.

I would like to thank the following group members for the great conversations making my working days enjoyable: Ozan Arslan, Elena Nickel, Dustin Hebecker, Peter Kövesárki and Elizabeth Nuncio.

Last but not least, I would like to express my deep appreciation to my parents in Thailand for the unconditional love and constant support. With their power of love all my dreams always become true. I also cannot forget all my family members and friends who always share a part of my soul. My special thank must go to the solar eclipse in 1995 for inspiring me to study physics and keeping me on my toes.