Universität Bonn

Physikalisches Institut

Optimization of the single top-quark production analysis in the *Wt* channel at ATLAS using kinematic fitting and neural networks

Irina Cioară

The optimization of the search for the associated production of a top quark and a real W boson (Wt channel) is presented. The lepton + jets decay topology of this channel produces one b-quark jet, one charged lepton, two light-quark jets and one neutrino. The main sources of background for the Wt signal are top-quark pair production and events with a W boson and extra jets. An artifcial neural network is used to separate signal from background. The signal extraction procedure is optimized by including a kinematic fit of the Wt signal and studying only subsets of events that have a cleaner event topology. These consist of events that contain a hadronically decaying top quark or events in which the light-quark jets are matched to truth particles from the Monte Carlo information. The expected signal significance, including systematic uncertainties was evaluated. A decrease in systematic uncertainties by up to 20% and increase of the signal significance by 23% compared to the previous approach is observed when selecting only events with at least two matched light-quark jets.

Universität Bonn

Physikalisches Institut

Optimization of the single top-quark production analysis in the *Wt* channel at ATLAS using kinematic fitting and neural networks

Irina Cioară

Contents

Introduction

The Standard Model (SM) of particle physics has been the most popular physics theory of the last decades mostly because of the fact that it offers answers to fundamental questions regarding what are the elementary particles that make up the visible universe and how they interact with each other. As of July 2012, when the ATLAS and CMS experiments at the Large Hadron Collider reported the discovery of the Higgs boson, one can state that all particles predicted by the Standard Model have been experimentally found. This, however, does not make it a complete theory, in the sense that there are still open questions that need to be addressed, such as: the so-called hierarchy problem, the nature of neutrinos or the lack of a renormalisable theory for gravity.

One of the most interesting particles in the Standard Model, and also the heaviest, at 173 GeV, is the top quark. Its existence was first postulated in 1973 by Kobayashi and Maskawa but it was experimentally found only in 1995 at the Tevatron proton-antiproton collider. Because it is very short lived, the top quark decays before forming any bound state, thus allowing the determination of its properties with a very high precision. This is very important since, for example, top-quark mass measurements can provide constraints on the mass of the Higgs particle, which represents a self-consistency test for the Standard Model.

At hadron colliders, top quarks are produced either via the strong interaction, in top-antitop quark pairs, or via the weak interaction, in what is usually called single top-quark production. There are three possible channels for single top-quark production: the *s*-channel, the *t*-channel and the *Wt* channel. The last one is also referred to as associated production, because, along with the top quark, an on-shell *W* boson is produced. Single top-quark production is a very interesting topic, because it offers direct access to the measurement of the V_{tb} element in the CKM matrix and may also give information about physics beyond the SM. Measurements of single top-quark production were first done at the Tevatron. Both CDF and D0 published their results in 2009, yielding a production cross-section for the *s*- and *t*-channel combination of $\sigma = (2.76\,^{+0.58}_{-0.47})$ pb at 1.96 TeV centre-of-mass energy [1].

The high energies at the Large Hadron Collider made it possible for measurements of individual single top-quark production channels to be performed. At the LHC, the most common one of these production mechanisms is the *t*-channel, which is expected to constitute 76.2% of the total single top-quark production. The ATLAS collaboration reported a $\sigma = (95 \pm 2\,\text{(stat.)} \pm 18\,\text{(sys.)})$ pb cross-section for this channel, after analysing 5.8 fb^{-1} of data collected at $\sqrt{s}$ =8 TeV [2].

The LHC also made it possible for the *Wt* channel, which at $\sqrt{s} = 7$ TeV constitutes only about 18% of the single top-quark production cross-section, to become visible. The CMS collaboration recently

reported a measurement of the Wt cross section in the dilepton decay channel [3]. The result is in good agreement with the theoretical predictions and has over 5σ significance. ATLAS has seen, until now, only evidence of the associated production of a *W* boson and a top quark in the dilepton channel. The measured cross-section is $\sigma = (16.8 \pm 2.9\,(\text{stat.}) \pm 4.9\,(\text{sys.}))$ pb [4] and was determined using 2.05 fb^{-1} of the data collected at 7 TeV centre-of-mass energy.

The research presented in this thesis is a contribution to the search for single top-quark associated production in the lepton + jets decay mode. Unfortunately, this measurement is dominated by systematic rather than statistical uncertainties so it is of high importance to develop new analysis methods and tools that will contribute to the validation of the results. The main difficulty of this analysis and, as previous studies suggest, also the main cause of systematics uncertainties, is the difficulty in separating signal from background. The main sources of background are $t\bar{t}$ production and events in which an on-shell *W* boson is produced in association with jets, because of many similarities between the final states of these events and the *Wt* final state. In order to achieve better signal and background separation, the analysis strategy uses a neural network that is "trained" to distinguish between signal and the main sources of background. 2D distributions of neural network outputs are created. The resulting templates are then fitted to the data and the cross-section, systematic uncertainties and significance of the result are calculated.

One of the possible ways to improve the isolation of the *Wt* signal from the main background sources is using kinematic fitting. Its implementation for the *Wt* channel using the KLFitter package and also the method validation and an evaluation of its performance can be found in [5]. This thesis investigates and quantifies the improvement that can be achieved by including variables obtained using kinematic fitting by looking at 4.7 fb^{-1} data collected by the ATLAS detector in 2011. In comparison to previous studies, systematic uncertainties were also evaluated. The analysis is performed in the region where most of the signal is concentrated (events that contain exactly 3 jets), but the effect of including the 4-jet bin in the analysis and using a dedicated $t\bar{t}$ kinematic fit for that region was also evaluated. Besides the improvement coming from better reconstructing the events using kinematic fitting, additional gain can be obtained when selecting only subsets of events that have a cleaner event topology. This idea was also studied in this thesis, in particular by extracting signal only from events that contain a hadronically decaying top quark or events in which the light quark jets are matched to truth particles from the Monte Carlo information.

The necessary information in order to understand the theoretical and experimental basis, along with the main steps of the performed studies and their results are described in this thesis. Chapter 2 familiarizes the reader with the Standard Model of particle physics, as well as introducing some general unit conventions and main properties of the elementary particles. The focus is, however, on the top quark. The next chapter, 3, presents the machinery that produces and collects the necessary data for this analysis, namely the ATLAS detector. A special section is dedicated to the description of how particles are identified in the detector and how the physical objects are reconstructed. The first part of Chapter 4 offers insight into the event topology of the investigated decay mode of the *Wt* channel and compares that to the main sources of background. The second part focuses on the analysis strategy itself and the necessary tools, starting with a description of the samples that were used (from Monte Carlo simulation or collected data) and the selection that is performed in order to select only events that have a final state that corresponds to that of the lepton+jets *Wt* hypothesis. Chapter 5 presents the principles of kinematic fitting and a brief description on how this is included in the analysis. More studies on how to improve signal and background separation and constrain systematic uncertainties by selecting only regions with a cleaner event topology and increased signal fraction are given in Chapter 6. In the end, Chapter 7 draws the conclusions of this thesis and summarises the results of the performed studies.

Theoretical concepts

The field of high energy physics provides a good laboratory for understanding the properties and interaction mechanisms of elementary particles. The term "elementary" refers to objects for which, using the available experimental methods, no internal structure has been yet discovered. Because of the really small scales that one wants to probe, very high energies are necessary in order to conduct such experiments.

This chapter offers an overview of the theory related to the field of particle physics. The first section introduces units of measure and quantities that are often used in experimental high energy physics. Section 2.2 gives an overview on the Standard Model, while Section 2.3 focuses on the physics of the top quark. The information presented here is summarised from [6] or [7].

2.1 Particle physics basics

Some quantities, notations and units that are important when talking about experimental particle physics are described in the following paragraphs. Getting an overview of these concepts is very important before looking into more detail at the theory that deals with particles and their interactions.

Units of measure

One of the most important characteristics of particle physics experiments is the energy at which the interactions take place. In atomic physics one takes as a unit the energy gained by an electron when passing through a potential difference of 1 Volt. This is denoted with eV and is equivalent to 1.6×10^{-19} J. However, when looking at very small scales the energies are much higher, so more frequently the multiples MeV, GeV or TeV are used.

Additionally, in experimental particle physics one defines the so called natural units. This system redefines units in such a way that two of the most widely used constants in relativistic quantum mechanics (the reduced Plank constant $\hbar = 1.054 \times 10^{-34}$ J s and the speed of light $c = 2.99 \times 10^8 \, \mathrm{m\,s^{-1}}$) are equal to one. Setting $c = \hbar = 1$ leads to a simplification of the usual equations. This convention will be used throughout the thesis; in particular energy, momentum and mass will be given in MeV or GeV.

Typical observables in decays, scattering and collisions

Two very important quantities when identifying a particle are its mass and lifetime. The lifetime (τ) represents the average time that a particle at rest lives before decaying.

The average lifetime is directly linked to the total decay width of a particle (Γ) by the formula:

$$\Gamma = \frac{\hbar}{\tau}.$$

This represents the probability per unit time of the considered particle to decay and can be computed as the sum of the partial widths of all the distinguishable final states: $\Gamma = \sum_i \Gamma_i$. This is measured in eV. Using these quantities one can compute the probability of a certain particle to decay into a particular final state. This quantity is called branching fraction and is defined as:

$$B_i = \frac{\Gamma_i}{\Gamma}$$

for a particle with total width Γ decaying into a particular channel i.

When going from individual particles to certain types of processes one has to define other quantities in order to describe the interaction. An example for such a quantity is the cross section. This is a measure of how likely it is for a certain type of interaction to take place. It is usually measured in barns ($1\,\text{b} = 10^{-24}\,\text{cm}^2$) and denoted with σ.

The rate of any given reaction that can take place in a particle physics experiment depends on the cross section of the considered process and other parameters that are specific to the experiment. In collider physics these parameters are: the number of particles per bunch, the collision frequency and the vertical and horizontal widths of the beam profile. These are all contained in a quantity named luminosity. It is denoted by $\mathcal{L}$ and can be defined as:

$$\mathcal{L} = \frac{R}{\sigma},$$

where R is the interaction rate and σ is the cross section of the process. In order to directly measure the number of expected events one can integrate the previous formula over time, yielding

$$N_{events} = \sigma L,$$

where $L = \int_0^t \mathcal{L}\mathrm{d}t$. The integrated luminosity, L, is a measure of the amount of data collected in a certain experiment and has the dimensions of an inverse cross section (fb^{-1}, nb^{-1}, ...).

2.2 The Standard Model of particle physics

The Standard Model (SM) of particle physics is a theory that encompasses the current understanding on the structure of matter, in particular, on which particles are elementary and how they interact with each other. Two types of particles are included in the Standard Model: fermions (matter particles) and bosons (force carrier particles).

This model has so far proven to be very successful, experimental results from LHC and many other experiments being in good agreement with the theoretical prediction. This can be seen in figure 2.1, where several Standard Model total production cross section measurements are compared to the theoretical expectations. The ATLAS results are, so far, in excellent agreement with the theoretical calculations over five orders of magnitude.
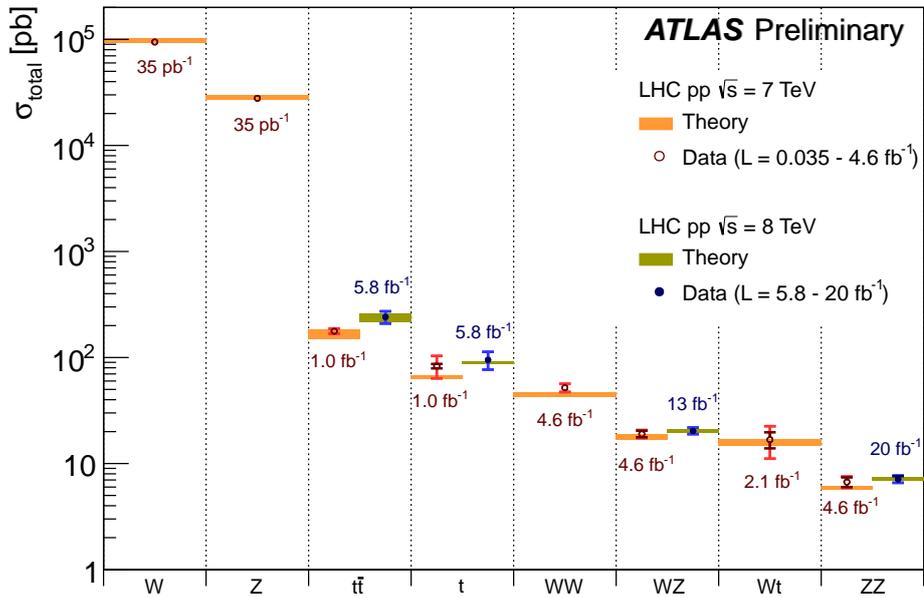
Figure 2.1: Summary of several Standard Model total production cross section measurements, compared to the corresponding theoretical expectations[8].
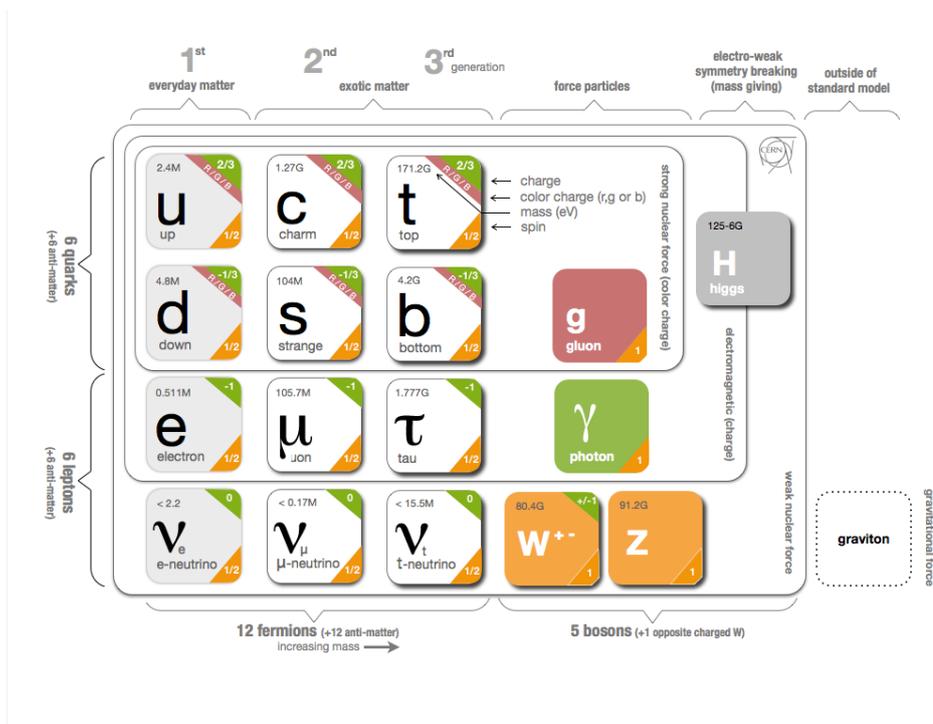
Figure 2.2: Constituent particles of the Standard Model: quarks, leptons and gauge bosons, including properties like mass, electric charge and spin [9].

2.2.1 Ingredients of the Standard Model

Figure 2.2 gives an overview of the constituents of the SM and also their properties and classification. Quite often fermions are termed as "the fundamental building blocks of matter".

They are spin 1/2 particles that can be classified into leptons and quarks. Each of these subclasses has 6 particles that can be arranged into three generations and for each of these particles there exists also a corresponding antiparticle that has the sign of charge and flavour numbers reversed.

The lightest charged lepton is the electron. It has been discovered in 1897 by J. J. Thomson and has a mass that is 1836 times smaller than that of the proton. Its charge (1.602×10^{-19} C) is used as the standard unit of electric charge for other subatomic particles. In addition to the electron, two more charged leptons were added to the Standard Model, the muon and the tau lepton. Their masses are much larger than that of the electron, hence their lifetime is quite small (of the order of 10^{-6} s for the muon and 10^{-13} s for the tau).

The other three leptons are neutral particles with very small masses called neutrinos. Pauli postulated the existence of neutrinos after studying the continuous energy spectrum of nuclear β-decay. Their direct detection happened much later when looking at anti-neutrinos produced in nuclear reactors interacting with protons and resulting int the creation of a positron and a neutron.

Three different types of neutrinos exist, reaffirming the symmetrical nature of the Standard Model and supporting the separation of the leptons into generations of distinct flavour.

The other type of fermions are called quarks. The six quarks are: up, down, charm, strange, top and beauty. They carry fractional charges and are usually organised in pairs containing one quark with charge $+\frac{2}{3}e$ (up-type quarks) and one with $-\frac{1}{3}e$ (down-type quarks). Unlike leptons, quarks cannot exist in a free state and are only observed in bound states named hadrons (the top quark is somehow special in this regard; this will be addressed in detail in Section 2.3). This behaviour is referred to as quark confinement and it stems from the conservation of an additional quantum number called colour charge. Quark exist in one of three so-called colours (red, green, blue), while anti quarks will be anti-red, anti-blue or anti-green. The hadronic bound state formed by this type of fermion must be colour neutral, leaving the simplest possibility of $q\bar{q}$ (mesons) or qqq (baryons) combinations.

2.2.2 Particle interactions

So far all interactions that occur in nature can be explained using only four fundamental forces: strong, electromagnetic, weak and gravity. Each of them has a different strength and acts on a different scale. The Standard Model includes the first three forces and their carriers and does not incorporate gravity. Although the gravitational force has an infinite range, for particle physics its effects are very small and can be neglected.

Before getting into details and describing each type of interaction, it is certainly worth mentioning one of the main tools used in particle physics to depict such processes: Feynman diagrams.

Feynman diagrams are precise graphical representations of mathematical expressions that allow us to better visualise particle interactions and also provide an easier set-up for the calculations of transition amplitudes between different quantum states. In this schematical view each particle is represented by a line (solid lines for fermions and wavy or dashed for bosons) and each crossing of this lines constitutes an interaction vertex. A detailed description of Feynman rules for calculating matrix elements can be found in [6].

The SM interactions and their corresponding exchange particles will be described in the second half of this section. The electromagnetic and weak interaction are presented together, since in the Standard Model Lagrangian these two are unified and characterised by the $SU(2) \times U(1)$ gauge group .

Strong interaction

The gauge field theory describing the strong interaction is called Quantum Chromodynamics (QCD). It represents the $SU(3)$ component of the Standard Model. As the name suggests, this interaction is based on the exchange of colour charge, which is carried by gluons. Due to the fact that leptons don't carry any colour they are not affected by the strong force.

The Feynman diagram for the fermion-gluon chromodynamic vertex is shown in 2.3a. In this quark-gluon interaction the incoming and outgoing fermions can have different colours but the same flavour. Because of colour charge conservation, the gluon will thus carry one unit of the colour charge of one quark and the opposite colour charge of the other quark. Since gluons carry colour themselves, they can directly couple to each other, giving rise to vertices such as 2.3b.

These interactions between the quark current and the gluon fields as well as the gluon self-coupling appear in the QCD Lagrangian. The fundamental parameters of this theory are the masses of the quarks and the strong coupling α_s.

The coupling strength α_s is proportional to the square of the colour charge. For interactions involving small distances (thus large momentum transfer, q^2) α_s is small, while for interactions that have a small 4-momentum transfer, the coupling strength gets larger.
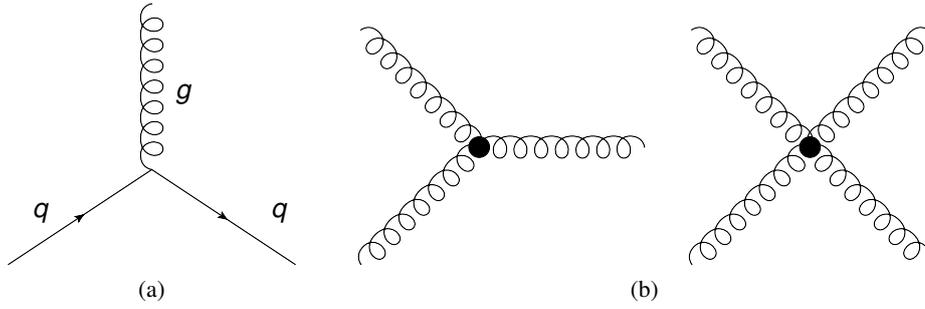
Figure 2.3: The fundamental vertices for QCD: (a) quark-gluon coupling, (b) gluon-gluon coupling.

Factorisation and parton distribution functions

One of the basic properties of QCD is the fact that it is possible to separate the dynamics of different scales. This is known as factorisation. In hard scattering between hadrons, this means that the partons from each hadron involved in the process interact at short distances (where $\alpha_s(Q) << 1$), while the system of partons in the hadron that interacts strongly at nuclear distances is non-perturbative.

The cross-section for a final state X resulting from proton-proton scattering can be written as:

$$\sigma_{pp \to X} = \sum_{ijk} \int dx_1 dx_2 dz f_i(x_1, \mu) f_j(x_2, \mu) \times \hat{\sigma}_{ij \to k}(x_1, x_2, z, Q^2, \alpha_s(\mu), \mu) D_{k \to X(z,\mu)},$$

where f_i are the parton distribution functions (PDFs), i, j, k are the indices corresponding to the partons and $\hat{\sigma}_{ij \to k}$ is the hard parton cross-section that can be calculated in powers of α_s in perturbative QCD. $D_{k \to X(z,\mu)}$ denotes a fragmentation function (or jet algorithm) that makes the transition from the perturbative hard partons to the final state observed particles. All quantities are functions of the renormalisation and factorisation scale $\mu_R = \mu_f = \mu$ that are often chosen to be the same [7].

Parton distribution functions in the protons are available for quarks, antiquarks and gluons. They represent the momentum fraction x_i carried by the parton i. Since PDFs cannot be calculated in perturb-

ation theory, they are extracted from fits to experimental data. Various groups worldwide produce PDFs sets. A library providing an interface to all major PDF sets can be found at [10].

Production rates of all interesting processes rely on the calculation of $\hat{\sigma}_{ij \to k}$. A first estimate of the hard parton cross-section is the leading order approximation in perturbative approach. For more precise theoretical calculations at least the next-to-leading order (NLO) radiative corrections have to be taken into account.

Electroweak interaction

The electromagnetic force has an infinite range and manifests itself through the exchange of a virtual photon. The quantum field theory that deals with this interaction type is quantum electrodynamic (QED). The main process that occurs in QED is the absorption or emission of a photon by a charged particle.

The mediators for weak interaction processes are the $W^{\pm}$ and Z bosons. These are heavy particles with masses of around 80 to 90 GeV. The weak interaction allows the flavour change of quarks and violates parity and charge conjugation symmetry.

When talking about electroweak interactions one must distinguish between left- and right-handed fermion fields. The left-handed fields have weak isospin $T_3 = \pm 1/2$ and form doublets:

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

while the right-handed fields have $I = 0$ and can be written as:

$$e_R, u_R, d_R, \quad \mu_R, c_R, s_R, \quad \tau_R, t_R, b_R.$$

The singlets are invariant under the weak isospin transformation. The SM does not include right-handed neutrinos and also considers all neutrinos to be strictly massless.

This structure, $SU(2)_L \times U(1)_Y$, forms the group of gauge transformations under which the Lagrangian stays invariant. Because of the mass difference between the photon and the other gauge bosons ($W^{\pm}, Z^0$) the symmetry must be broken. This is done by including the Higgs mechanism which is equivalent to a single Higgs fields, constituting a doublet under $SU(2)_L$. The complete electroweak Lagrangian can then be written as:

$$\mathcal{L}_{ew} = \mathcal{L}_{WB} + \mathcal{L}_F + \mathcal{L}_H + \mathcal{L}_Y,$$

where the first term is related to the gauge field, $\mathcal{L}_F$ is for fermion fields, $\mathcal{L}_H$ includes the previously mentioned Higgs mechanism and the last term is related to Yukawa interactions.

Yukawa interactions with the Higgs field are responsible for the origin of quark masses and mixings in the SM. When looking at the flavour changing quark interaction (via $W^{\pm}$ exchange) in the fermion Lagrangian, a unitary quark mixing matrix is introduced. This is called Cabibbo-Kobayashi-Maskawa (CKM) matrix and is a 3×3 matrix with elements $(V_{CKM})_{ij}$ that relate a left-handed up-type quark of generation i with a left-handed down-type quark of generation j:

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

The relevant parameters for this unitary matrix are three rotation angles and one complex phase. There are many parametrisations available for the CKM matrix. One of them is the Wolfenstein para-

metrisation:

$$V_{CKM} = \begin{pmatrix} 1 - \lambda^2/2 & +\lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & +A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + O(\lambda^4).$$

In this representation the expansion is done using the small parameter $\lambda \simeq 0.2$ and CP-violation becomes visible for a non-zero value of η. One can see from the structure of the CKM matrix that transitions between same generation quarks are favoured. In particular, a top quark decays almost exclusively into a *b*-quark.

The entries in the CKM matrix, together with the masses of the Standard Model particles, are some of the parameters of the electroweak theory. Measuring them with very high precision is important for testing the validity of the Standard Model theory and also because it helps constraining the other parameters by looking at the loop contributions to the electroweak observables.

2.3 Top-quark physics

The existence of the top quark has been established even before its experimental discovery, in 1995, by making precision measurements of electroweak vector boson masses and couplings.

This section describes the main aspects of top-quark physics; the information summarised here relies mostly on the latest top quark review available from the Particle Data Group [11].

2.3.1 General properties

The top quark is the heaviest particle included in the Standard Model. Together with its down-type partner, the beauty quark, the top quark forms a weak isospin doublet. Its electrical charge is $+\frac{2}{3}e$ and $T_3 = \frac{1}{2}$.

The top quark lifetime is of the order 0.5×10^{-24} s, which is smaller than the typical time necessary for the hadronization process, making it impossible for the top quark to form bound states. Because of this, its properties have been investigated in much detail and its mass has been measured with a very good precision.

2.3.2 Production mechanism

In hadron collisions the production of the top quark occurs either through the strong interaction, in which case a top and an antitop quark are produced, or through the weak interaction.

Top quark pair production

Producing a top - antitop quark pair by colliding hadrons occurs via one of the two QCD processes: quark annihilation ($q\bar{q} \longrightarrow t\bar{t}$) or gluon fusion ($gg \longrightarrow t\bar{t}$). In proton-antiproton colliders, such as the Tevatron, the first process is dominating, while the later one only accounts for 15% of the total number of top quark events. Theoretical calculations for the $t\bar{t}$ cross section at the Tevatron with a center of mass energy of 1.96 TeV at approximate NNLO yields a 7.3 pb for an assumed mass $m_t = 173$ GeV [12].

At the Large Hadron Collider, in *pp* collisions, 80% of the top quark pairs were produced through gluon fusion at a center of mass energy of 7 TeV and this will get to 90% once the design centr-of-mass energy is reached. For the LHC energy regime, recent cross section calculations have been done at next-to-next-to-leading order in QCD, including resummation of next-to-next-to-leading logarithmic (NNLL) soft gluon terms. For $\sqrt{s} = 7$ GeV the result is $\sigma_{t\bar{t}} = (177 ^{+10}_{-11})$ pb [13]. The corresponding

uncertainties correspond to the sum in quadrature of the scale and PDFs uncertainties. For $\sqrt{s} = 8\,\text{GeV}$ the cross section has been evaluated to $\sigma_{t\bar{t}} = (253\,^{+13}_{-15})\,\text{pb}$. The assumed top-quark mass is 172.5 GeV for both results. So far all the measurements are in good agreement with the theoretical predictions.

Single top-quark production

The production of single top quarks occurs less often than $t\bar{t}$ production and therefore has been much more difficult to measure. Feynman diagrams of the three different production mechanisms at leading order are shown in figure 5.8. All single top processes involve the electroweak Wtb vertex and, can therefore be used to measure the V_{tb} CKM matrix element.
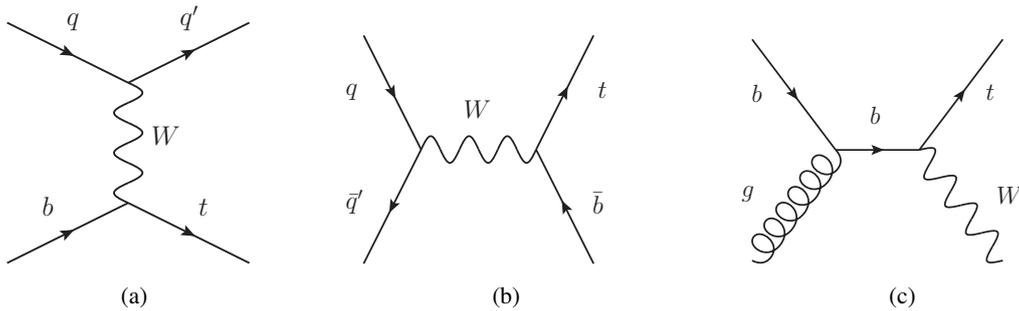
Figure 2.4: Feynman diagrams of single top-quark production processes at leading order: (a) t-channel production, (b) s-channel production and (c) Wt production.

The cross sections of the three modes of producing single top quarks only add up to half of the $t\bar{t}$ cross section at the LHC starting energy. Of these, the t-channel is the dominant one. Shown in figure 2.4a, $qb \longrightarrow q't$ processes are mediated by a virtual W boson. Even at 7 TeV, the t-channel cross section is expected to be almost 40 times as large as at the Tevatron. Approximate NNLO theoretical predictions yield $\sigma_{t-\text{channel}} = 41.7$ pb for t and 22.5 pb for $\bar{t}$. Note that for events coming from $p\bar{p}$ collisions, the production cross sections for top and antitop are identical. This is not the case in pp events because the colliding parton can only be an antiquark if it is coming from the sea quark-antiquark pairs. Thus it is more likely to produce t quarks than $\bar{t}$ at the LHC.

The second type of single top-quark production occurs in the s-channel and is again mediated by an off-shell W boson (figure 2.4b). $q\bar{q}' \longrightarrow t\bar{b}$ events constitute only a small fraction of the total number of events in which a single top quark is produced (roughly 5%). The very low cross section, $\sigma_{s-\text{channel}} = 3.2$ pb , will make it very difficult for this channel to be measured even with the large amount of data collected at the LHC.

One example for a Wt channel leading order diagram is depicted in figure 2.4c. The b quark, that interacts with a gluon, will emit an on-shell W boson and will change into its up-type partner, the top quark.

At next-to-leading order the Wt channel gets additional contributions from diagrams in which a top quark from an internal line can become on-shell. These diagrams actually represent leading-order $t\bar{t}$ processes in which one of the top quarks decays into a b quark and a W boson. Given the very high cross section of top-pair production, when including the formerly described diagrams, the next-to-leading order correction becomes larger than the leading order calculations. More details about this issue and how the Wt channel is then defined at NLO in order to be able to make theoretical predictions, as well as to use this in the generation of MC events, can be found in [14] and [5].

2.3.3 Decay modes

The top quark electroweak decay probability is dominated by the two-body channel $t \longrightarrow bW$, the value of the corresponding CKM matrix element V_{tb} being close to 1. Given its very high mass, the top is the only quark that can produce an on-shell W boson. One advantage of the fact that the weak decay takes place before the hadronization process occurs is the fact that the width of the $t \longrightarrow bW$ transition can be calculated without involving non-perturbative QCD [15]. In particular, we can write Γ_t as predicted in NLO calculations from the Standard Model as

$$\Gamma_t = \frac{G_F m_t^3}{8\pi \sqrt{2}} \left(1 - \frac{m_W^2}{m_t^2}\right)^2 \left(1 + 2\frac{m_W^2}{m_t^2}\right) \left[1 - \frac{2\alpha_s}{3\pi}\left(\frac{2\pi^2}{3} - \frac{5}{2}\right)\right],$$

if we neglect terms of the order m_b^2/m_t^2, α_s^2 and $(\alpha_s/\pi)M_W^2/m_t^2$.

Another consequence of the top quark decaying so rapidly is the fact that the spin information is transferred directly to the decay products. This leads to the possibility of studying the top polarisation by looking at angular distributions of the decay products.

ATLAS and the LHC

The TeV energy regime was first explored with the Tevatron accelerator at Fermilab. The Tevatron collided protons and anti-protons, reaching a maximum centre-of-mass energy of 1.96 TeV. Data taking ended in September 2011.

Taking particle physics experiments even further, the Large Hadron Collider (LHC) was designed to investigate proton-proton collisions with centre-of-mass energies up to 14 TeV. The first collisions were recorded in 2009 and data taking continued until 2013, when the LHC was shutdown for upgrades that will allow the design energy to be reached.

The first part of this chapter presents an overview of the accelerator and its main experiments. In the second half the ATLAS detector is presented, giving a description of its main components and how these are used for detecting particles.

3.1 The Large Hadron Collider

The LHC is a proton-proton collider operated by the European Center for Nuclear Research (CERN) and located on the border between France and Switzerland. The 27 km long accelerator is located 100 m underground and was designed to reach a luminosity $L = 10^{34}\,\mathrm{cm^{-2}\,s^{-1}}$ and beam energies of 7 TeV.

The Large Hadron Collider was designed to reach this high centre-of-mass energy so that it provides a perfect set-up for performing experiments that can test the validity of the Standard Model and also search for other particles predicted by theories and models of physics beyond SM, such as supersymmetry.

The LHC has so far delivered 5.5 fb^{-1} of data in 2011 at $\sqrt{s} = 7$ TeV and 22.8 fb^{-1} at the maximum centre-of-mass energy reached (8 TeV) in 2012 [16], [17]. Operation was stopped in 2013 for almost two years in order to upgrade the machine for dealing with increasing energies. Since the start of its operation, the LHC has provided data that lead to many important findings in particle physics, such as: the discovery of the Higgs boson or the observation of first evidence for the very rare decay $B_s \rightarrow \mu^+\mu^-$.

The accelerated particles are either protons or lead ions. For the Pb ion collisions the expected maximum centre-of-mass energy is even higher, going up to around 1150 TeV.

Before being injected in the LHC ring, the protons first are accelerated until they reach an energy of 450 GeV. In order to do that they are circulated from the proton source through several smaller accelerators, including a linear one (LINAC 2) and three circular ones (the Proton Synchrotron Booster, the Proton Synchrotron and the Super Proton Synchrotron). Particles are then injected in the LHC ring in bunches. Each beam consists of 1350 bunches (in the current operating conditions) and each bunch

contains approximately 10^{11} protons. After a 25 minutes time of acceleration that allows them to reach the desired energy, the proton beams circulate in the accelerator for about ten hours.

Along the LHC tunnel there are four main experiments, located at the interaction points of the colliding beams. These are shown in figure 3.1. Two of them are very large general purpose detectors (ATLAS and CMS) while LHCb focuses on *b* physics and studies charge-parity violation by looking at the decays of *B* hadrons and ALICE is dedicated to the study of data resulting from heavy ion collisions.
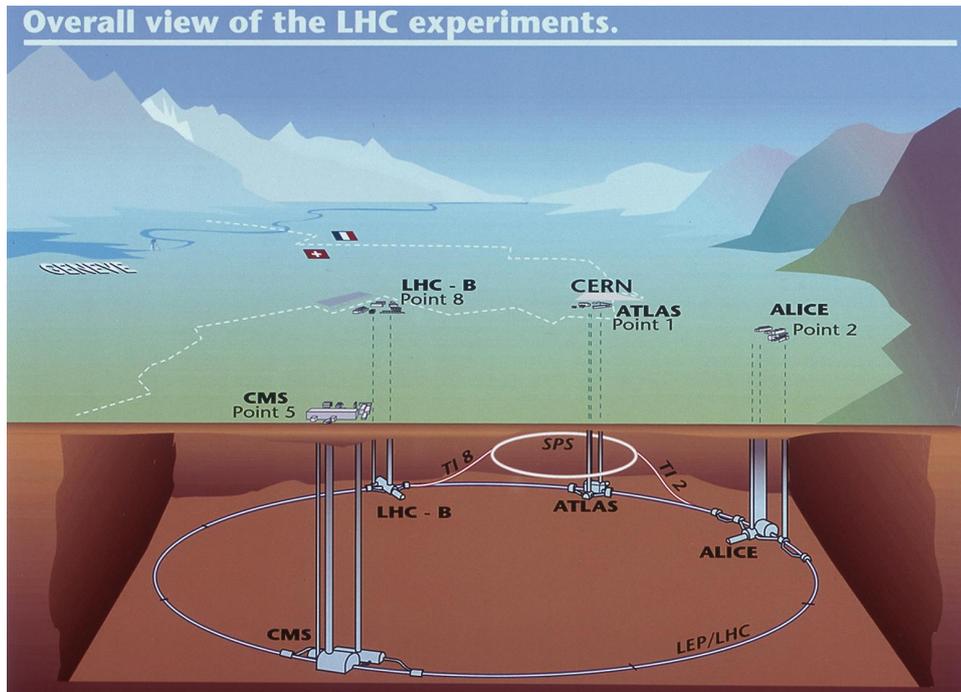
Figure 3.1: The Large Hadron Collider.

Since the dataset used for the analysis presented within this thesis was collected with the ATLAS detector, the following section is dedicated to a more detailed description of this experiment. Further details about the other detectors can be found in [18].

3.2 ATLAS experiment

ATLAS stands for "A Toroidal LHC ApparatuS" and is the name of one of the two multipurpose detectors along the LHC ring. Scientists from over 170 research institutes across the world are involved in the collaboration that analyses the data collected by the ATLAS detector.

The ATLAS detector was designed such that it provides high efficiencies for most physics processes of interest at Large Hadron Collider. It has an impressive size, weighing 7000 tons and having a length of 46 m and 25 m diameter.

3.2.1 ATLAS coordinate system

The ATLAS coordinate system is right-handed and has the origin in the nominal interaction point. The beam direction defines the *z*-axis. The perpendicular plane, also referred to as the transverse plane, is

determined by the *x*- and *y*-axes. The *x*-axis points towards the centre of the LHC ring while the *y*-axis has a vertical direction, pointing upwards.

Quantities such as transverse momentum, transverse energy or missing transverse energy are defined in the *xy* plane. The use of this information allows new constraints on the event kinematics, such as energy and momentum conservation in the transverse plane.

The symmetric geometry of the detector makes the use of a spherical coordinate system convenient. The azimuthal angle is denoted with ϕ and is measured in the *xy* plane (around the *z*-axis). The polar angle θ is measured with respect to the beam axis. Another way of describing the angle of a particle with respect to the *z*-axis is by using pseudorapidity. This is defined as

$$\eta = -\ln[\tan\frac{\theta}{2}].$$

For highly relativistic particles ($\beta \rightarrow 1$) this is a good approximation of the rapidity *y*, which is often used in theoretical calculations and is defined as

$$y = \frac{1}{2}\ln\frac{E + p_z}{E - p_z}.$$

Distances ΔR are often measured in the pseudorapidity-azimuthal angle space as:

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

3.2.2 ATLAS detector

The detector has a structure typical for colliding beam experiments, with all detector components directly surrounding the beam pipe and being symmetrically located around the interaction point, providing almost full azimuthal angle coverage.

Figure 3.2 shows a view of the full detector and its main components: the inner detector, electromagnetic and hadron calorimeters, the muon system and the magnet system. A brief description of each of these systems is given in the following paragraphs. More detailed information can be found in [19] and [20].

Inner detector (ID)

The detector component closest to the beam pipe is the inner detector. This is the central tracker of the detector and plays an important role in momentum and vertex measurement, as well as in charge identification. The main components of the inner detector are the pixel detector (Pixel), semiconductor tracker (SCT) and transition radiation tracker (TRT). All of them are located in a 2 T magnetic field generated by a solenoid magnet and cover a range of pseudorapidity $|\eta| \leq 2.5$.

The **pixel detector** consists of 3 barrel layers and 3 disks located in the forward regions of the detector. Each pixel layer is segmented in R-ϕ and *z* and, typically, each track crosses at least three of these layers. The barrel layers contain about 67 millions pixels and the ones located on the endcaps approximately 13 million. Each of the pixels has a size of $50 \times 400\ \mu m^2$. The position resolution of the Pixel in R-ϕ is 10 μm for both barrel and endcap regions. The corresponding *z*/*R* resolution is 115 μm.

The **SCT** is a silicon microstrip tracker consisting of 4 cylindrical barrel layers and 18 planar endcap disks. The readout is done via readout strips that are placed every 17 μm on the 60 m^2 of material. In the transverse direction to the strips, the measurement of the position of electrically charged particles is done with a precision of up to 17 μm. The *z*/*R* resolution for the barrel/endcap regions is 580 μm.
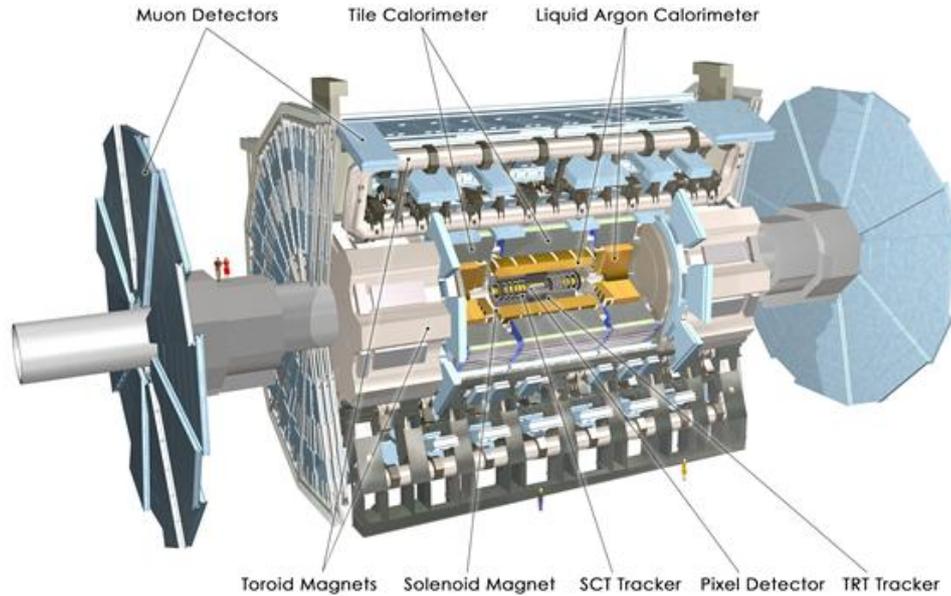
Figure 3.2: Overview of the ATLAS detector.

The **transition radiation tracker** is the largest component of the inner detector, having a volume of about $12\,\mathrm{m}^3$. Its coverage extends up to $|\eta| \leq 2$. The TRT is composed of $4\,\mathrm{mm}$ diameter straw tubes. Each of these straw tubes acts as an individual drift chamber in which the signal wire is located in the center of the straw[1]. The straws are arranged along the direction of the z-axis. In the barrel region they only provide little information in the longitudinal direction but do measure information from the transverse plane. In the forward region the straws are uniformly distributed in ϕ and only provide measurements in z and the azimuthal direction. In total, there are over $50\,000$ straws in the barrel region. The endcaps regions have approximately $300\,000$ straws that are 4 times shorter. Each of them is read-out separately.

Calorimeters

Calorimeters are located between the inner detector and the muon system. They contribute to particle identification and are used for measuring the energy of the particles. In order to provide good energy measurements for determining the missing transverse energy in the event, the calorimeters cover pseudorapidities between -4.9 and 4.9. The system consists of an electromagnetic calorimeter and a hadron one.

The **electromagnetic calorimeter (ECAL)** is composed of three parts. The central region, $|\eta| < 1.475$ is covered by the barrel, while the endcaps each have two coaxial wheels that cover, in total, the region $1.375 < |\eta| < 3.2$. The ECAL uses liquid argon as an active material and lead as an absorber. A complete ϕ coverage and no azimuthal cracks are obtained by using an accordion geometry. In the barrel, the electromagnetic calorimeter has a total thickness of more than 24 radiation lengths, X_0, while in the forward and backward regions this is larger than $26X_0$.

For the regions that require a very high precision (pseudorapidity in the range $[-2.5, 2.5]$) the ECAL is divided into three different sections: the strip section, the middle section and the back section. The strip

---

[1] The wires inside the tubes have a $0.03\,\mathrm{mm}$ diameter and are made of gold-plated tungsten.

section acts as a preshower detector and plays an important role in particle identification and precision measurements in η. All these regions have fine granularity, with the minimum being reached by the first sector, with $\Delta\eta \times \Delta\phi = 0.003 \times 0.1$. In the endcaps, the granularity is coarser, but it is still sufficient for measurements such as jet reconstruction and E_T^{miss} measurements.

The energy resolution for the ATLAS ECAL is [21]:

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 0.7\%,$$

where the energy independent term is related to errors coming from calibration or non-uniformities in the read-out systems or other imperfections in the calorimeter and the first term parametrizes the fluctuations that appear due to the physical development of the shower [7].

The **hadron calorimeters** have a design that is optimised for the detection of jets. They combine different materials in order to provide a good containment of the hadron showers. In the acceptance range $|\eta| < 1.7$, in the barrel region, scintillating-tiles are used. These make up the so-called tile calorimeter. This uses steel as absorber material and scintillating tiles as active material.

In the $1.5 < |\eta| < 3.2$ range a hadron endcap calorimeter is located. This is again a liquid argon calorimeter that uses copper absorber.

Starting from $|\eta| = 3.1$, and up to $|\eta| = 4.9$, the high-density forward calorimeter (FCAL) is located. This detector component is made of copper in the first layers and for the outer layers tungsten is used.

The design energy resolution for the hadron calorimeter is:

$$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E}} \oplus 3.0\%.$$

Muon system

The role of the muon system is, as the name suggests, to identify muons and measure their momenta. The muons are, excepting the neutrinos that do not interact with the detector material, the only particles that do not get stopped in the calorimeters and reach the outer part of the detector. Their momentum is measured by magnetic deflection of the muon tracks in the magnetic field generated by the toroid magnets.

The muon trigger is fired when a particle passes the Thin Gap Chambers (forward region) or the Resistive Plate Chambers (central regions). The combined coverage range of these detectors is $|\eta| \leq 2.4$.

In order to measure the curvature of the tracks Monitored Drift Tubes are used, while at the ends of the detector the coordinates are precisely measured using Cathode Strip Chambers. Each drift tube has a resolution of $80\,\mu$m. In the forward region the resolution is even better, reaching down to $60\,\mu$m.

Magnet system

The magnet system has two components: the central solenoid magnet and the toroid system. Both parts consists of superconducting magnets. The first one is located outside the inner detector and provides the 2 T magnetic field that is necessary for bending the tracks of high momentum particles. The track curvature offers information on the sign of the electric charge of the particle and is used for momentum measurements.

The toroid magnet is the system that gives the name of the detector because of its unique size and design. This is used for bending muons. It consists of 3 parts, one located in the central region and the others are on each of the endcaps.

Each of them has 8 separate coils that create 4 T magnetic fields. The working point temperature for these magnets is 4.7 K, so they have to be placed in a cryostat. Because of its very large size (25 m length), the barrel toroid has a separate cryostat for each of the coils.

Trigger and Data Acquisition

The large amount of information collected by the previously described detector components requires a very sophisticated system for selecting only the interesting events and recording this data. For AT-LAS, the trigger and data acquisition system (TDAQ) manages to select one interesting event out of approximately 200 000 others and has three levels that contribute to the event selection.

The level one trigger is entirely hardware based. Interesting objects are selected based on information collected by subsets of detectors that have reduced granularity. These are usually high transverse momentum leptons[2], jets, τ leptons decaying into hadrons and large missing and total transverse energies. Another important feature of this trigger segment is the fact that it has a latency of only 2 µs. This is defined as the necessary time, starting from the interaction time, for the trigger decision to reach the front-end electronics. For the events that pass this selection step the full detector information is read from the electronics into readout drivers and then into readout buffers.

The second level trigger is designed to reduce the rate from 75 kHz to ~3 kHz. This is done by selectively accessing the data by making use of the information provided by the primary trigger. This means that only information from the detector components located in "regions of interest" (centred around the indicated objects) is used. Due to this feature, the second level trigger makes a decision using only a few percent of the full event data. Events that do not pass this trigger are discarded, while the other ones are transferred by the data acquisition system to storage associated with the event filter. The latency of this component is between 1–10 ms

The third and last part of the trigger system makes the final selections of events that are kept for offline analysis. The information is reduced by a factor of 10 compared to the output of the second level. This corresponds to a data rate of ~300 MB/s for fully recorded events [22].

The raw data recorded in one year by the ATLAS detector amounts to about 3000 Tbytes.

3.2.3 Particle identification and object reconstruction with ATLAS

Particle identification is based on the interactions that take place in the detector material. The layered construction of the ATLAS detector allows for the identification of both neutral and charged particles. Figure 3.3 shows the signatures of representative particles in the ATLAS detector. Electrically charged particles, like electrons, protons or muons, will leave a track in the tracking system. Because of the magnetic field generated by the solenoid, the tracks will be bent. Based on the radius of curvature, one can calculate the momentum of the particle and determine the sign of the electric charge. In the case of photons, there will be no visible interaction in the inner part of the detector, but there will be a significant energy deposition in the electromagnetic calorimeter. Because of their higher masses, hadrons will reach further in the detector, depositing their energy in the HCAL. Muons are the only particles that are detectable only by the outermost layers of the detector. Neutrinos do not interact at all with the ATLAS systems and are therefore detected only indirectly through the imbalance in energy conservation in the transverse plane.

Because of the top-quark decay configuration, a *Wt* event will already contain most of the particles detectable by ATLAS. The corresponding object definitions for electrons, muons, jets and missing trans-

---

[2] unless stated otherwise, here and throughout the thesis "lepton" refers to electrons and muons.
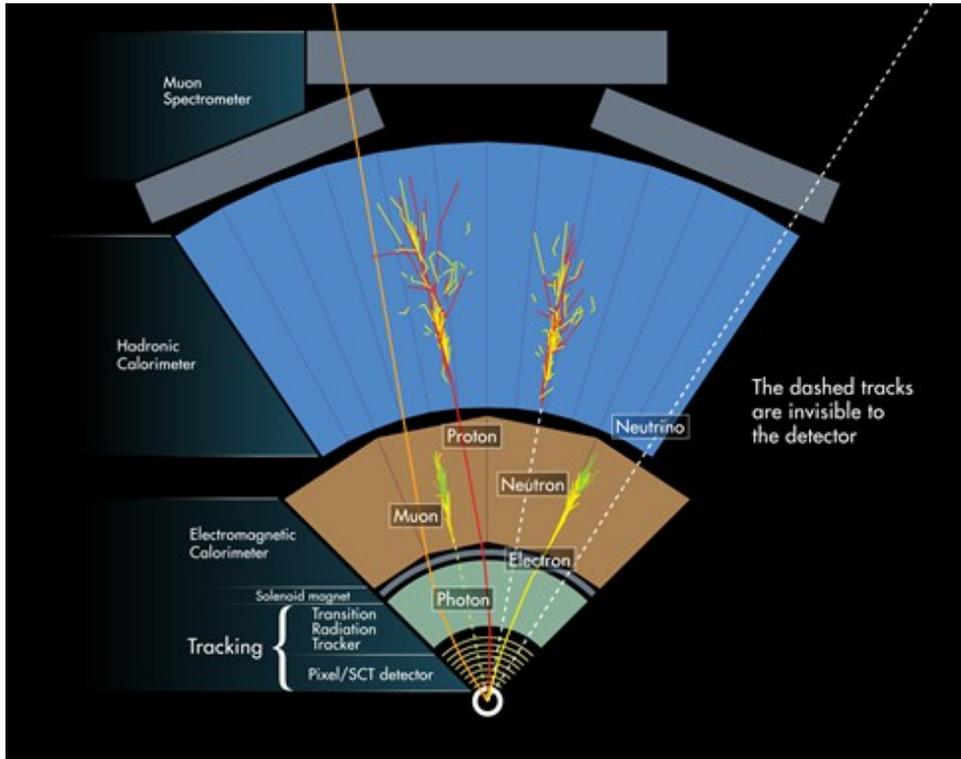
Figure 3.3: Illustration of particle detection in the subsystems of the ATLAS detector.

verse energy are described in the following part. These follow the recommendation of the ATLAS top reconstruction group [23], [24].

Electrons

The transverse energy threshold for selecting electron candidates is $20\,\text{GeV}$ for $\sqrt{s} = 7\,\text{TeV}$ and the pseudorapidity range is between -2.5 and 2.5. The reconstruction algorithm for these objects matches tracks to each of the selected electromagnetic clusters. The tracks must also fulfil quality requirements that depend on variables such as: distance of closest approach to the primary vertex, number of hits in different detector regions or fraction of high-threshold hits in the transition radiation tracker.

Another requirement that further helps the selection is the so called isolation criterion. Because the electron from a top-quark event most likely comes from the $W \longrightarrow e\bar{\nu}$ decay, it can be differentiated from electrons that come from heavy flavour decays or $\gamma \longrightarrow e^+e^-$. This is done by checking the sum of transverse energy within a cone of ΔR around the electron cluster or the scalar sum of transverse momenta of the tracks found in the cone.

This thesis includes only reconstructed electrons that are selected by applying a collection of additional cuts that have the role of suppressing background and include cutting on information from the transition radiation tracker. The analysis specific electron definition imposes even more constraints on the electron candidates that are selected by the criteria described above, such as: the requirements for the object to come from the central region of the detector ($|\eta_{cluster}| < 2.47$)[3] and to have a transverse

---

[3] The region between the barrel and the endcaps, $1.37 < |\eta_{cluster}| < 1.52$, is excluded because of large drops in reconstruction efficiency and energy resolution.

energy of $E_{cluster}/\cosh(\eta_{track}) > 25\,\text{GeV}$. The isolation cuts are made on E_T within a cone of $\Delta R = 0.20$ and p_T with $\Delta R = 0.30$ and are chosen as to keep a constant efficiency of 90%.

In order to reproduce the energy resolution of the collision data, a smearing procedure is applied on all Monte Carlo events. The measured electron energy in data is calibrated in order to correct the cluster energy. Data collected in studies on J/Ψ, W and Z was used for determining the electron energy resolution and energy scale uncertainty [25].

Muons

Muon candidates are required to have at least $p_T = 25\,\text{GeV}$. Their tracks are fit separately in the inner detector and muon spectrometer; combined muons in which the inner detector track is matched to the one reconstructed by the muon system are required in the top analyses. The isolation criteria are similar as for the electrons. The sum of missing transverse energy inside a $\Delta R = 0.2$ cone is required to be less than $4\,\text{GeV}$, while the p_T sum is considered for a cone of $\Delta R = 0.3$ and must be $< 2.5\,\text{GeV}$.

Additionally, other cuts are required. These include: selecting muons that have the distance z_0 with respect to the primary vertex smaller than $2\,\text{mm}$, considering the acceptance $|\eta| < 2.5$ and additional hit requirements. Muons that are closer than $\Delta R = 0.4$ to an already identified jet are discarded.

Jets

The reconstruction of jets is based on the anti-k_t algorithm [26] using topological clusters [27]. The distance parameter is set to $R = 0.4$. Furthermore, jets are removed if they are within $\Delta R = 0.2$ of an already accepted electron and have to be located in the $|\eta| < 2.5$ region of the detector. Jets that have $p_T < 25\,\text{GeV}$ are rejected.

An additional cut, meant to suppress pile-up, is applied on the jet vertex fraction. This discriminant combines tracks and their primary vertices with calorimeter jets and gives a measure of the probability that a jet originated from a particular vertex [28]. The cut value for this quantity is set to $|JVF| > 0.75$.

Missing transverse energy

The missing transverse energy (E_T^{miss}) is assigned to be the measured energy of the particles that escape detection. E_T^{miss} is calculated using calorimeter clusters and muon tracks. Each calorimeter cell is calibrated according to the objects that they are associated to. The calculation is done by vectorially summing up the p_T of the identified photons, electrons, muons and jets and even the energy coming from cells that are not included in cluster reconstruction.

Analysis setup and strategy

In one of the single top-quark production mechanisms the top quark is produced in association with an on-shell *W* boson. This was described in detail, together with top-quark pair production and the other two single top-quark production mechanisms in Section 2.3.2. The first part of this chapter presents the description of the $Wt \rightarrow$ lepton + jets final state and the main sources of background. The techniques and tools used to separate signal from background in this analysis are presented in Sections 4.3, 4.4 and 4.5.

4.1 Lepton + jets decay mode of the Wt channel

Events in which *Wt* single top-quark production occurs contain in the final state a real *W* boson, as well as the decay products that come from the top-quark decay. Already in Chapter 2 it was mentioned that the CKM matrix element V_{tb}, that gives the probability of a top-quark to decay into a *b*-quark and a *W* boson, is almost 100%. This leads to a *bWW* intermediate state of the *Wt* process. The leading order Feynman diagrams of such events are shown in figure 4.1. The bottom quark becomes visible in the detector and produces a jet, while the *W* bosons further decay via one of the well-known channels: leptonic or hadronic. The leptonic decay $W \rightarrow l\nu$ (with $l = e, \mu, \tau$) occurs in about 32.4% of the cases, while the hadronic one, $W \rightarrow q\bar{q}'$ is twice as common.

Our analysis focuses on events in which the two *W* bosons have different decay modes, namely one decays into a lepton and a neutrino while the other one decays into two light quarks. This channel has a branching fraction $\mathcal{B}_{\text{lepton+jets}} = 43.5\%$, which is comparable to the one of the pure hadronic channel ($\mathcal{B}_{\text{hadronic}} = 46.2\%$) and much larger than $\mathcal{B}_{\text{dilepton}}$, which is only 10.3%. Although it is produced most often, an all-hadronic final state is difficult to study because it only consists of jets and this usually suffers from higher uncertainties and also makes it harder to separate from background coming from QCD processes.

The dilepton channel has cleaner features, consisting of a large amount of E_T^{miss}, two isolated leptons and a jet produced by a *b* quark (*b*-jet). The downside of this channel is the small branching fraction but this is compensated by the good signal-background separation and smaller uncertainties. This is also the reason why evidence and discovery from the ATLAS and CMS collaboration respectively have been first announced in the dilepton channel.

The lepton + jets decay mode signature consists of: missing energy from the neutrino that escapes detection, an isolated lepton from the decay of one of the *W* bosons, as well as three jets (one of which
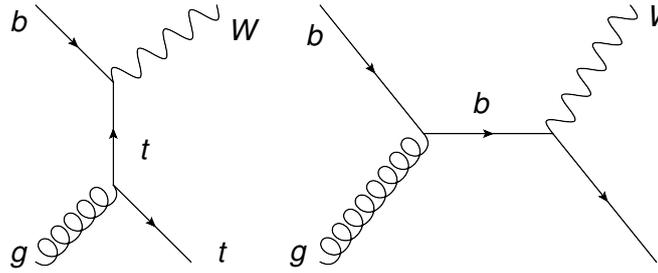
Figure 4.1: Leading order Feynman diagrams of the *Wt* associated production.

has to be *b*-tagged). Requiring the selected events to have all these characteristics in the final state is the first step that is made for separating signal from background.

An important aspect that one has to consider for any final state involving leptons is the fact that not all leptons are directly detectable. For the *Wt* analysis, only events in which a muon or an electron appears are selected. The tau leptons are very short lived and decay almost immediately, either leptonically or hadronically. This makes such events more difficult to detect and distinguish from background and this is why *Wt* events containing a τ lepton are not explicitly included in the signal sample. Events with τs decaying into hadrons are most likely treated as background while the ones in which the tau lepton decays into a μ or an *e* are recovered as signal in the muon and electron channel respectively. The additional neutrinos from the τ decay then contribute to the missing transverse energy.

According to the lepton + jets decay mode presented above, signal events contain 3 jets in the final state. However, sometimes, one of the light-quark jets can be misidentified or other additional jets can occur, so the 2- and 4-jet topologies are also taken into account. The *Wt* search in events containing two jets has been investigated in detail in [29] and is not included in the analysis presented in this thesis. A brief study of the effect of including 4-jet bin events to the analysis was performed for this thesis, more precisely in Section 5.

4.2 Sources of background

Events coming from other processes represent background for the *Wt* analysis if they contain reconstructed physical objects that are common to the *Wt* lepton + jets signature. In particular, the presence of any on-shell heavy boson (*Z* or *W*) or a top-quark is relevant. The main sources of background for the previously described signal is presented in the following paragraphs.

Multijet (QCD)

The multijet background, sometimes called QCD, consists of events in which no on-shell top-quark, *W* or *Z* boson are produced. These processes have a cross-section that is many orders of magnitude higher than any other background process. It is therefore very important to ensure that the cuts that are applied for separating signal from background have a high rejection against multijet events.

QCD events with a topology similar to the signal are very difficult to simulate due to the very high cross-section. Because of this reason, the available Monte Carlo samples for this type of background have limited statistics and therefore are not used in the analysis. Even after the event selection the multijet background is still not negligible and data-driven methods are used to estimate the amount of QCD background. For the muon channel the so-called matrix method is used, while while the multijet

contribution in the electron channel is estimated with the so-called jet-electron model. Both methods, as well as a comparison between the two are described in detail in [29].

Top-quark pair production

Top-quark pair production is one of the background channels that is most difficult to separate from Wt events. One of the reasons for that is the much larger cross-section for top-antitop processes than the single top-quark cross-section. In particular, for $\sqrt{s} = 7\,\mathrm{TeV}$, $t\bar{t}$ events are about 10 times more likely to be produced than events coming from Wt associated production.

Top-antitop production is one of the two largest sources of background, together with W+jets. Separating $t\bar{t}$ events has proved to be the most difficult part of this analysis, mainly because the signature in the lepton+jets decay mode is identical to the Wt final state, with only one additional b-jet. This final state topology indicates that requiring the signal events to only contain exactly one b-tagged jet is a good first step for reducing the $t\bar{t}$ background contribution, at least in the 3-jet bin. In the 4-jet bin, an additional jet is expected in the detector, so the signatures of the two processes are almost identical. This makes $t\bar{t}$ separation even more difficult for events that contain four jets in the final state.

W + jets

The production of a W boson and additional jets is the background with the second largest cross-section, orders of magnitude larger then the one for single top-quark production. When the W boson decays leptonically, into one lepton and a neutrino, and in addition to that one or more jets are produced, the W + jets final state is very similar to the one of Wt events, containing all the important signatures (missing energy, a lepton and jets). However, the signal for W + jets events is concentrated mainly in the lower jet multiplicity bins, the production of each new jet being suppressed by a constant factor. This is referred to as "Behrends scaling" [30]. Thus, cutting on the number of jets when selecting events will greatly reduce background coming from this type of process.

Z + jets

Although it has a cross-section similar to the W + jets one, the Z + jets background is one of the channels that gets separated best from the signal. The Z boson decays into two leptons or hadronically, into a quark and an antiquark pair, and does not imply any E_T^{miss}. Therefore, requiring one lepton in the final state of the signal sample and missing transverse energy will already greatly reduce the Z + jets contribution. Also, Behrends scaling is valid for this process so the jet multiplicity is a useful tool for further suppressing this background channel.

Other single top-quark production channels

In this category the contributions come from the other single top-quark production mechanisms: s-channel and t-channel. The Feynman diagrams for these processes are shown in figure 5.8, Section 2.3.1.

The s-channel has a lower cross-section than that of the Wt channel. Its signature, when the top quark decays semi-leptonically, is very similar to the signal one, except for the fact that instead of the two light quark jets, a b-jet is produced. Because of that, requiring signal events to only have one b-tagged jet greatly reduces this background, making the s-channel the process with the smallest background contribution.

The *t*-channel final state is also quite similar to the *Wt* signal one, the difference coming from the fact that, in the case when the top quark decays semi-leptonically, only one light-quark jet appears. The *t*-channel signal is therefore mostly concentrated in the 2-jet bin. Although the cross-section is 4 times higher than the signal one, the *t*-channel background is also one of the minor backgrounds of this analysis.

Diboson (*WW*, *ZZ* and *WZ*)

Diboson refers to events in which a pair of heavy vector bosons is produced, namely *WW*, *ZZ* or *WZ*. In *WW* events, if one of the bosons decays leptonically and the other one hadronically, the final state is identical to the *Wt* one, except for one additional b-jet. For the *ZZ* events, the same considerations as for *Z* + jets apply, so cutting on E_T^{miss} and requiring only one lepton decreases its contribution. In the case of *WZ* processes, for a hadronically decaying *Z* and a leptonically decaying *W*, the signature is similar to the one of *W* + jets events. Although separation is not straightforward for all of these three processes, this channel has a small cross-section and is therefore not among the problematic background channels.

4.3 Data samples / Monte Carlo (MC) simulations

This section describes the datasets that were used for the analysis presented in this thesis. For the majority of the background processes, as well as the signal, the samples consist of Monte Carlo simulations. Only the QCD contribution is estimated using data-driven methods, as was already mentioned in the previous section.

4.3.1 Data

The data sample consists of collision data collected by the ATLAS detector in 2011, at a centre-of-mass energy of $\sqrt{s} = 7\,\text{TeV}$. The integrated luminosity is $L = 4.7\,\text{fb}^{-1}$. The sample contains events from the muon channel, that are selected from the "Muons" stream using a muon trigger, and from the electron channel, coming from the "Egamma" stream and passing an electron trigger.

Good run lists (GRL)

The data collected by the different detector components is first checked by the Data Quality (DQ) group. Flags are assigned to the detector components, triggers and reconstructed objects for indicating if data taking was successful (e.g. if detectors and reconstruction were working correctly). All DQ flags are then combined into good run lists (GRLs) which contain all the luminosity blocks for which the data is suited for analysis and the corresponding luminosity of each block. A luminosity block is the unit in which the luminosity is measured and typically corresponds to one or two minutes of data taking.

The GRLs are provided by the ATLAS Data Quality group, but specific requirements corresponding to every analysis are also used. The good run lists for this analysis were provided by the top reconstruction group and, additional to the common ones, they include flags related to tracking, vertexing and *b*-tagging.

4.3.2 Monte Carlo datasets

Some default values are set for parameters such as particle masses and widths when generating Monte Carlo events. These include the mass of the top quark, $m_t = 172.5\,\text{GeV}$, and the masses of the *Z* and

W bosons, $m_Z = 91.2\,\text{GeV}$ and $m_W = 80.4\,\text{GeV}$. The corresponding widths are set to $\Gamma_t = 1.32\,\text{GeV}$, $\Gamma_W = 2.09\,\text{GeV}$ and $\Gamma_Z = 2.50\,\text{GeV}$.

The Monte Carlo samples that are used in this analysis come from general-purpose MC generators, such as: MC@NLO [31], AcerMC [32], HERWIG [33], POWHEG [34], PYTHIA [35] and ALPGEN [36] and are provided by the ATLAS production group. For parton showering and modelling of under-lying events PYTHIA and HERWIG/JIMMY [37] are available. The corresponding PDF sets used for these generators are: CTEQ6.6 for MC@NLO, CTEQ6L [38] for ALPGEN and LO* MRST [39] for AcerMC and HERWIG.

All the samples described in the following paragraphs are normalised to the theoretical cross-sections using K-factors. These are higher order QCD corrections to leading order cross-section calculations.

Another aspect that has to be taken into account for the samples is the pile-up configuration. Pile-up interactions occur at high luminosity conditions, when more than one interaction takes place per bunch crossing. The MC pile-up has to be therefore corrected to the pile-up conditions from the collison data. This procedure is called pile-up reweighting and is done for all MC samples.

This thesis consists of two different studies. One that tries to improve the reconstruction by using kinematic fitting, that is presented in Chapter 5, and another one in which the improvement of the analysis by selecting events with cleaner topology is investigated. Between these two studies, the recommendations of the Top Working Group for which default samples (in terms of MC generators) to use for the single top-quark production and $t\bar{t}$ production channels has changed. For these two channels, both MC generator options are listed below.

Single top-quark production

Single top-quark production events were generated, for the KLFitter study, using MC@NLO for the *s*- and *Wt* channel and AcerMC for the *t*-channel. For parton showering MC@NLO is interfaced with HERWIG/JIMMY while AcerMC uses PYTHIA. The simulated *Wt* decay includes all possible channels, while for *s*- and *t*-channel only the lepton+jets decay mode events are generated.

After the change in default samples recommendations, for the *s*- and *Wt* channel samples generated with POWHEG interfaced to PYTHIA were used. For the *t*-channel the AcerMC sample is still the default generator.

Top-quark pair production

For the top-quark pair production, in the first sample configuration, events were generated using MC@NLO interfaced to HERWIG/JIMMY for parton showering. This was done separately for events with a final state involving only hadrons and the non-fully hadronic decay mode.

The new recommendation supports the use of POWHEG plus PYTHIA. Also for this case, the all-hadronic and non-all hadronic decay modes are treated separately.

W + **jets production**

The *W* + jets production samples are first classified in *W* + light flavour (WLF) and *W* + heavy flavour (WHF). Both processes are simulated with ALPGEN interfaced to HERWIG/JIMMY. WLF events consist of one leptonically decaying *W* boson plus a number of jets that corresponds to 0 to 5 partons.

The WHF generated events can be either from $W + b\bar{b}$+jets, $W + c\bar{c}$+jets or $W + c$+jets configurations. For the first two possibilities the number of additional jets varies from 0 to 3, while for the last one it goes from 0 to 4. Separate samples are available for all these combinations.

The W + light flavour and W + heavy flavour samples are not strictly separated in terms of flavour content; some of the events in one of the sample can be possibly found in the other. In order to avoid double counting because of this overlap, a procedure called "heavy flavour overlap removal (HFOR)" is applied. A description of this tool can be found in [40].

Z + **jets production**

Z + jets events are simulated using the ALPGEN generator interfaced to HERWIG/JIMMY. The considered events must have a dilepton invariant mass between 40 and 2000 GeV. Z bosons decaying into a pair of electrons, muons or taus are considered and separate samples for each of these decay modes plus a certain number of partons (ranging from 0 to 5) are provided.

Diboson production

For the simulation of the WW, WZ and ZZ, the used generator is HERWIG. All the decay modes are considered for the W and Z bosons. The only requirement that is applied is to have at least one lepton in the final state with a transverse momentum higher than 10 GeV and $|\eta| < 2.8$.

4.4 Event selection

The selection of events that is used in the actual analysis is done on three levels. First, some general event cleaning cuts are applied. This includes selecting only events in which the primary vertex has more than four associated tracks (vertex cleaning) and rejecting all events that have at least one "bad" jet (either coming from non-collision events or due to calorimeter problems). Also, a "LAr cleaning" has to be applied. This refers to an acceptance gap that appeared because of a problem in the front-end electronics of the liquid argon calorimeter. All events that contain at least one object in that region of the calorimeter are removed from the analysis. For the electrons this step is included in the object quality flag.

The next step in the selection is the pretag selection. Exactly one lepton is required in both the electron and muon channel, with $p_T^l > 25$ GeV. Because of the neutrino that escapes detection, a cut on the missing transverse energy is also applied. This is set to 30 GeV for the electron channel and 25 GeV for the muon one. Because most of the signal is concentrated in the 3-jet bin and also the 4-jet bin topology is very similar to the $t\bar{t}$ one, exactly three jets are required, with $p_T > 25$ GeV and $|\eta| < 2.5$.

In order to suppress the multijet background a cut is also applied on the transverse mass of the lepton and the neutrino. For two particles with very low masses ($m_1 \simeq m_2 \simeq 0$), the transverse mass, m_T, is defined as:

$$m_T = 2E_T^1 E_T^2 (1 - \cos\phi),$$

where $E_T^{1,2}$ are the transverse energies of the particles and ϕ is the angle between them in the transverse plane. The cuts that are applied on this quantity are: for the electron channel $m_T(l\nu) > 30$ GeV and for the muon channel a triangular cut is applied $m_T(l\nu) + p_T^{miss} > 60$ GeV.

For identifying jets originating from b quarks, this analysis uses the MV1 algorithm. This is a neural-network based algorithm that has as an input the output weights of other b-tagging algorithms [41], [42]. The last part of the selection, also called tag selection, consists of requiring only one b-tagged jet identified with the MV1 tagging algorithm at an operating point corresponding to a 70% b-tagging efficiency. This is useful for separating top-antitop events, as discussed in Section 4.2. The transverse momentum of the selected b-jet has to be larger than 25 GeV.

	# of events
Wt	1660 ± 41
single top	1933 ± 44
W + light flavour	3719 ± 61
Z + jets, WW, ZZ, WZ	1430 ± 38
W + heavy flavour	11452 ± 110
$t\bar{t}$	15700 ± 130
Multijets	2300 ± 1100
TOTAL Exp.	38200 ± 1200
DATA	37503

Table 4.1: Event yields for the 3-jet bin, after the events selection.

Table 4.1 shows the number of selected events in the 3-jet bin for signal and all background channels, as well as for data. One can see from the event yields that the number of *Wt* events is very low compared to almost all other channels, even after the event selection. At the end of this selection procedure, there should be good agreement between data and Monte Carlo. This is checked by looking at control plots for each variable that is used in the analysis. Examples of such control plots are shown in the next chapter, when the first studies are presented.

4.5 Analysis strategy

This section describes the steps that are done in order to get from the selected data to actual physical results that one can interpret with respect to the theoretical predictions. Each search channel has an analysis path that is optimized to obtain results with the lowest possible uncertainties and highest significance. In the case of the $Wt \rightarrow$ lepton + jets, due to the low cross-section of this process and many sources of background, the analysis requires the use of more sophisticated analysis techniques and tools, that are described in the following part of this section.

Due to the similar features between the *Wt* processes and the background channels, and the very small signal fraction even after the event selection was performed, a "cut and count" analysis is not possible. Instead, in order to separate signal from background, multivariate analysis techniques have to be applied. In particular, for this thesis, an artificial neural network (NN) is used. This is part of the NeuroBayes (NB) package [43], that is described in detail in the next section.

The output of the neural network is one single discriminant that combines the separation power of all the variables that go into the training. The trained network is then used in order to calculate this discriminant for signal, all the background channels and the data.

4.5.1 Neuro Bayes package

The NeuroBayes neural network package is a tool based on Bayesian statistics, that performs multivariate analysis on correlated data. This package consists of a three-layer feed-forward neural network and a preprocessor that deals with the input variables. The three-layer structure refers to the three levels of the neural network, more precisely: the input, hidden and output layers. In a feed-forward neural network the information flows from the input nodes to the output node, through the nodes in the hidden layer. The number of nodes in the input layer is equal to the number of variables that are fed to the network plus one additional bias node. For the hidden layer, the number of nodes can be set by the user in order
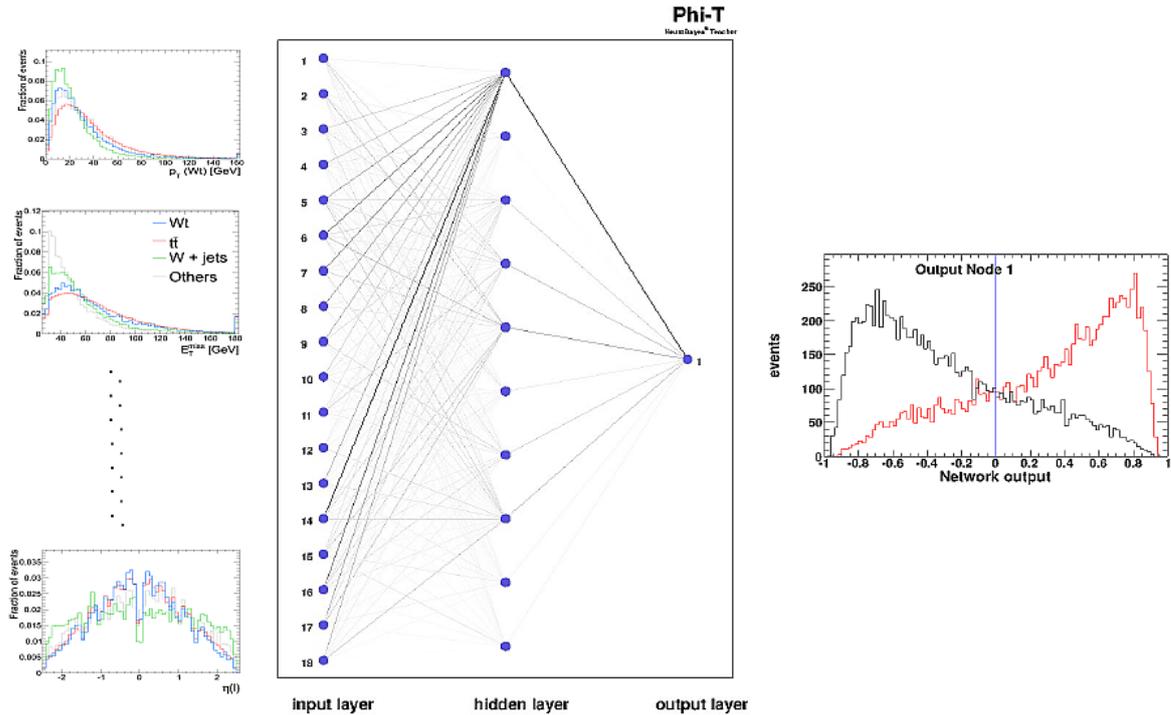
Figure 4.2: Overview of a three-layered neural network showing the input, hidden and output layers, as well as example input variable distributions and a distribution of the neural network discriminant.

to ensure an optimal performance of the neural network (NN). The output layer contains only one node, which corresponds to the NN discriminant mentioned before. All these elements are illustrated in figure 4.2. It is visible in that picture that, although the initial variables do not show significant separation between the Wt signal and background, the constructed discriminant does manage to separate fairly well the two event types. The output variable from the neural network ranges between -1 and 1, where -1 corresponds to background and 1 denotes signal-like events.

Preprocessing

The first step that is done when using the NN is preprocessing. All the variables that may show some separation between signal and background are fed into the neural network. The maximum number of input variables that can be handled by NeuroBayes is 300. After being preprocessed, these variables are ranked according to their significance and only the ones that show significance above a certain threshold (that can be set by the user) are used for the training.

The first step that occurs during preprocessing is a non-linear transformation over the interval $[-1, 1]$ of each variable and another transformation such that the distribution has a Gaussian shape. After that, the correlation matrix for all variables and the correlation to target of the whole set are calculated. The total significance of this correlation is also computed. The next step is to remove one variable and check what the loss of correlation to target is. The variable that shows least loss of information is then removed from the set and considered the least significant variable. This procedure is than repeated until all variables from the list have been classified.

For each variable four different quantities are computed:

- the **additional significance**, that is calculated with the procedure described above and is the criterion for the ranking of the variables;

- the **single significance**, that equals the total significance when only this variable is included;

- the **significance loss**, that is determined as the significance that is lost when this variable is excluded from the training;

- the **global correlation**, representing the total correlation of the variable with respect to all the others.

NeuroBayes training output

The neural network training is done using the most significant variables selected by the preprocessor. The output file of the NN training contains a lot of plots that are useful in order to evaluate how well the signal gets separated. The main figure of merit is the NeuroBayes output distribution. This type of plot, that is also shown on the right side of figure 4.2, is an overlay of the histograms for signal (in red) and background (in black) and shows a line at 0. A good separation of the network would show the red curve pushed to the right side of the plot, towards 1, and the black curve towards -1, and as few events as possible in-between.

Another output figure of the network is the plot of signal purity (defined as the ratio between signal and signal plus background), in bins of the NN output. If the points are located on the diagonal, one can conclude that NN is well calibrated in that training, confirming the interpretation of the NN output as a probability.

Other plots that are present in the output file are the purity vs. signal efficiency plot for different cuts on the NN discriminant or the signal efficiency vs. total efficiency when cutting on the same quantity. The latter one is shown in figure 4.3. From that plot one can determine a quantity called "Gini index", that is a measure of the quality of that training. For example, in the plot that is shown here, the Gini index is defined as the ratio between the blue area and the area below the diagonal. The diagonal line corresponds to a random sorting of the events. The other line delimiting the white area corresponds to a completely correct separation, in which, when applying the cuts on the NN discriminant, first all the background is cut away, resulting in a completely pure sample. The white area corresponds to the physical region and is limited by the choice of signal and background fractions in the sample. In our analysis we chose a 50:50 signal to background ratio. That translates into an upper limit for the Gini index of 50%.

The output file also contains the correlation matrix of all input variables and a separate section for each variable, in which plots such as the signal purity or the purity vs. efficiency of that variable are shown.

Neural network analysis in the Wt channel

For the $Wt \rightarrow$ lepton + jets analysis, even with the use of a neural network, the Wt signal cannot be separated from the two main sources of background (W + jets and $t\bar{t}$) just in one training. It is also visible in the distribution of individual variables that the signal distribution is, in most of the cases, located in-between the two main background contributions. Hence, two neural networks are separately trained to distinguish Wt from $t\bar{t}$ and Wt from W + jets events. 2D distributions are then constructed from the output of these neural networks and then used for signal extraction. Two-dimensional distributions for the signal and main background channels can be seen in figure 4.4.
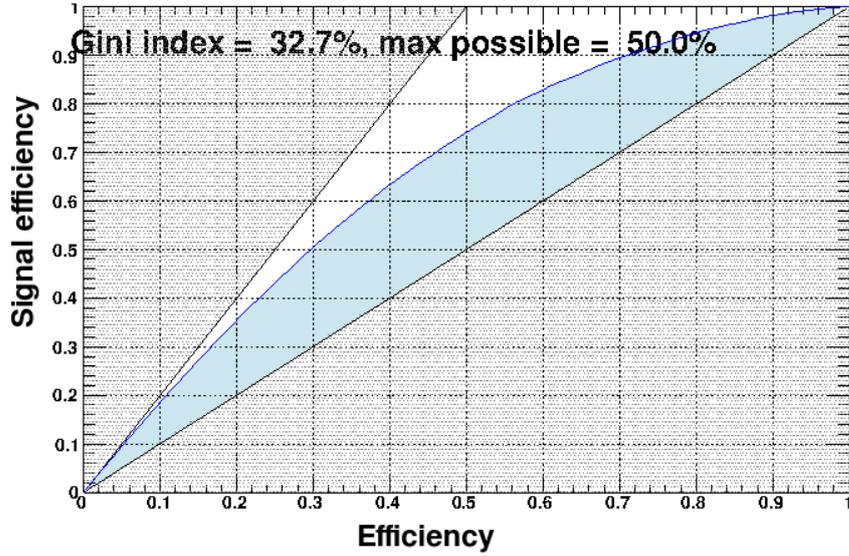
Figure 4.3: Signal efficiency vs. total efficiency plot for the NN output.

Extracting signal from the entire two-dimensional distribution is done by using a binned maximum likelihood fit. This method has the advantage that the two main backgrounds can be constrained independently because they dominate different parts of the phase space. Also, in comparison to an analysis in which a cut on the NN output distribution is applied, the 2D approach benefits from the usage of all events, improving the final significance.

For technical reasons the fit is actually performed on an one-dimensional representations of the 2D distributions (e.g. all the bins of the 2D histogram are arranged in a one dimensional one). The fitting package that was used is the Bill Fitter, which is a common tool used by other single top-quark research groups in ATLAS that is briefly described in the next section.

4.5.2 Bill Fitter

A binned maximum likelihood fit is used in this analysis in order to extract the *Wt* signal and calculate the significance of the measurement. The templates for the fit consist of the one dimensional mapping of the 2D distributions. In total there are 7 templates: 5 coming from the Monte Carlo samples (for *Wt*, $t\bar{t}$, single top-quark production in the *s*- and *t*-channel, *Z* + jets and diboson, and *W* + jets), one template for the multijet background and one for the combined electron and muon data. In order to perform the fits, a dedicate fitting package is used.

The Binned Log Likelihood (Bill) fitting package was developed by the University of Wuppertal for the signal extraction in the *t*-channel analysis [44]. A detailed description of the package can be found in [45].

The fitting procedure is based on a likelihood function that consists of Poisson likelihoods for each bin of the template histograms and a Gaussian term including the a-priori knowledge about the background rates. The likelihood can be written as:

$$L(\beta^s; \beta_j^b) = \prod_{k=1}^{M} \frac{e^{-\mu_k} \cdot \mu_k^{n_k}}{n_k!} \cdot \prod_{j=1}^{B} G(\beta_j^b; 1, \Delta_j)$$

with

$$\mu_k = \mu_k^s + \sum_{j=1}^{B} \mu_{jk}^b, \qquad \mu_k^s = \beta^s \cdot \tilde{\nu}_s \cdot \alpha_k^s \quad \text{and} \quad \mu_{jk}^b = \beta_j^b \cdot \tilde{\nu}_j \cdot \alpha_{jk}.$$

Here M denotes the maximum number of bins and B the background processes numbers. The indices s and b stand for signal and background. The predicted number of signal/background events is denoted with $\tilde{\nu}_s/\tilde{\nu}_j$. Observed and expected events numbers for channel k are denoted with n_k and μ_k respectively. α_k^s denotes the number of signal events falling in channel k while α_{jk} is the equivalent in background events.

For this function, the negative logarithm is evaluated using Minuit and from that the maximum likelihood is extracted. Using the results of this fitting procedure one can extract scale factors for the different channels. For signal, the Wt scale factor, β^s, relates the theoretical cross-section for this process to the observed value. Uncertainties of this results are evaluated using 20 000 Monte Carlo generated pseudo-experiments. The ensemble of pseudo-experiments includes all systematic uncertainties. In each of the pseudo-experiments the number of generated events for a certain process, n_j, is determined by throwing a random number acording to a Poisson distribution with mean μ_j. For the studies presented in this thesis no cross-section measurements were included, but systematic uncertainties and expected significance of the results were compared. More details about how the significance is determined are given in the next chapter, when the first results are presented. Asymmetric systematic uncertainties will shift the mode (and mean) of the distribution, indicating a bias. This is also calculated for each systematic uncertainty. However, since the values for this bias are typically very low, they will not be shown in the tables listing systematic uncertainties.
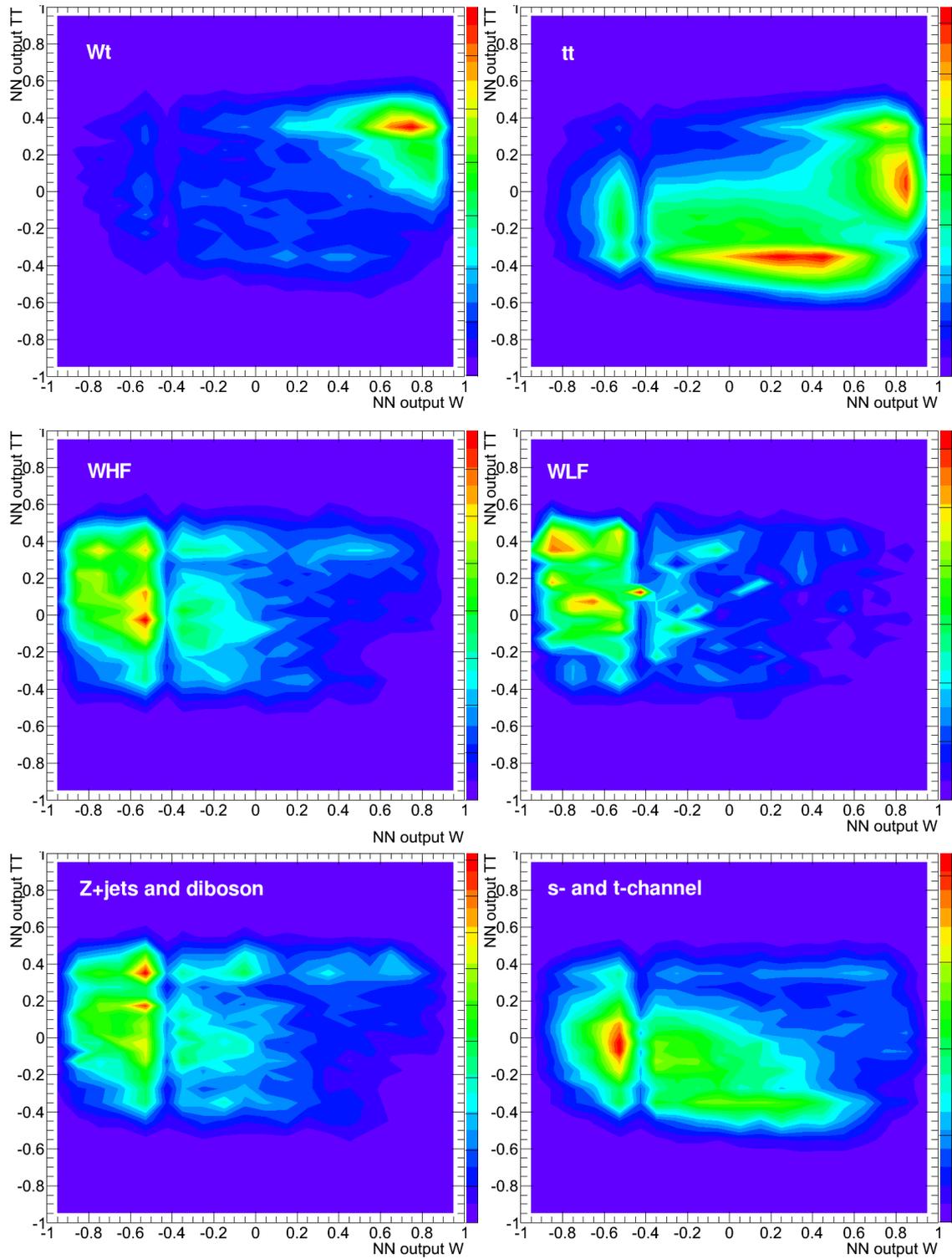
Figure 4.4: 2D distributions of the NN discriminants; On the *y*-axis there is the output of the neural network trained to distinguish *Wt* events from *tt̄*, while the *x*-axis represents the resulting discriminant of the NN trained against *W* + jets; only kinematic variables are used.

Kinematic Fitting

Kinematic fitting is a method in which the measured physical quantities are fitted using assumed known event topologies [46]. The basic principle of this method and a short description of the KLFitter package are given in the first section of the chapter. Sections 5.2 and 5.3 present the output variables of the package and how they can be used in the single top-quark analysis. The last section summarizes the results of a study that evaluates the effect of including information from kinematic fitting in the current analysis strategy.

5.1 KLFitter and its application in the Wt channel analysis

The principle of kinematic fitting is based on the association between the measured objects and a certain model that has a specific event topology. The fit parameters are the four momenta of the model particles and their initial values are the measured associated observables. The final values of the fit parameters are then "corrected" values of these quantities. The aim of performing such fits is to provide better estimates for some of the observables. Kinematic fitting also presents a way to test the fit hypothesis and correctly assign the measured objects to the particles that constitute the final state of the model.

5.1.1 The KLFitter package

Several software packages that use this method were developed. The one that is used in this analysis is called KLFitter [47] and was developed by the University of Göttingen. It was first constructed for the application of kinematic fitting in the analysis of top-quark pair production (lepton + jets channel) events. The package can be adapted for other analysis configurations as well, and its implementation for single top-quark Wt is detailed in [5].

The KLFitter uses a likelihood approach to kinematic fitting. In this approach the likelihood function gives the probability of measuring the observed values, given the configuration of the assumed event topology:

$$\mathcal{L}(\text{measurement}|\text{model}(\text{parameters})).$$

The likelihood function can be separated into two parts: one taking into account the kinematic constraints and the other one taking into account the measurement resolution of the observables that enter the fit.
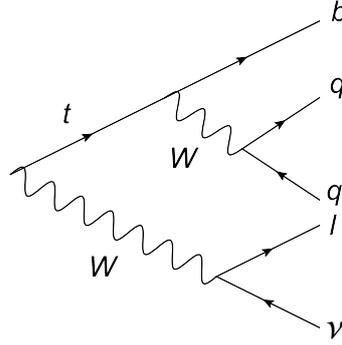
Figure 5.1: Decay chain of the *Wt*-channel in the lepton + jets decay mode.

The mass constraints are imposed through Breit-Wigner functions that parametrize the probability of a particle to have a certain mass which is equal to the invariant mass of the decay particles. Given the known value of the width Γ and pole mass M, this function can be written as:

$$BW(m|M,\Gamma) = \frac{1}{(m^2 - M^2)^2 + M^2\Gamma^2}.$$

The second component of the likelihood function is the product of the transfer functions (TFs) corresponding to the individual likelihood of measuring a value of one observable, taking into account the true value of the model parameter associated with it. For the main observables, the transfer functions are provided by the KLFitter package. This includes TFs for:

- the energy of quark and gluon jets, electrons and photons;

- missing transverse momentum ($x-$ and $y-$ component);

- the direction of light-quark and *b*-quark jets (in η and ϕ);

- the p_T of the muon.

These transfer functions are parametrized either by double-Gaussian functions (for jets and electron energy) or simple Gaussians (for $p_{x,y}^{miss}$, p_T^{μ}, photon energy and jet angles). Except for the missing transverse momentum TF, all the other ones are provided in bins of pseudorapidity.

5.1.2 Application of kinematic fitting in the *Wt* channel analysis

For the *Wt* analysis the model consists of the considered event topology of the *Wt* channel, namely the lepton + jets decay mode. Particles expected to be found in the final state (one lepton, one neutrino, two light quarks and one bottom quark) are associated to the observed quantities (a lepton, missing transverse energy, two light-quark jets and one *b*-tagged jet). The decay chain of the $Wt \longrightarrow bqql\nu$ can be seen in figure 5.1

The likelihood function of the *Wt* channel can be written as:

$$\mathcal{L} = \mathcal{L}_{\text{constraints}} \cdot \mathcal{L}_{\text{TF}}.$$

The component imposing the mass constraints, $\mathcal{L}_{\text{constraints}}$, will include conditions for all three decay vertices present in the model hypothesis (one for the top quark and one for each W boson) and can be

written as follows:

$$\mathcal{L}_{\text{constraints}} = BW(l\nu|M_W) \cdot BW(jj|M_W) \cdot BW(bW|M_{top}).$$

$\mathcal{L}_{\text{TF}}$ represents the product of the individual transfer functions of the *b*-quark and two light quarks energies, the lepton energy and the measured components of the neutrino momentum ($p_{\nu,x}$ and $p_{\nu,y}$). This can be written as:

$$\mathcal{L}_{\text{TF}} = \prod_{i \epsilon E_b, E_q, E_{q'}, E_l, p_{\nu,x}, p_{\nu,y}} TF(i_{\text{measured}}|i),$$

where *i* corresponds to the observables listed above.

Fit hypothesis and implemented fits

When fitting the *Wt* lepton + jets topology it is important to distinguish between two different fit hypotheses:

- **t**$W \longrightarrow$ **Wb**$W \longrightarrow$ **b**$l\nu qq$, in which the *W* boson coming from the top quark decays leptonically and

- **t**$W \longrightarrow$ **Wb**$W \longrightarrow$ **qqb**$l\nu$, when the prompt *W* decays leptonically and the boson coming from the $t \longrightarrow Wb$ vertex decays hadronically.

This differentiation has to be made because of the constraint imposed on the invariant mass of the decay products of the top quark by introducing the $BW(bW|M_{top})$ term in the likelihood calculation, as mentioned before.

The implemented fits that are relevant for this channel are the *Wt* system fit for both leptonic and hadronic hypothesis, an individual fit of the hadronic top quark and also a $t\bar{t}$ lepton + jets fit (only available in the 4-jet bin since the top-antitop final state contains at least two *b*-tagged jets and two light quark jets).

Permutations

Another important aspect that has to be taken into account when talking about kinematic fitting is the fact that the association of the physical objects included in the model to the observed quantities may not be unique. Since *b*-tagging information is not included in the actual *Wt* fitting procedure, ambiguities arise from the associations between the flavour[1] of the measured jets and the corresponding model partons.

In 3-jet events there are three possible options of "pairing" the model parameters and the observables. Since there is no difference considered between the two light quark jets and no *b*-tagging information is used at this step, the only thing that differentiates the 3 permutations is which jet is associated to the bottom quark from the model. With the same reasoning, we can deduce that in the 4-jet bin, the number of permutations will be equal to 12.

At this point information from *b*-tagging can be included in the KLFitter. This is done by applying a *b*-tagging weight, w_b, to the likelihood of each permutation:

$$\mathcal{L}_b = w_b \cdot \mathcal{L}.$$

The applied weight is calculated as the product of the individual *b*-tagging probabilities of each parton. These probabilities depend on the efficiency and rejection of the *b*-tagging working point used when measuring the jet flavours. In the case of a jet that is tagged as a *b*-jet and that originates from a bottom

---

[1] "flavour" here refers to whether a jet is *b*-tagged or not

quark this probability will be equal with the efficiency, ϵ, of the tagger while in the case in which a jet that comes from a light quark is b-tagged, the probability will be equal to $1/R$, where R is the rejection rate. For this analysis the MV1 tagger was chosen with an efficiency of 70%.

All permutations are fitted and then ranked according to their b-tag weighted likelihoods, $\mathcal{L}_b$.

5.2 KLFitter output variables

The KLFitter output variables can be categorized into two classes. The first category consists of so-called "likelihood variables". These include the likelihood of a fit and the b-tag weighted likelihood. Besides $\mathcal{L}$ and $\mathcal{L}_b$, the ratio between the b-tag weighted likelihood of one permutation and the sum of the weighted likelihoods for all permutations is computed. This is called the normalized b-likelihood. For convenience, the natural logarithm of the first two variables will be considered. Additionally, the individual components of the likelihood (transfer functions and Breit-Wigner functions) can be accessed.

The second type of variables that are used for the analysis are the kinematic variables of the fitted particles. These are the transverse momentum, energy, mass and position information (ϕ and η) corresponding to the individual particles (two light quarks, one bottom quark, one lepton and one neutrino), as well as for the reconstructed particles (W boson and top quark) and for the Wt system.

In these studies only kinematic variables corresponding to the best ranked fit permutation were used. The likelihood variables included correspond to the three permutations with the highest likelihoods. In the Wt fit, such variable sets are computed for both fit hypotheses, leptonically or hadronically decaying top, and for all permutations.

5.3 Kinematic fitting in the Wt analysis

The main issue of the Wt lepton + jets channel analysis is the difficulty to separate signal from background. In order to do that, as described in Section 4, neural networks are trained to differentiate between Wt events and the $t\bar{t}$ and W+jets events representing the main background sources. One idea for improving the separation is to include the output variables from the KLFitter that were described in the previous section in the list of initial variables used for the network training. This was done for events containing 3 and 4 jets.

5.3.1 3-jet bin analysis

The evaluation of whether including KLFitter information improves the results of the analysis was done by checking the separation power of the individual neural networks and the expected signal significance and main systematic uncertainties resulting after the signal extraction procedure and comparing the results obtained when adding the fitted variables to the ones from the standard analysis.

Neural network analysis in the 3-jet bin

The list of variables that are used in the default analysis consists of:

- variables coming from the 4-vectors of all particles (top quark, hadronic and leptonic W, b, l, ν_l and the two light-quark jets),

- reconstructed variables of the Wt system,

- angles between particles (in η and ϕ planes) as well as distances in the $\eta - \phi$ space,

- differences between the invariant masses of the W and t and the corresponding PDG values,

- event shape variables (combined kinematic variables for many or all particles in the final state), such as aplanarity, centrality or sphericity of the event.

These variables are all fed into the neural networks that are trained to separate between Wt and $t\bar{t}$ and Wt and W + jets respectively. The variables that are kept for the neural network trainings after the preprocessing step are listed in table 5.1. Control plots for the three the most significant variables in the Wt vs. $t\bar{t}$ and two of the Wt vs. W + jets variables, are shown in figure 5.2. These show all the background and the signal contributions stacked. The histograms are normalized with the individual cross-sections. The muon and electron channels are summed. The data points are shown in black and it is visible from these distributions that there is good agreement between data and Monte Carlo. In figure 5.3, the same variables are shown in overlay plots. These are area normalised distributions that only show the signal and the two main sources of background. The rest of the background channels are added into one distribution that is labelled as "Other".

Wt vs. $t\bar{t}$		Wt vs. W + jets	
Variable	Significance	Variable	Significance
$M(W^{had})$	34.4	$\|M(W^{had}) - M(W_{PDG})\|$	41.9
$p_T(Wt)/p_T(ljet2)$	21.7	$\|M(W^{had}) - M(t^{had}\|)$	21.5
$p_T(ljet2)$	18.5	E_T^{miss}	21.2
$p_T(l)$	12.7	$\|M(lJet2) - M(t^{lep})\|$	20.9
$p_T(ljet1)$	9.6	$\eta(l)$	19.3
$p_T(t^{had})$	7.6	$p_T(Wt)$	16.8
$M(Wt)/M(lJet1)$	6.8	$p_T(jet1)$	15.1
$\Delta\phi(\nu, bJet)$	6.9	$\Delta\eta(ljet1, bJet)$	12.4
$\|M(W^{had}) - M(t^{had})\|$	6.7	$\Delta\eta(l, bJet)$	9.9
$p_T(l)/p_T(ljet2)$	6.6	$p_T(ljet2)$	9.5
E_T^{miss}	5.3	$\|(m(W^{had}) - m_t^{PDG})\| + (\|m(t^{lep\backslash had}) - m_{PDG}^t\|)$	7.8
		$M(lJet1)/M(t^{had})$	6.9
		$\Delta\phi(l, W^{had})$	6.3
		$\Delta\phi(\nu, W^{had})$	7.8
		$\Delta R(ljet2, bJet)$	4.5
		$M(Wt)/M(bJet)$	6.5

Table 5.1: Most significant variables selected by the preprocessor in the 3-jet bin default analysis, ranked according to their additional significance. Only variables with significance higher than 5σ are selected for the training. "ljet1" and "ljet2" stand for the hardest and second-hardest light-quark jet and the superscripts *had* or *lep* refer to hadronic or leptonic decaying particles.

The output variables of the KLFitter package have been already described in the previous section. These have been added to the initial list of variables that goes into the preprocessing of the NN. The list of most significant variables for both trainings (Wt vs. $t\bar{t}$ and Wt vs. W + jets) is given in Table 5.2. Out of the 23 selected variables, 7 come from the KLFitter. Most of them are selected for separating the signal from W + jets background. Stack plots for the variables that come from kinematic fitting are shown in figure 5.4. The corresponding overlay plots can be seen in figure 5.5. KLFitter output variables also show good agreement between data and Monte Carlo.

The NN discriminants for both approaches are shown in figure 5.3. On the top row we can compare the separation between signal and $t\bar{t}$ background with and without KLFitter variables, while the bottom
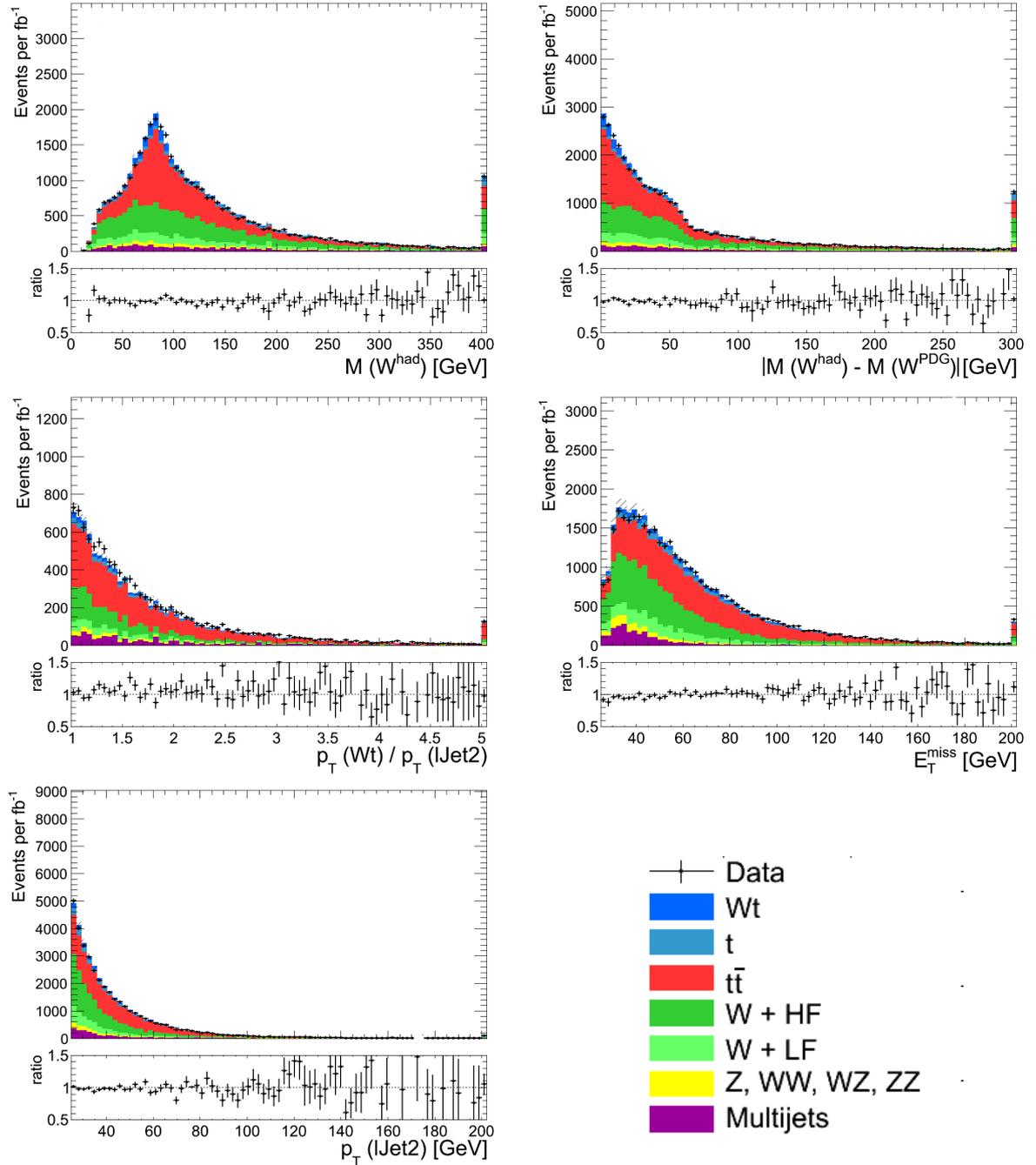
Figure 5.2: Control plots of the most significant kinematic variables that are selected in the preprocessing step of the NN training against $t\bar{t}$ and W+jets. The ratio between data and Monte Carlo is also shown for each plot.
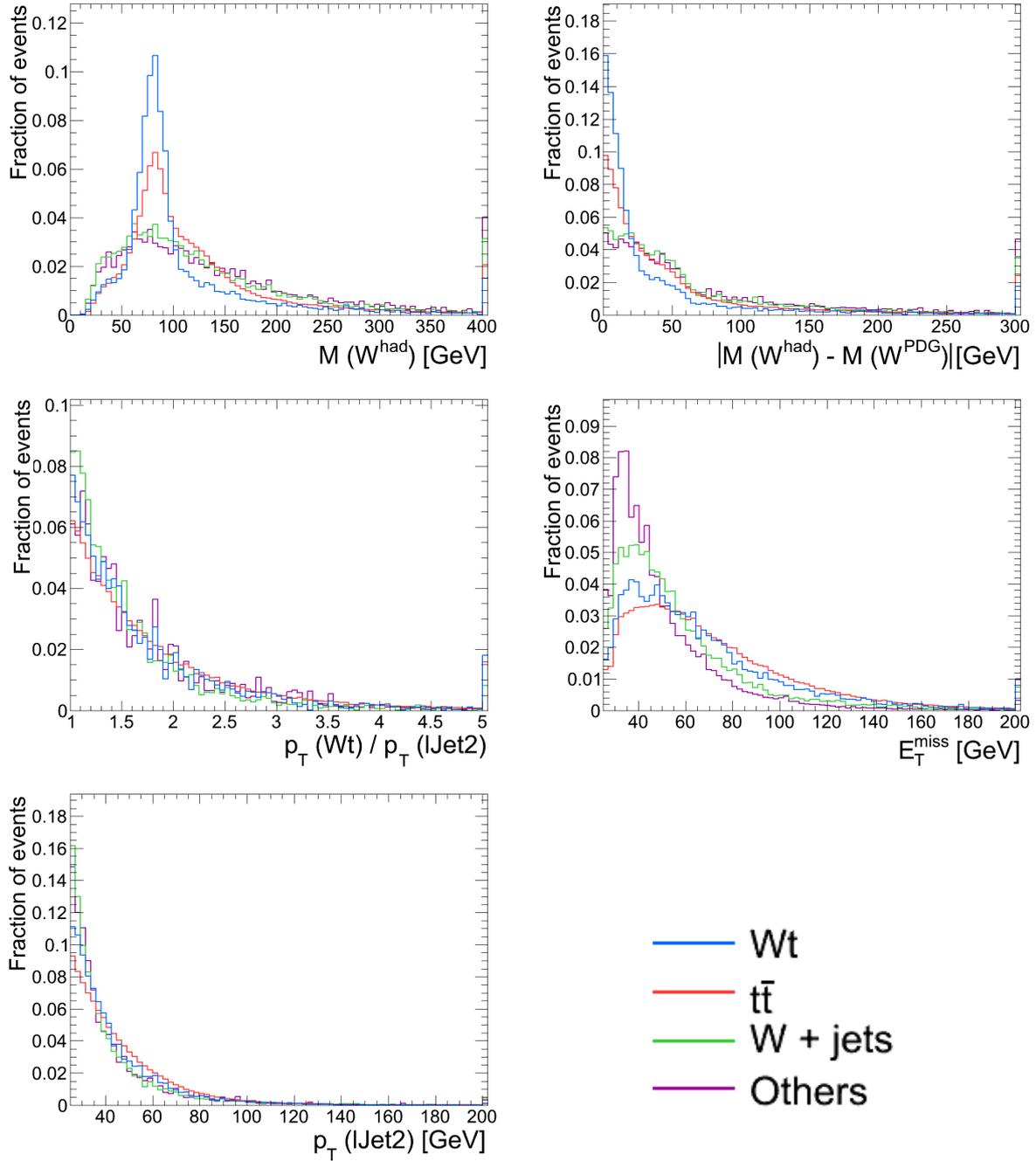
Figure 5.3: Overlay plots of the most significant kinematic variables that are selected in the preprocessing step. All distributions are normalised to the area.
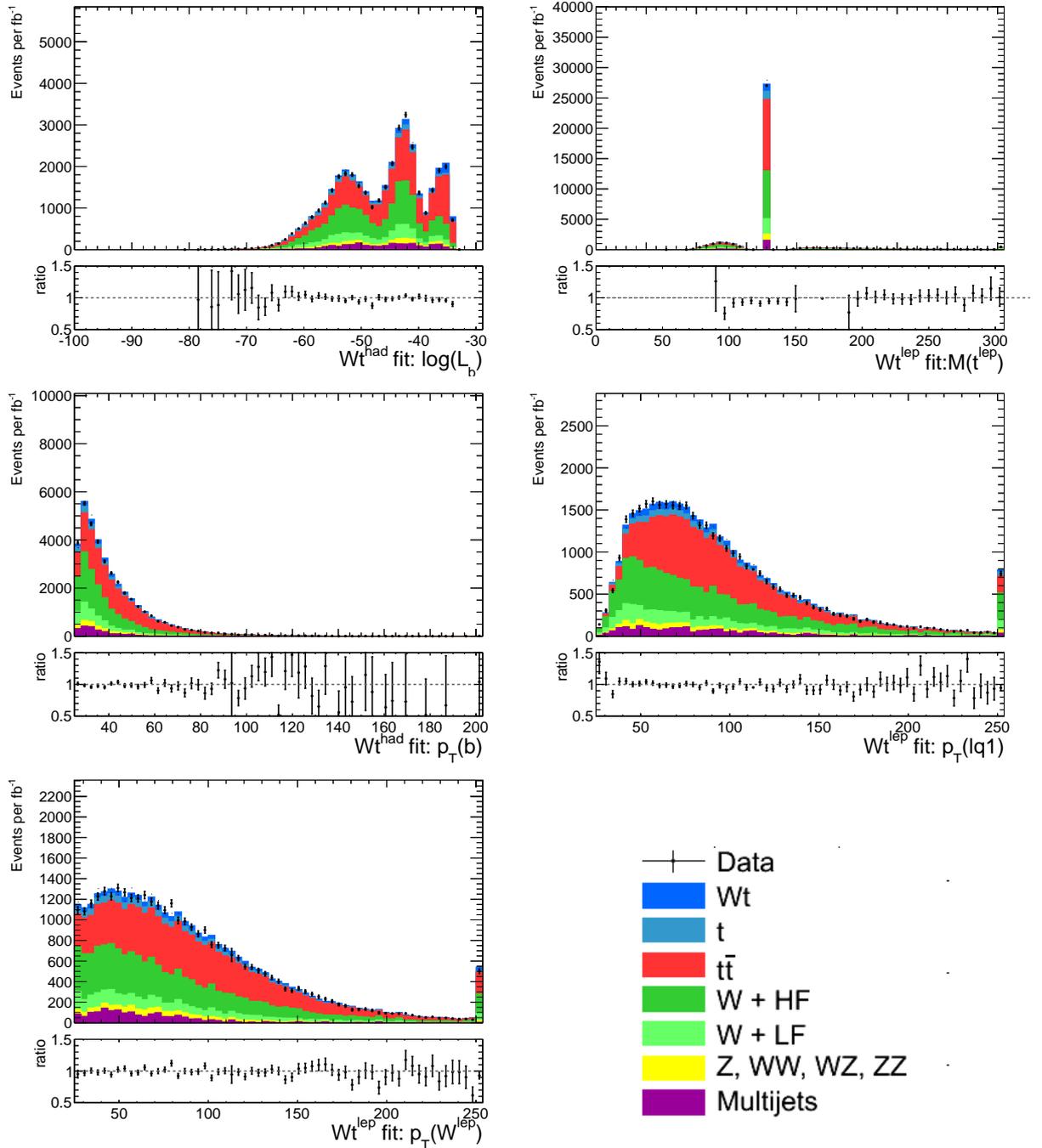
Figure 5.4: Control plots of the most significant KLFitter variables that are selected in the preprocessing step of the NN training against $t\bar{t}$ and W+jets. The ratio between data and Monte Carlo is also shown for each plot.
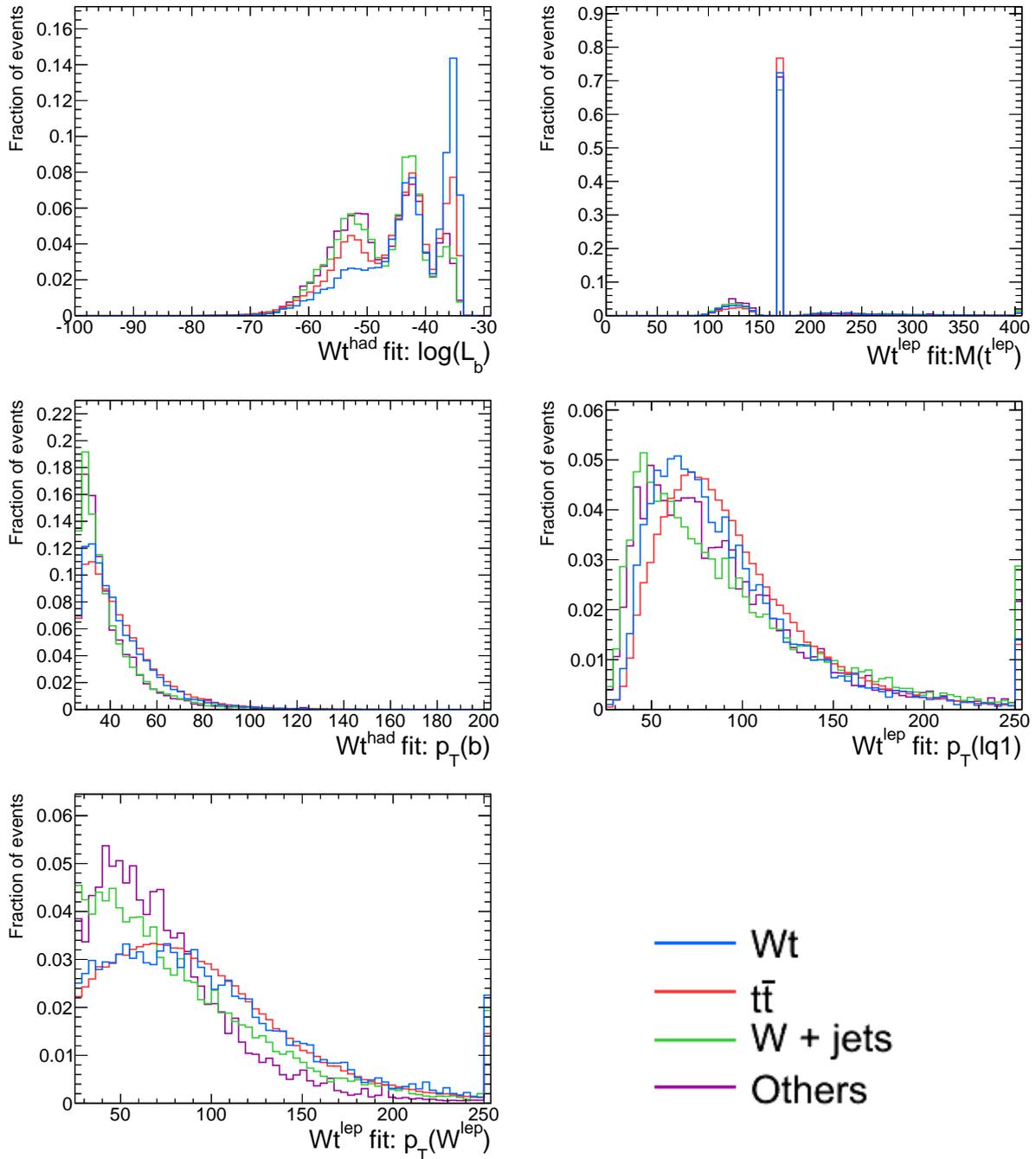
Figure 5.5: Overlay plots of the most significant KLFitter variables that are selected in the preprocessing step. All distributions are normalised to the area.

Wt vs. $t\bar{t}$		Wt vs. W + jets	
Variable	Significance	Variable	Significance
$M(W^{had})$	34.4	Wt^{had} fit: $log(\mathcal{L}_b)$	42.1
$p_T(Wt)/p_T(ljet2)$	21.1	E_T^{miss}	22.1
$p_T(ljet2)$	18.8	$\|M(t^{lep}) - M(t_{PDG})\|$	23.1
$p_T(l)$	12.2	$\eta(l)$	18.6
$M(lJet1)/M(t^{lep})$	11.7	$p_T(Wt)$	17.9
Wt^{lep} fit: $p_T(ljet1)$	9.7	$p_T(jet1)$	16.3
E_T^{miss}	7.8	$\|M(W^{had}) - M(W_{PDG})\|$	16.4
Wt^{lep} fit: $M(t)$	6.4	Wt^{had} fit: $p_T(b-quark)$	14.3
$\Delta R(ljet1, ljet2)$	5.4	$\Delta R(ljet1, bJet)$	11.4
Wt^{lep} fit: $E(b-quark)$	5.9	$M(bJet)/M(t^{lep})$	10.1
		Wt^{lep} fit: $p_T(W^{lep})$	7.2
		Wt^{lep} fit: $log(\mathcal{L})$	7.0
		Wt^{had} fit: $M(W^{lep})$	6.9

Table 5.2: Most significant variables selected by the preprocessor in the 3-jet bin analysis, ranked according to their additional significance; variables from the KLFitter are included, ranked according to their significance.

row corresponds to trainings against W + jets background. The Gini index for the $t\bar{t}$ trainings increases from 17.7% to 17.9% when adding the KLFitter variables. The same small change can be seen in the Gini indices of the Wt vs. W + jets network, that goes from 31.7% to 32.0%. Since these small changes in the separation power of the neural networks are not sufficient for evaluating the effect of including the KLFitter variables, signal extraction is performed for both cases and the systematic uncertainties and signal significance are calculated.

Signal extraction in the 3-jet bin

In order to perform the binned likelihood fit using the Bill fitting tool described in the previous chapter, we first need to construct templates from the 2D distributions of the NN discriminants from the $t\bar{t}$ and W + jets training. These are included in Appendix B. Figure 5.6 shows the overlay and stack plots of the templates of the combined (kinematic + KLFitter) analysis.

Systematic uncertainties

The evaluation of systematic uncertainties is done following the recommendations of the ATLAS top group.

The dominant systematic uncertainties are found to be:

- jet energy scale (JES)

- initial and final state radiation (IFSR)

- jet energy resolution (JER)

- b-tagging heavy scale factors (BTAGB)

- generator systematics (including parton shower systematics)

NN training using only kinematic variables   **NN training including output variables from the KLFitter**
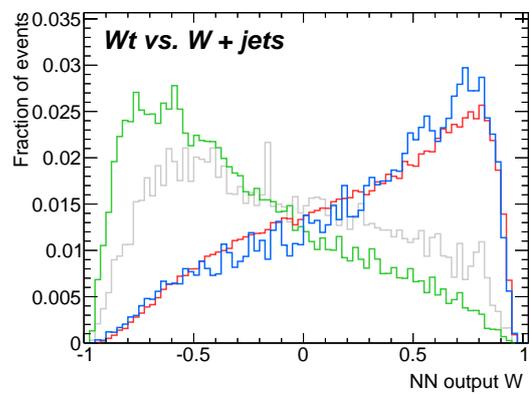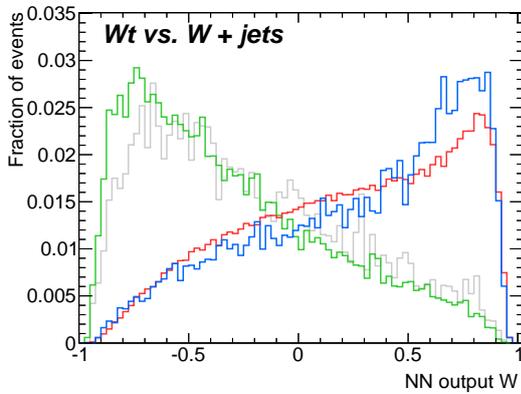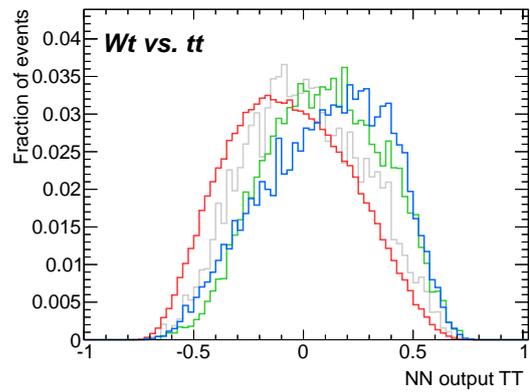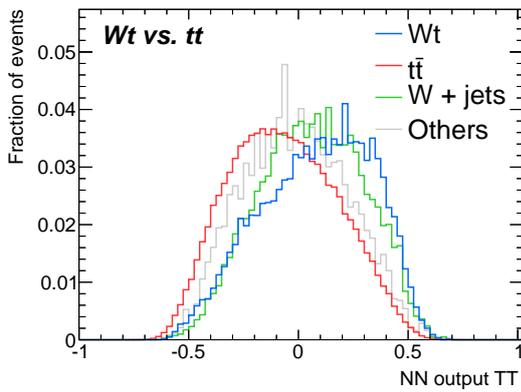
Table 5.3: Overlay plots of the neural network distributions, when the NN is trained against $t\bar{t}$ (upper row) and W + jets (bottom row), using only kinematic variables (left column) and including output variables from the KLFitter (right column).
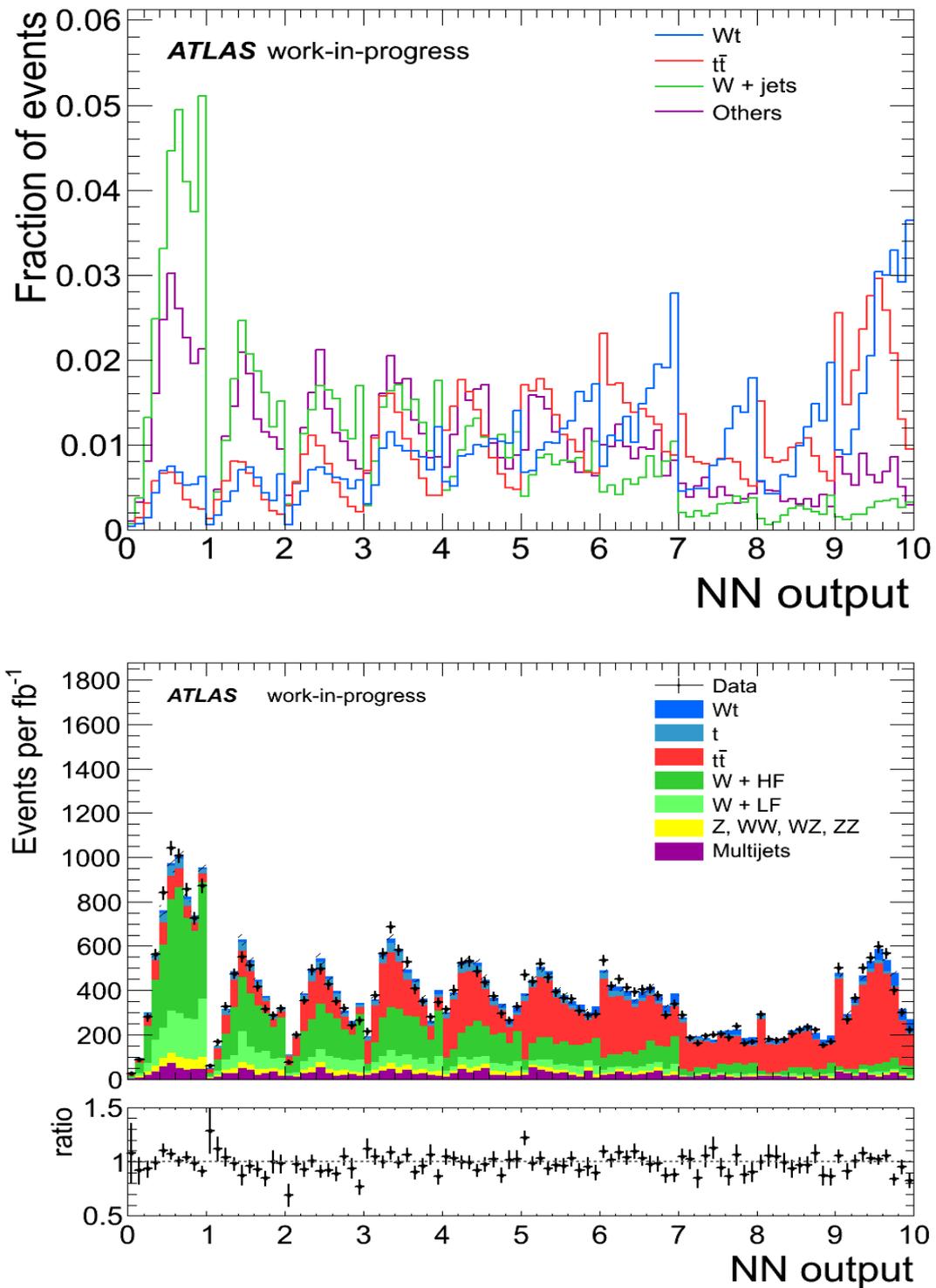
Figure 5.6: Unrolled 2D distributions of the NN discriminants when the neural network is trained to separate *Wt* signal from $t\bar{t}$ and from *W* + jets; KLFitter output variables are also included in the training. The top plot shows the normalized overlay of signal and background channels and the bottom distribution is the corresponding stack plot with the corresponding ratio between data and Monte Carlo.

In ATLAS effects coming from *b*-quark fragmentation, hadronisation and underlying soft radiation are studied using different Monte Carlo event generation models.

Initial and final state radiation of gluons are very common processes at hadron colliders. These produce additional jets in the event, besides the ones coming from hard scattering. The ISR/FSR uncertainty is, along with jet energy scale related systematics, the largest source of systematic uncertainties in the *Wt* analysis.

The JER systematic accounts for uncertainties in the reconstruction efficiency of jets while the jet energy scale uncertainty, as the name suggests, is connected to the jet energy scale calibration that is done in order to to correct the energy and momentum of the jets measured in the calorimeter to those of the jet at the hadronic scale. The jet energy scale uncertainty depends on p_T and η. The flavour composition of the jet and also the presence of close-by jets is taken into account. The jet energy scale and resolution corrections, are also propagated consistently to the missing transverse energy calculation. This is also done for leptons [48].

For all the jets in Monte Carlo events, a recalibration is done based on their flavour. The scale factors and corresponding uncertainties are provided by the ATLAS flavour tagging group. Variations of these scale factors are applied separately for the heavy and light flavour and then evaluated. The heavy flavour scale factors systematic uncertainties is also among the highest ones.

More systematic uncertainties are taken into account in this analysis:

- electron/muon energy resolution (EER, MER)

- electron/muon energy scale (EES, MES)

- lepton identification and trigger efficiency (ESF, MSF)

- jet vertex fraction (JVF)

- jet reconstruction efficiency (JEF)

- *b*-tag light flavour scale factors (BTAGC, BTAGL)

- MC generator

- *Wt* diagram removal/subtraction (DSDR)

- *W*+jets shape and normalisation.

However, compared to the dominant ones, these have a small effect. In the systematic uncertainties tables that are included in the thesis, the computed values for all uncertainties are shown.

Significance

The expected significance is calculated using hypothesis testing. *Q*-values are estimated using 50 000 pseudo-data events for the signal + background hypothesis (signal = *Wt* as predicted by the Standard Model) and background-only hypothesis. The definition for the *Q*-value is:

$$Q = -2(\ln L_L(S + B) - \ln L_L(B)),$$

where $L_L(S + B)$ is the value of the likelihood function at $\beta^s = 1$ (signal according to *Wt* Standard Model predictions) and $L_L(B)$ is the likelihood corresponding to $\beta^s = 0$ (no *Wt* signal).

Sys Name	up[%]	down[%]
Data stat.	16	−16
MSF	2	−1
ESF	2	−1
WJetsShapePtjmin10	2	−2
BTAGL	2	−2
DSDR	2	−2
JVFSF	2	−2
xsection	2	−2
BTAGC	4	−3
WJetsShapeIqopt3	4	−4
mcstat	10	−10
PS	12	−12
JER	29	−29
BTAGB	39	−39
JES	40	−40
IFSR	65	−65
Total(sys)	94	−93
Total(sys+stat)	95	−95

Table 5.4: Systematic uncertainties for the default analysis. The definitions for the abbreviations can be found in 5.3.1.

Sys Name	up[%]	down[%]
Data stat.	16	−16
MSF	3	−1
DSDR	2	−2
BTAGL	4	−2
JVFSF	2	−2
WJetsShapePtjmin10	2	−2
WJetsShapeIqopt3	3	−3
ESF	4	−3
BTAGC	6	−5
PS	10	−10
mcstat	11	−11
JES	27	−28
JER	29	−29
BTAGB	40	−39
IFSR	66	−66
Total(sys)	90	−89
Total(sys+stat)	92	−90

Table 5.5: Systematic uncertainties for the analysis performed including KLFitter variables.

The Q-value distributions are constructed and showed in figure 5.7. The results for the p-value and significance calculations for the analysis approach that includes KLFitter variables are summarised in table 5.6. In the default approach, the expected significance is about 1.2 standard deviations.

By looking at these results one can conclude that the effect of only adding variables from kinematic fitting does not make a significant difference. However there are other things that can be tried, such as, using the newly implemented $t\bar{t}$ kinematic fit for the 4-jet bin and the effect of adding events with 4 jets to the analysis.

Q-value	p-value	Significance [σ]
-51.9	0.091	1.3

Table 5.6: Expected Q-value, p-value and significance for the 3-jet bin measurement, when KLFitter variables are included in the NN training.

5.3.2 Including the 4-jet bin

For the 4-jet bin, separation between Wt and $t\bar{t}$ is more difficult than in the 3-jet bin. Up to now, including the 4-jet bin did not improve the analysis results. With the new KLFitter $t\bar{t}$ fit, it is worth checking what the effect of including 4-jet events in the selected sample that is used for signal extraction. This could be beneficial also because it will increase the total number of events, and hence decrease statistical uncertainties.

In order to achieve better separation, variables from a dedicated $t\bar{t}$ kinematic fit are added to the
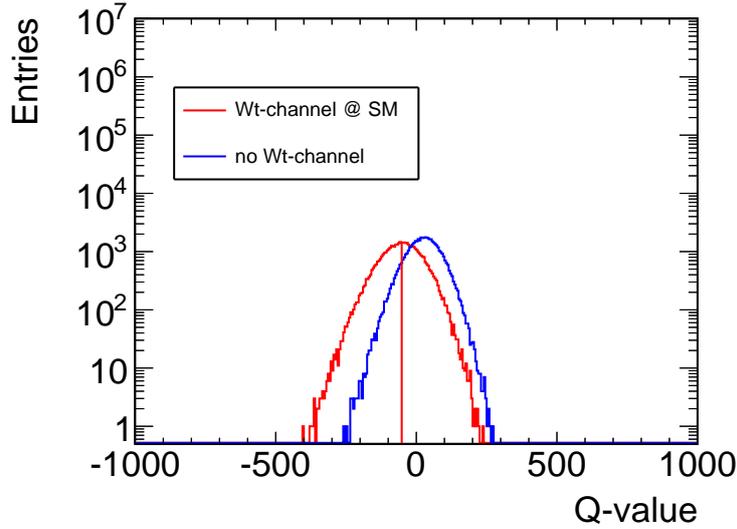
Figure 5.7: *Q*-value distributions derived from pseudo experiments, for the signal hypothesis (in red) and background only hypothesis (in blue).

complete list of variables that is fed to the preprocessor. In total, the KLFitter output for this fit gives 79 new variables. After the preprocessing step, 11 variables with significance higher than 3σ are selected for the training. A list of these variables, together with the corresponding significance is given in Table 5.7. Three of the selected variables come from kinematic fitting, out of which the most significant one is the likelihood of the $t\bar{t}$ fit. This has a significance of 13.6σ.

Variable	Significance [σ]
$t\bar{t}$ fit: $log(\mathcal{L})$	13.6
$\Delta M(bJet, W^{had})$	10.8
$p_T(l)$	8.7
$p_T(Wt)$	9.6
$p_T(bJet)$	6.9
$t\bar{t}$ fit (2nd perm.): norm.$\mathcal{L}_b$	6.0
$\Delta R(l, bJet)$	6.3
$m_T(W^{had})$	4.6
Wt^{had} fit: $m_{inv}(l\nu)$	4.6
$p_T(W^{had})$	4.3
$p_T(Wt)/p_T(lJet1)$	3.4

Table 5.7: Most significant variables selected by the preprocessor in the 4-jet bin analysis, ranked according to their significance; variables from the KLFitter are included.

The discrimination of the neural network for the kinematic/ kinematic + KLFitter approaches in the 4-jet bin can be seen in figure 5.8a / 5.8b. The signal distribution, represented by the blue line, is slightly pushed toward the right side of the plot in the second plot. The difference between the corresponding Gini indices is 2.5%, kinematic variables bringing a little more separation than just the default analysis. By performing the same steps as presented in the 3-jet bin dedicated section, one constructs a two-dimensional distribution of the NN discriminants, that is then unrolled and used as a template for signal extraction.
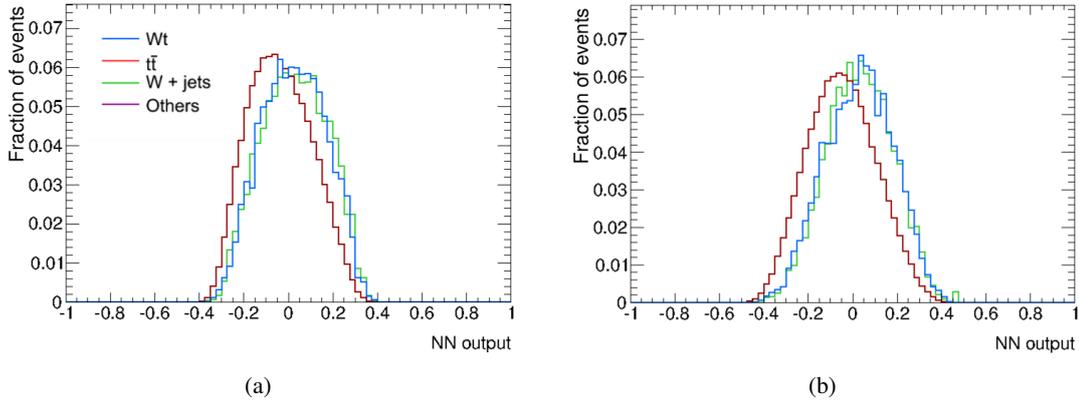
Figure 5.8: NN output overlay plots for Wt vs. $t\bar{t}$ training (a) without and (b) with including variables from the $t\bar{t}$ kinematic fit.

The binned maximum likelihood fit is not performed in the 4-jet bin only, but instead, signal is extracted from the combination of the 3- and 4-jet bins. The systematic uncertainties of the combination are listed in Table 5.8. One can see that, as expected, the statistical error on data is lower for the combination of the two channels. Also, the overall statistical + systematic uncertainty is lower by roughly 5%.

When calculating the expected signal significance, the result is comparable to the one obtained from the 3-jet bin only. There are no previous studies that combine the 3- and 4-jet bin when extracting the signal so that one could compare results from default analysis to the ones presented above, in which KLFitter information is used. However, combining the two jet bins brings a 5% improvement in terms of uncertainties and a signal significance of 1.3, which is identical to the result obtained from the 3-jet bin only.

Sys Name	up[%]	down[%]
Data stat.	14	−14
WJetsShapeIqopt3	1	−1
xsection	1	−1
WJetsShapePtjmin10	2	−2
DSDR	2	−2
MSF	5	−3
BTAGL	5	−3
ESF	6	−4
BTAGC	6	−4
JVFSF	7	−5
mcstat	10	−10
PS	10	−10
JER	25	−25
JES	40	−40
IFSR	42	−42
BTAGB	48	−47
Total(sys)	85	−82
Total(sys+stat)	86	−83

Table 5.8: Full systematic uncertainties table of the 3- and 4-jet bin combination.

Separating samples with a cleaner event topology

Besides improving reconstruction, in order to decrease systematic uncertainties and increase signal significance, it is important to also improve the selection. By only selecting events that have a cleaner event topology, such as those in which the top quark decays hadronically (which will be called "hadronic top-quark events" in the following) or events in which at least two light quarks are matched to the ones coming from the $W \to qq$ decay (or "matched events"), one can limit the combinatorial background and also increase the signal fraction in the sample. Matching refers to the correct association between measured jets and the true quarks in the final state of the $Wt \to$ lepton+jets decay. The first three sections of this chapter present the methods used for selecting this type of events and also a subset that combines these two features. Results are presented in Section 6.4.

6.1 Separating events that have a hadronic top-quark

An artificial neural network is used for separating events that have a hadronic top-quark. The training sample consists of Wt Monte Carlo generated events. Those events in which a hadronic top can be reconstructed (based on truth information) make up the signal sample. The rest of the events are used as background sample. In the first step 300 variables are considered as input for the neural network, including the output variables of the KLFitter, as described in Chapter 5. According to the preprocessor, 15 variables have a significance above 5σ. These are selected for the training. They are listed, together with their significance, in table 6.1. Control plots for the selected variables can be seen in Appendix A. 7 out of the 15 selected variables come from the KLFitter. The likelihood of the fit for the two fitted hypotheses, corresponding to the hadronic and leptonic decay of the top-quark, are the most significant variables.

The distribution of the neural network discriminant (NN_{had}) is shown in figure 6.1. This shows a fairly good separation between signal and background. The corresponding Gini plot can be seen in figure 6.2. The Gini index for this training is about 32%. The training is then applied to all samples in order to classify the events from all other channels into events that are similar to the ones that have a hadronic top-quark and events that do not contain a hadronic top-quark in their event topology. This is done based on the value of the neural network output. If $NN_{had} > 0$ we classify the event as having a hadronic top quark and if $NN_{had} < 0$ we consider that the event did not contain any hadronically

Variable	Significance [σ]
Wt^{had} fit: $\log(\mathcal{L}_b)$	46.9
Wt^{lep} fit: $\log(\mathcal{L}_b)$	48.2
$\lvert M(t^{lep})\text{-}M(t_{PDG})\rvert$	17.6
Wt^{lep} fit: BW(W^{lep})	12.8
Wt^{lep} fit: TF(b-quark)	16.2
E_T^{miss}	13.6
Wt^{had} fit: p_T(b-quark)	9.7
Wt^{had} fit: norm. $\mathcal{L}_b$	8.6
$M(t^{lep})$	7.4
$\Delta R(\nu, bJet)$	7.4
$E(t^{had})$	6.4
$M(t^{had})$	5.1
$\Delta M(W^{had}, t^{lep})$	6.2
Wt^{had} fit: p_T(W^{lep})	6.0
$\Delta R(ljet2, bJet)$	5.4

Table 6.1: Most significant variables selected by the preprocessor when a NN is trained for separating events that have a hadronically decaying top-quark, ranked according to their significance; variables from the KLFitter are included.

decaying top quark. This separation corresponds to a cut on the network discriminant along the blue line shown in figure 6.1.

In order to get an idea of the number of events that we consider as hadronic top-quark events, one can look at the table 6.2. This contains the number of events for each category, in each of the background samples. As expected, events that have $NN_{had} > 0$ will have a higher fraction of Wt events compared to the signal fraction in the full data set. This subset of events will then be used for signal extraction. Systematic uncertainties and expected significance in that region will be compared to the ones obtained when using the full dataset or other selected subsets, such as the one containing matched events, that will be described in the next section.

	Number of events	
Channel	$NN_{had} > 0$	$NN_{had} < 0$
Wt	950	860
W + light flavour	1600	2100
W + heavy flavour	5100	6400
Z + jets, WW, ZZ, WZ	710	720
Single top	430	1500
Multijets	950	1330
$t\bar{t}$	6300	9400

Table 6.2: Event yields for the 3-jet bin, after the event selection. The column in the middle shows the number of events with a neural network discriminant > 0 and the one on the right gives the number of events that have been classified as not having a hadronic top-quark.
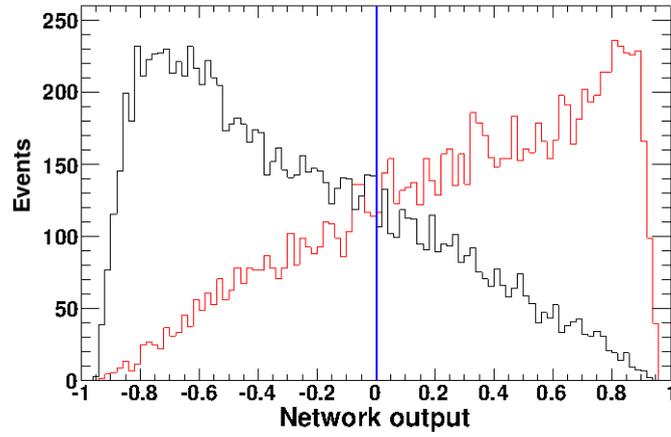
Figure 6.1: Neural network output distribution; events that have a hadronic top quark are considered as signal and the remaining ones are considered as background.
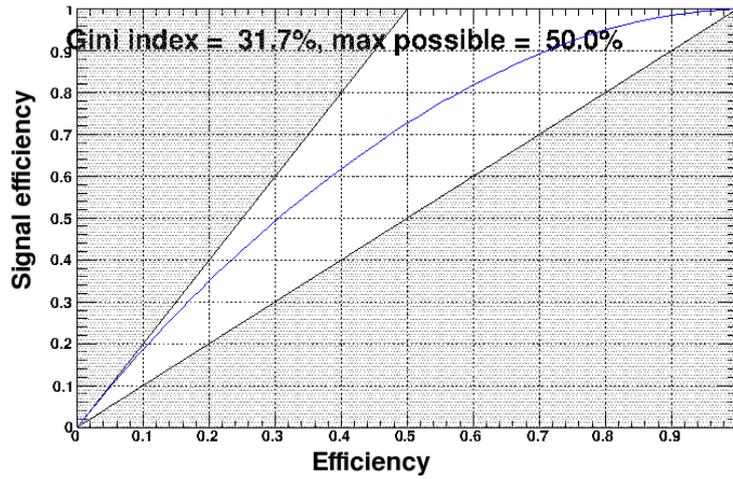
Figure 6.2: Gini plot for the neural network that was trained to distinguish events that have a hadronic top-quark from the ones that do not.

6.2 Separating matched events

The strategy for separating events that have at least two light-quarks matched is similar to the one used for separating events that have a hadronic top-quark. Quark-jet matching refers to the correct association between measured jets and the true quarks in the final state of the $Wt \rightarrow$ lepton+jets decay. The matching is done by a simple ΔR mapping of the light-quark jets to the true particles from Monte Carlo information. More details on this procedure can be found in [5]. For this analysis we chose to select events that have at least two matched light-quarks since there are indications that reconstructing the hadronic W boson would already give information about the full topology.

For separating matched events, another neural network is trained. Truth information is used for defining the signal sample. This is represented by Wt Monte Carlo events that are matched, while the considered background sample consists of the rest of the Wt events. The variables that are chosen after preprocessing and that show a significance above 5σ are listed in table 6.3. Out of the 17 selected variables, 8 are KLFitter output variables. The separation between signal and background is very good in this training. The neural network output (NN_{match}) is shown in figure 6.3. Again a blue line is drawn at the 0 value of the discriminant. This represents the cut that will be used for all samples in order to split them into matched and not-matched events. The Gini plot, included in figure 6.4, shows a value of 37.8% out of a maximum of 50% for the Gini index.

Variable	Significance [σ]
$\|(M(W^{had})\text{-}M_W^{PDG})\| + (\|M(t^{lep \backslash had})\text{-}M_{PDG}^t\|)$	76.4
Wt^{lep} fit: $\log(\mathcal{L}_b)$	23.1
$\|M(t^{lep})\text{-}M(t_{PDG})\|$	27.9
Wt^{had} fit: $p_T(light\text{-}quark2)$	19.4
t^{had} fit (2nd perm.): norm. $\mathcal{L}_b$	11.1
Wt^{had} fit (2nd perm.): $\log(\mathcal{L})$	13.1
Wt^{had} fit: BW(W^{had})	11.4
$p_T(W^{had})$	12.7
$\Delta R(ljet2, ljet1)$	8.4
$\|(M(W^{had})\text{-}M_W^{PDG})\|$	9.2
t^{had} fit: $\Delta\phi(t, W)$	6.0
$\Delta R(l, \nu)$	8.7
Wt^{lep} fit: E(l)	6.6
$\Delta R(ljet2, bJet)$	6.3
$\Delta R(Wt, l)$	6.2
Wt^{lep} fit: TF(light-quark2)	5.7
$p_T(bJet)$	6.1
$\|M(t^{had})\text{-}M(t_{PDG})\|$	6.3

Table 6.3: Most significant variables selected by the preprocessor when a NN is trained for separating events that have at least two matched light-quarks, ranked according to their significance; variables from the KLFitter are included.

The number of events that look like matched events in the other background samples, as well as the number of events that are not matched, are listed in table 6.4. The Wt channel is the only one for which the number of events is higher when $NN_{match} > 0$. The other background contributions are greatly reduced in that region, the signal/background ratio going from 0.04 when looking at the whole phase
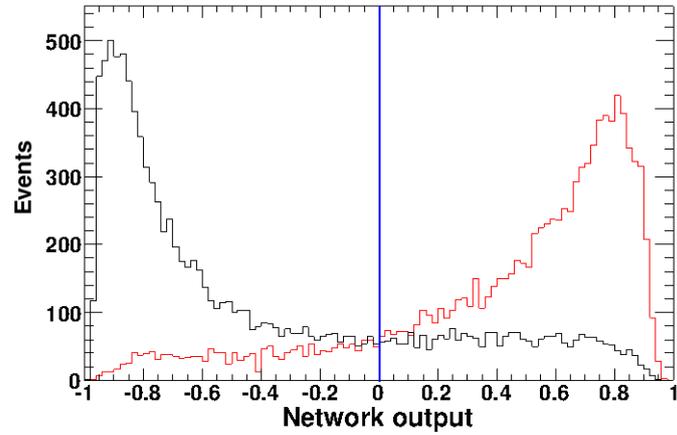
Figure 6.3: Neural network output distribution; events that have at least two matched light-quarks are considered as signal and events where the two light-quarks are not matched represent background.
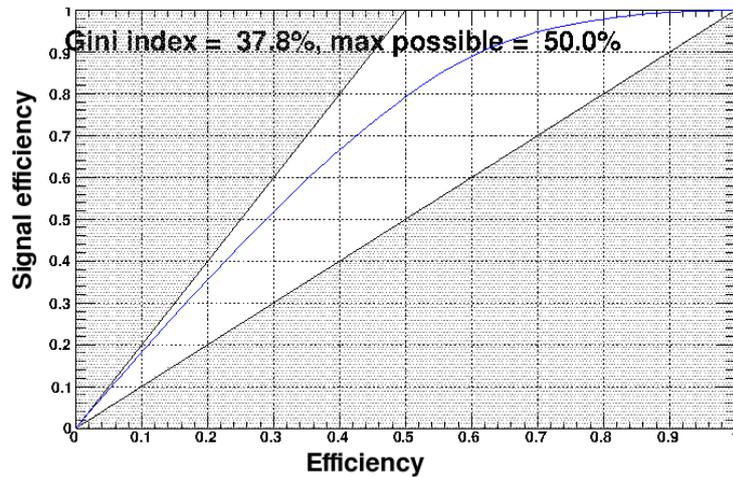
Figure 6.4: Gini plot for the neural network that was trained to distinguish events that have at least two matched light-quarks.

space to 0.09 when only matched events are selected. It is therefore beneficial to try to extract signal from this set of events that have a much cleaner event topology.

	Number of events	
Channel	$NN_{match} > 0$	$NN_{match} < 0$
Wt	860	800
W + light flavour	690	3000
W + heavy flavour	2300	9200
Z + jets, WW, ZZ, WZ	310	1100
Single top	330	1600
Multijets	350	1900
$t\bar{t}$	5900	9800

Table 6.4: Event yields for the 3-jet bin, after the events selection. The column in the middle shows the number of events with a neural discriminant > 0 and the one on the right gives the number of events that have been classified as not having at least 2 light-quarks matched.

6.3 Separating matched events that have a hadronic top-quark

In order to better visualize where the signal is accumulates in terms of matched and hadronically decaying top-quark events, 2D distributions of the network discriminants were constructed for each background channel and for *Wt*. These can be seen in figure 6.5. NN_{match} is shown on the *y*-axis, while NN_{had} is represented on the *x*-axis. In all the channels the majority of the events are concentrated towards the lower-left side of the distributions (i.e. they are not matched and do not have a hadronic top). The only processes that show more events in the region close to the +1 value of the NN selecting matched events are *Wt* and $t\bar{t}$. In particular, the top-right corner of the plot (where events are matched and have a hadronic top) seems to have an even higher *Wt* fraction. This is another interesting region that was selected and that was used for signal extraction.

Selecting events that are matched and have a hadronic top was done in two different ways. Either a neural network is trained to separate them from the rest, or one can apply cuts on the 2D distributions that were shown before. Both options were tested in this study and showed similar results. Because of that, only results obtained using the latter method will be presented in this thesis. Also, the cut values on the neural network outputs were optimized as to find the best configuration (e.g. minimal systematic uncertainties and maximal signal significance).

6.4 Signal extraction in all selected regions

For all the selected regions, as well as for the baseline analysis (when no additional sorting of the events is done), total uncertainties, as well as the signal significance were evaluated and compared. The four considered regions are:

- default analysis (no additional selection),

- events containing a hadronic top-quark ($NN_{had} > 0$),

- events that have at least two matched light-quarks ($NN_{match} > 0$),
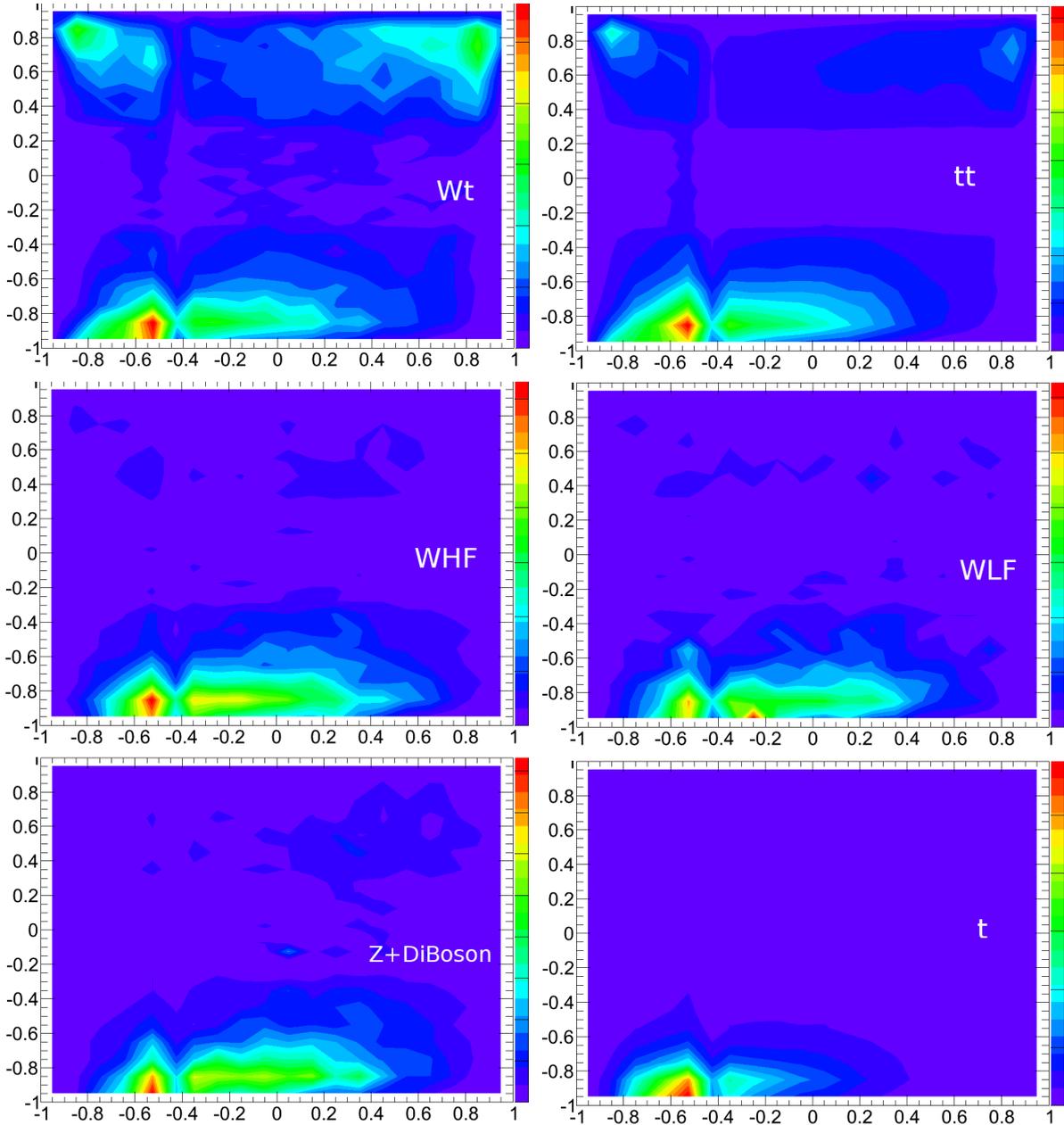
Figure 6.5: 2D distributions of the NN discriminants; On the y-axis there is the output of the neural network trained to distinguish events that have a hadronic top-quark from the ones that don't, while on the x-axis we have the resulting discriminant of the NN that identifies events in which at least two light-quark jets are matched.

- events that have a hadronic top-quark and at least two matched light-quarks ($NN_{had} > 0$ and $NN_{match} > 0.35$).
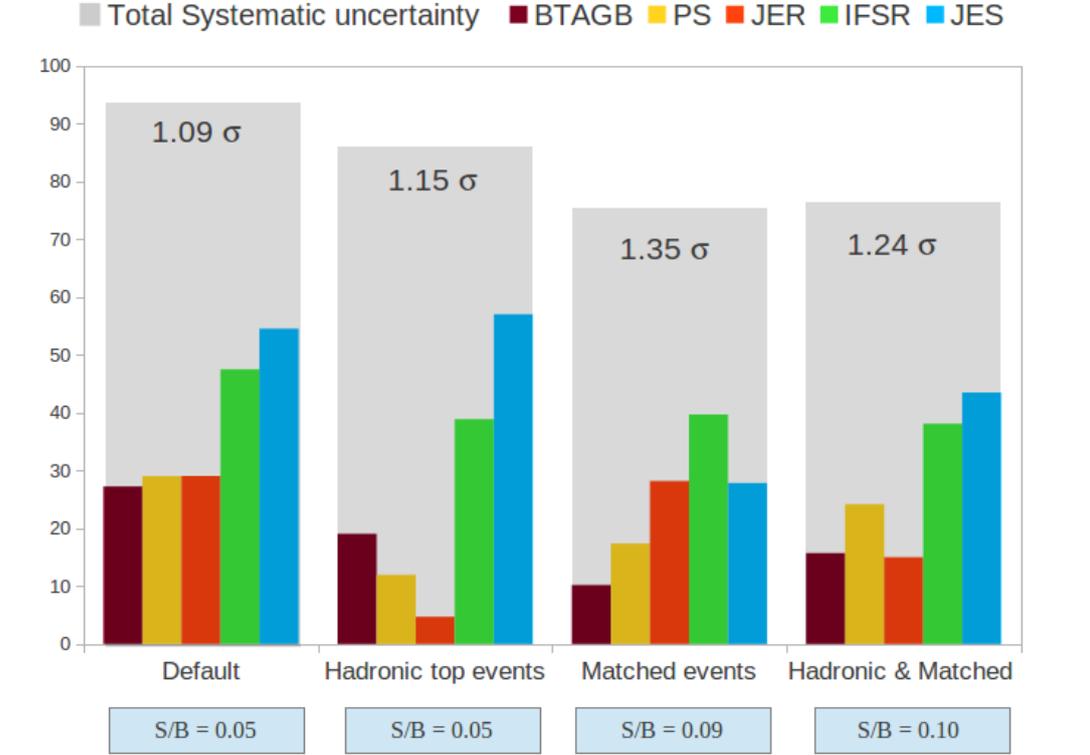
Figure 6.6: Systematic uncertainties and significance comparison for different selection regions.

The results of this comparison are shown in figure 6.6. The chart includes the main systematic un-certainties (jet energy scale and resolution, initial and final state radiation, MC generator and *b*-tagging scale factors) as well as the total systematic + statistical uncertainty (grey area) for each of the above mentioned regions. The signal significance is also computed and appears at the top of each column. At the bottom of the chart, for each region, the signal over background ratio is written. The full systematic tables for each subset of events are included in Appendix C.

It is visible from this chart how the systematics vary when selecting events with different topologies. In particular, the jet energy scale systematic uncertainty, that is dominating in the default approach, is reduced by more than 20% when only selecting matched events. The second highest systematic source, IFSR, is not strongly reduced by any of the selections and still goes up to almost 40% even in the best configuration.

Selecting only events that have a hadronic top-quark reduces the jet energy resolution uncertainty but only improves very little on the signal significance. Overall, events for which the light-quarks are matched give the results with the highest signal significance and lowest systematic uncertainties. All systematics are reduced compared to the default analysis. Selecting from this sample the events that also contain a hadronic top (results represented in the rightmost section of the chart) shows comparable total systematic uncertainties but a lower signal significance.

Summary

This thesis presents a contribution to the search for single top-quark associated production in the lepton+jets decay channel. The data set that was used consists of proton-proton collision data recorded by the AT-LAS detector at a centre-of-mass energy of 7 TeV, corresponding to an integrated luminosity of 4.7 fb^{-1}.

Although the lepton+jets decay mode accounts for 43.5% of all Wt events, there is no signal evidence found so far in this search channel. The final state of such a process contains a lepton, one neutrino, two light-quark jets and one b-quark jet. This is very similar to the signatures of other processes, such as $t\bar{t}$ or W+jets, that represent the main sources of background of this analysis. The cross-sections of the background channels are sometimes several orders of magnitude higher than the signal cross-section. Combined with the difficulty of separating signal and background, all these lead to high systematic uncertainties and low signal significance.

For separating signal and background, this analysis uses a multivariate analysis technique, namely an artificial neural network. This thesis describes step by step what the setup of the analysis is and investigates the effect of combining the neural network default approach with kinematic fitting (thus improving reconstruction) or finding a method to separate events with cleaner topology (improving selection).

The use of kinematic fitting for the Wt lepton + jets analysis was investigated before but was never done taking into account systematic uncertainties. This thesis focuses on assessing the effect of introducing the KLFitter output variables in the initial set of variables that are fed into the neural network by looking carefully at systematic uncertainties and the expected signal significance. In the 3-jet bin, results show that adding KLFitter variables only decreases the total uncertainties by roughly 3% and the signal significance goes from 1.2σ in the default approach to 1.3σ in the combined kinematic + KLFitter analysis. The baseline analysis does not include events that have 4 jets in the final state because previous studies have shown that this does not significantly improve to the final result. However, adding the 4-jet bin to the signal extraction was also investigated here since a dedicated kinematic fit for the $t\bar{t}$ events is available in the KLFitter and this might help to separate better the Wt signal from $t\bar{t}$ background. After including the additional 79 variables that come from the $t\bar{t}$ kinematic fit and combining the 3-and 4-jet for extracting the signal, we see that the total systematic uncertainty goes down to 86%, compared to 91% for the 3-jet bin only. Minor improvement is seen when looking at the expected signal significance.

The second category of studies focuses on improving the separation by selecting only events that have a cleaner event topology. In order to do that, two neural networks were trained to separate events in which a hadronic top-quark is present at truth level and events that have at least two light-quarks

matched to the measured jets. When looking at the preprocessing output for both networks, one sees that kinematic fitting plays a very important role for the separation of the events, indeed: most of the separation power comes from KLFitter output variables. By cutting on the output of these two neural networks three different regions are defined: one that enriches hadronic top-quark events, a set of events that are matched and a sample that shares both features (i.e. matched events that also contain a hadronic top-quark). In these selected subsets of events, signal is extracted and the systematic uncertainties and expected signal significances are calculated and compared to results from the default approach. The overall best results are obtained when selecting only events with at least two matched light-quark jets. The systematic uncertainties are reduced by up to 20% and the signal significance increases by 23%.

These results give strong indication that restricting the analysis to only well-reconstructed events helps to reduce systematic uncertainties. This will be taken into account when performing the analysis on the full 8 TeV data collected in 2012.

Control plots
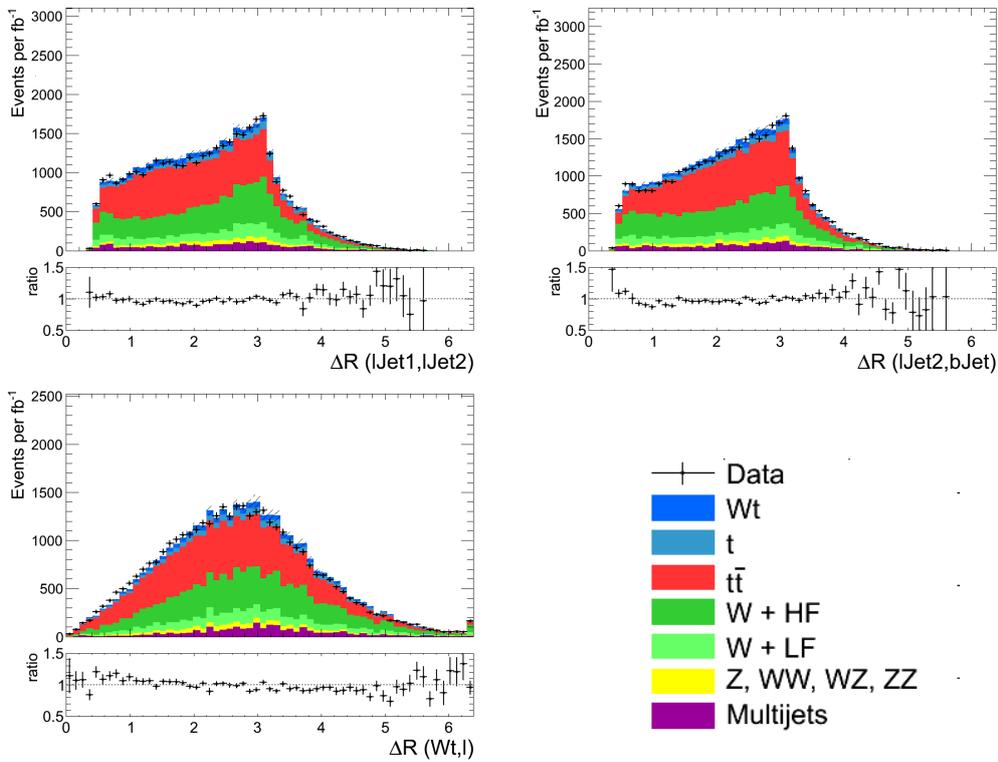
Figure A.1: Control plots of the most significant kinematic variables that are selected in the preprocessing step of the NN trainings that separate matched events and events that have a hadronic top. The top plots show the angle between the two light-quark jets (upper left), the second hardest light-quark jet and the b-jet (upper-right) and the lepton and the reconstructed *Wt* system (bottom plot). The ratio between data and Monte Carlo is also shown for each plot.
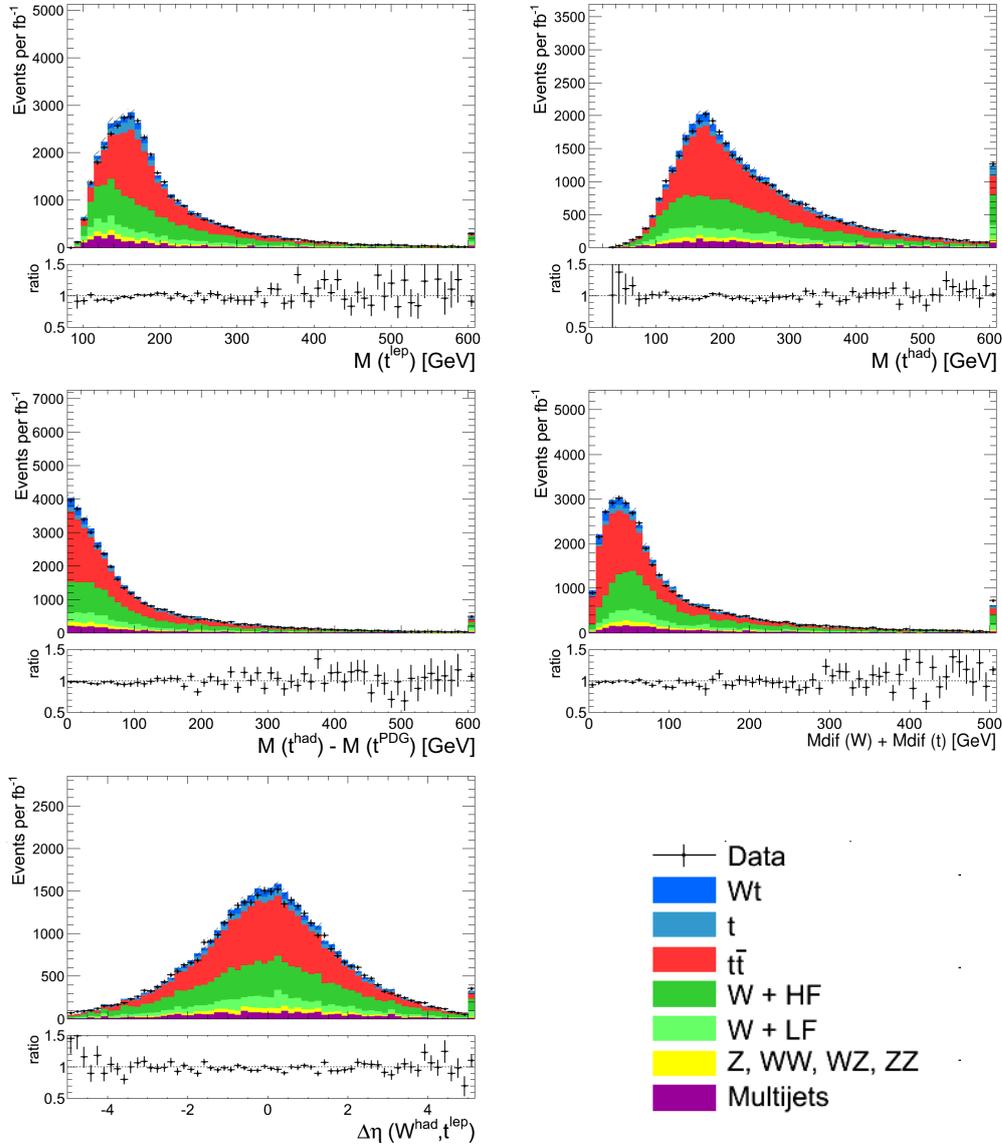
Figure A.2: Control plots of the most significant kinematic variables that are selected in the preprocessing step of the NN trainings that separate matched events and events that have a hadronic top. The ratio between data and Monte Carlo is also shown for each plot. On the top row, the reconstructed masses of the hadronic and leptonic top-quark are shown. The $2^{n}d$ row has, on the left, the mass difference between the reconstructed hadronic top quark mass and the default value given by the Particle Data Group. On the right, this difference is summed with the mass difference of the W boson.
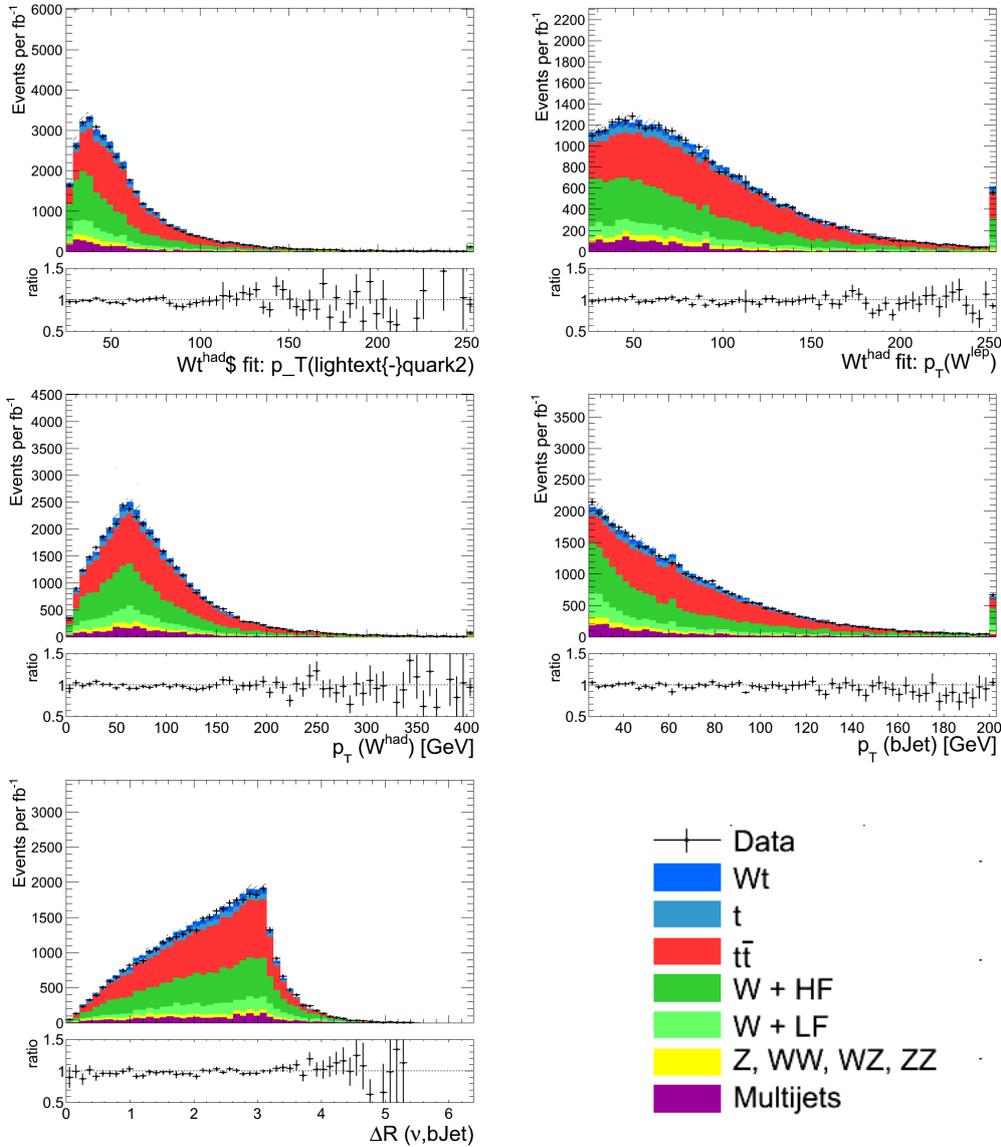
Figure A.3: Control plots of the most significant variables that are selected in the preprocessing step of the NN trainings that separate matched events and events that have a hadronic top. The ratio between data and Monte Carlo is also shown for each plot. The first row shows variables for the KLFitter, in particular the transverse momentum of two of the fitted particles. The remaining three plots are the transverse momentum of the hadronically decaying W boson (on the left) and of the b-jet (on the right). The bottom plot shows the angle between the neutrino and the b-jet.
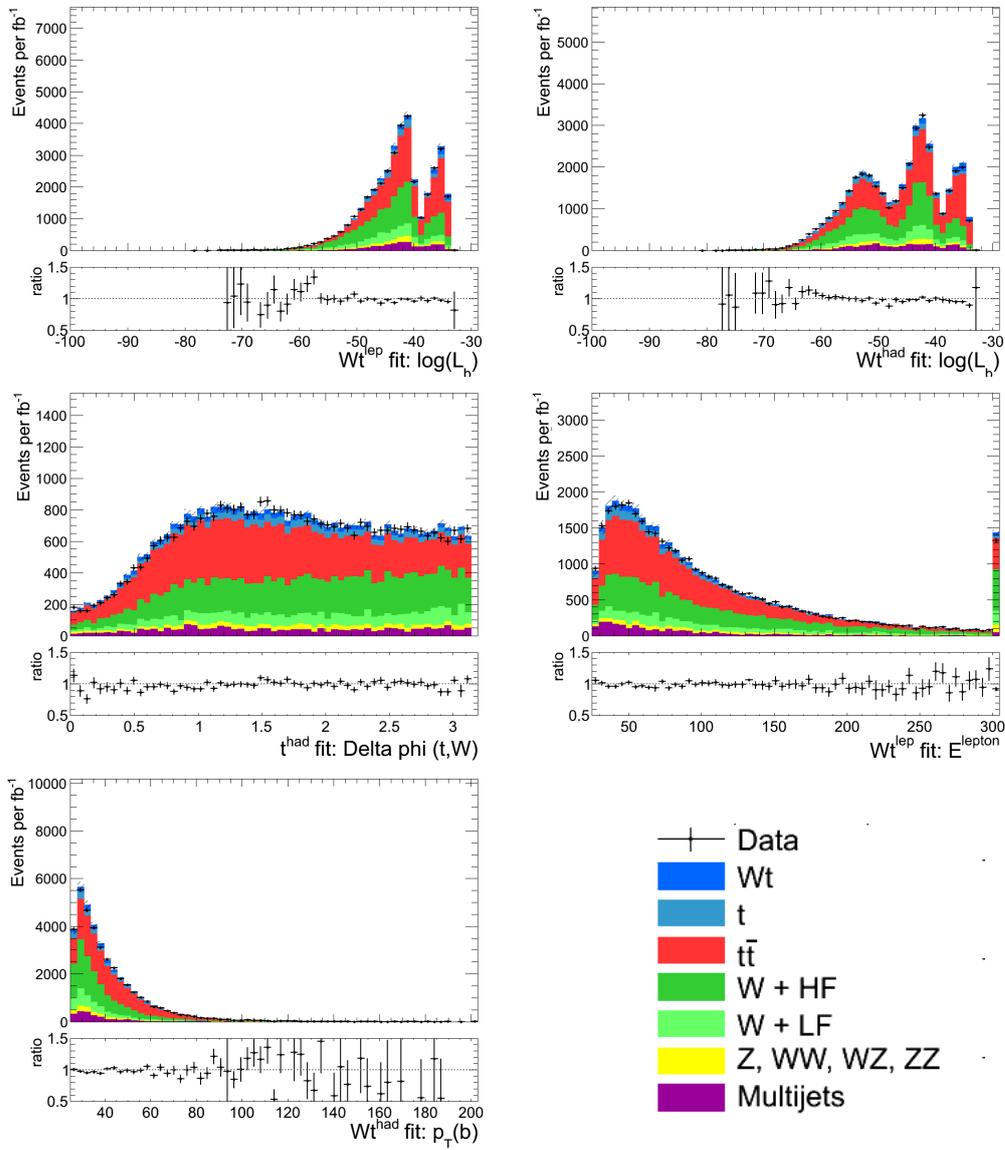
Figure A.4: Control plots of the most significant kinematic variables that are selected in the preprocessing step of the NN trainings that separate matched events and events that have a hadronic top. The ratio between data and Monte Carlo is also shown for each plot. All variables come from the KLFitter. On the upper row these are likelihood variables, while on the bottom part kinematic variables of the fitted particles are shown.

2D distributions of the neural network outputs
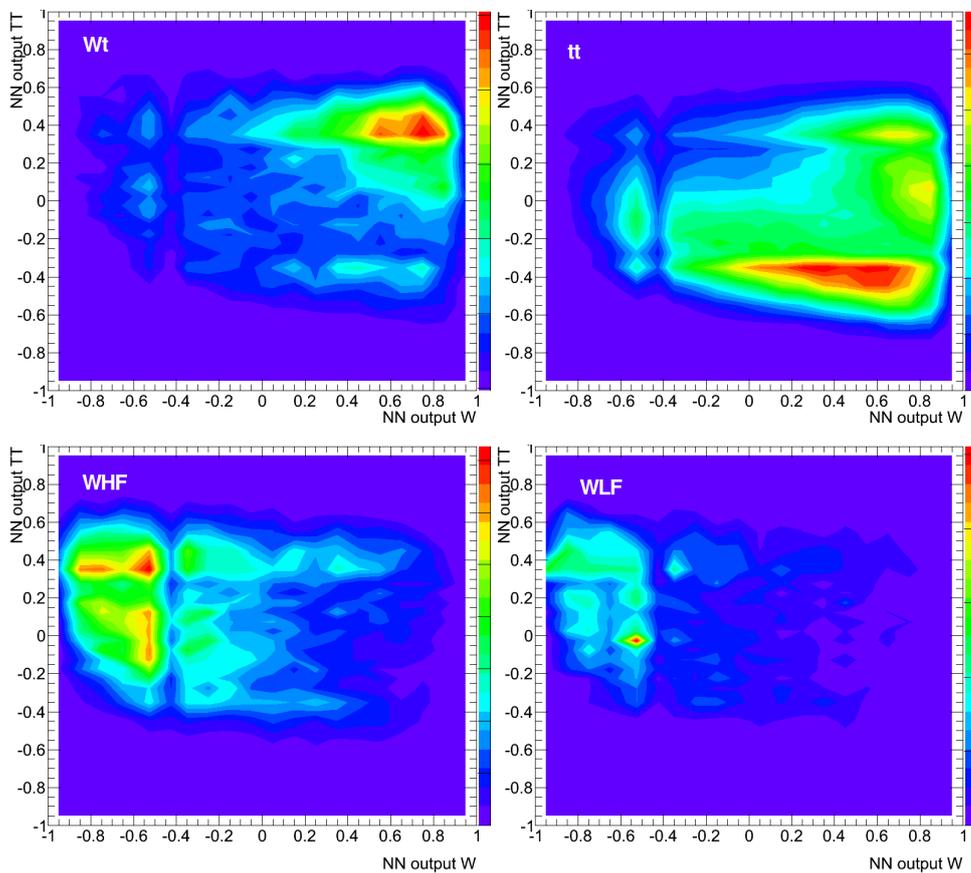
Figure B.1: 2D distributions of the NN discriminants; On the *y*-axis there is the output of the neural network trained to distinguish *Wt* events from *tt̄*, while the *x*-axis represents the resulting discriminant of the NN trained against *W* + jets; output variables from the KLFitter are included.
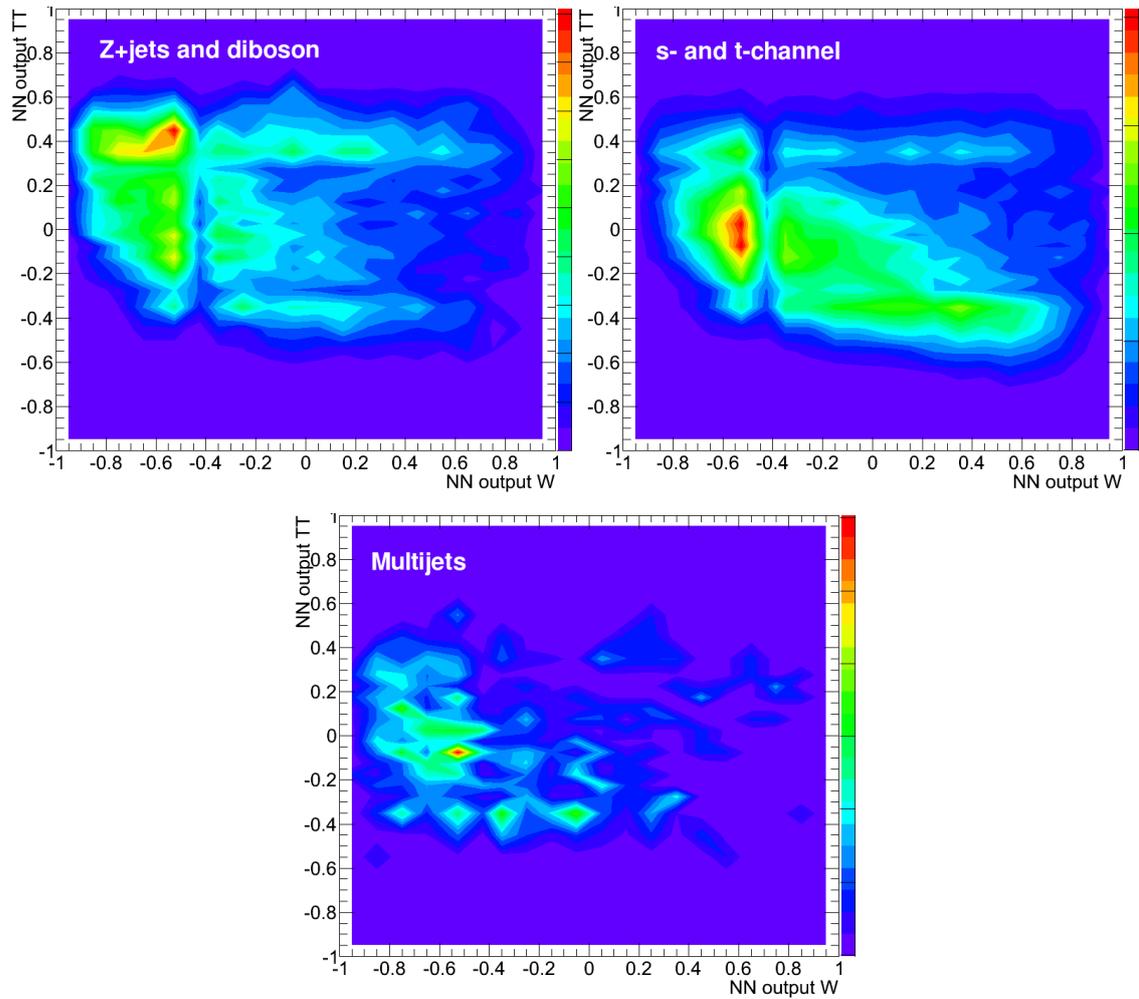
Figure B.2: 2D distributions of the NN discriminants; On the *y*-axis there is the output of the neural network trained to distinguish *Wt* events from *tt̄*, while the *x*-axis represents the resulting discriminant of the NN trained against *W* + jets; output variables from the KLFitter are included.

Systematic uncertainties

Sys Name	up[%]	down[%]
Data stat.	17	−17
BTAGL	2	−1
JVFSF	1	−1
WJetsShapePtjmin10	1	−1
MES	1	−1
ESF	2	−2
EES	2	−2
JEF	2	−2
MERID	2	−2
WJetsShapeIqopt3	2	−2
MERMS	2	−2
MSF	3	−3
EER	3	−3
DSDR	3	−3
BTAGC	5	−5
xsection	10	−10
mcstat	11	−11
BTAGB	27	−27
MC generator(PS)	29	−29
JER	29	−29
IFSR	48	−48
JES	55	−55
Total(sys)	93	−92
Total(sys+stat)	95	−94

Table C.1: Full systematics uncertainties table of the 3-jet bin default analysis.

Sys Name	up[%]	down[%]
Data stat.	22	−22
WJetsShapeIqopt3	1	−1
MERID	1	−1
MERMS	1	−1
EER	2	−2
BTAGL	4	−2
DSDR	2	−2
ESF	4	−2
JEF	3	−3
EES	3	−3
MES	3	−3
WJetsShapePtjmin10	3	−3
JVFSF	4	−4
MSF	5	−4
JER	4	−4
BTAGC	6	−5
xsection	10	−10
MC generator(PS)	12	−12
mcstat	15	−15
BTAGB	19	−19
IFSR	40	−40
JES	57	−57
Total(sys)	83	−79
Total(sys+stat)	86	−82

Table C.2: Full systematics uncertainties table of the 3-jet bin analysis for events containg a hadronic top quark.

Sys Name	up[%]	down[%]
Data stat.	21	−21
JEF	1	−1
WJetsShapePtjmin10	1	−1
MES	1	−1
WJetsShapeIqopt3	1	−1
MERMS	2	−2
DSDR	2	−2
ESF	2	−2
MERID	3	−3
MSF	3	−3
BTAGL	3	−3
EES	6	−4
BTAGC	4	−4
EER	4	−4
JVFSF	5	−6
BTAGB	10	−10
xsection	11	−11
mcstat	12	−12
MC generator(PS)	17	−17
JES	28	−27
JER	28	−28
Total(sys)	73	−72
Total(sys+stat)	75	−75

Table C.3: Full systematics uncertainties table of the 3-jet bin analysis for events with at least two light-quarks matched.

Sys Name	up[%]	down[%]
Data stat.	26	−26
DSDR	3	−3
EER	3	−3
JEF	3	−3
ESF	6	−4
MERID	4	−4
MERMS	6	−5
JVFSF	5	−5
BTAGL	7	−5
MSF	8	−7
BTAGC	8	−7
EES	10	−7
WJetsShapePtjmin10	7	−8
MES	8	−8
WJetsShapeIqopt3	8	−8
JER	13	−14
xsection	13	−13
BTAGB	16	−15
mcstat	17	−17
MC generator(PS)	24	−24
IFSR	38	−38
JES	43	−44
Total(sys)	72	−72
Total(sys+stat)	76	−76

Table C.4: Full systematics uncertainties table of the 3-jet bin analysis for events with at least two light-quarks matched and have a hadronic top quark.

Bibliography

[1] Tevatron Electroweak Working Group,
    "Combination of CDF and D0 Measurements of the Single Top Production Cross Section"
    (2009), arXiv: `0908.2171 [hep-ex]`.

[2] "Measurement of t-Channel Single Top-Quark Production in pp Collisions at $\sqrt{s}$ = 8 TeV with
    the ATLAS detector", tech. rep. ATLAS-CONF-2012-132, CERN, 2012.

[3] "Single Top associated tW production at 8 TeV in the two lepton final state",
    tech. rep. CMS-PAS-TOP-12-040, CERN, 2013.

[4] ATLAS Collaboration, "Evidence for the associated production of a W boson and a top quark in
    ATLAS at $\sqrt{s}$ = 7 TeV", *Phys. Lett.* B716 (2012) 142–159,
    DOI: `10.1016/j.physletb.2012.08.011`, arXiv: `1205.5764 [hep-ex]`.

[5] T. Loddenkötter, "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",
    BONN-IR-2012-06, PhD Thesis: University of Bonn, 2012,
    URL: `http://hss.ulb.uni-bonn.de/diss_online`.

[6] D. Griffiths, "Introduction to elementary particles" (2008).

[7] I. C. Brock and T. Schorner-Sadenius, "Physics at the Terascale" (2011).

[8] ATLAS Experiment Public Result, *Public Results (online)*, Accessed: 2013-08-01,
    URL: `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/CombinedSummaryPlots`.

[9] *Standard model infographic developed at the CERN Summer Student Webfest 2013*,
    Accessed: 2013-08-30,
    URL: `https://cds.cern.ch/record/1473657/files/SMinfographic_image.png`.

[10] The Les Houches Accord PDF Interface, *Parton distribution functions*,
    URL: `http://lhapdf.hepforge.org/`.

[11] J. Beringer et al., "Review of Particle Physics (RPP)", *Phys. Rev.* D86 (2012) 010001,
    DOI: `10.1103/PhysRevD.86.010001`.

[12] U. Langenfeld, S. Moch and P. Uwer, "New results for $t\bar{t}$ production at hadron colliders" (2009),
    arXiv: `0907.2527 [hep-ph]`.

[13] M. Cacciari et al., "Top-pair production at hadron colliders with next-to-next-to-leading
    logarithmic soft-gluon resummation", *Phys. Lett.* B710 (2012) 612–622,
    DOI: `10.1016/j.physletb.2012.03.013`, arXiv: `1111.5869 [hep-ph]`.

[14] S. Frixione et al., "Single-top hadroproduction in association with a W boson",
    JHEP 0807 (2008) 029, DOI: `10.1088/1126-6708/2008/07/029`,
    arXiv: `0805.3067 [hep-ph]`.

[15] S. Oliveira et al., "Electroweak corrections to the top quark decay",
Phys. Rev. D64 (2001) 017301, DOI: 10.1103/PhysRevD.64.017301,
arXiv: hep-ph/0011324 [hep-ph].

[16] ATLAS Collaboration, "Improved luminosity determination in *pp* collisions at $\sqrt{s}$ = 7 TeV
using the ATLAS detector at the LHC", *Eur. Phys. J.* C73 (2013) 2518,
DOI: 10.1140/epjc/s10052-013-2518-3, arXiv: 1302.4393 [hep-ex].

[17] ATLAS Collaboration, *Luminosity Public Results*, URL:
https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults.

[18] O. S. Bruning et al., "LHC Design Report. 1. The LHC Main Ring" (2004).

[19] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider",
JINST 3 (2008) S08003, DOI: 10.1088/1748-0221/3/08/S08003.

[20] ATLAS Collaboration,
"Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics" (2009),
arXiv: 0901.0512 [hep-ex].

[21] A. Talyshev, "Commissioning of the ATLAS Liquid Argon Calorimeter" (2009).

[22] A. Negri, "Evolution of the Trigger and Data Acquisition System for the ATLAS experiment"
(2012).

[23] Atlas Top Group, *Top Common Object (online)*, Accessed: 2013-07-20, URL:
https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/TopCommonObjects.

[24] V Boisvert, "Top quarks objects definition and performance at ATLAS",
tech. rep. ATL-PHYS-PROC-2013-064, CERN, 2013.

[25] ATLAS Collaboration, "Electron performance measurements with the ATLAS detector using
the 2010 LHC proton-proton collision data", *Eur. Phys. J.* C72 (2012) 1909,
DOI: 10.1140/epjc/s10052-012-1909-1, arXiv: 1110.3174 [hep-ex].

[26] M. Cacciari, G. P. Salam and G. Soyez, "The Anti-k(t) jet clustering algorithm",
JHEP 0804 (2008) 063, DOI: 10.1088/1126-6708/2008/04/063,
arXiv: 0802.1189 [hep-ph].

[27] W Lampl et al., "Calorimeter Clustering Algorithms: Description and Performance",
tech. rep. ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, 2008.

[28] D. W. Miller, A Schwartzman and D Su, "Jet-Vertex Association Algorithm",
tech. rep. ATL-COM-PHYS-2008-008, CERN, 2008.

[29] P. Seema,
"Search for single top-quark production in the *Wt* channel with 1 lepton and 2 jets at ATLAS",
BONN-IB-2013-01, Universität Bonn, 2013,
URL: http://brock.physik.uni-bonn.de/atlas_pub.php.

[30] G. Brandt et al.,
"Recommendations for measurements of vector boson production in association with jets"
(2012).

[31] S. Frixione and B. R. Webber,
"Matching NLO QCD computations and parton shower simulations", *JHEP* 0206 (2002) 029,
arXiv: hep-ph/0204244 [hep-ph].

[32]   B. P. Kersevan and E. Richter-Was, "The Monte Carlo event generator AcerMC version 2.0 with interfaces to PYTHIA 6.2 and HERWIG 6.5" (2004), arXiv: `hep-ph/0405247 [hep-ph]`.

[33]   G. Corcella et al., "HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)", *JHEP* 0101 (2001) 010, DOI: `10.1088/1126-6708/2001/01/010`, arXiv: `hep-ph/0011363 [hep-ph]`.

[34]   L Bellagamba et al., "Validation of the POWHEG MC Generator for the ATLAS Experiment", tech. rep. ATL-COM-PHYS-2010-699, CERN, 2010.

[35]   "Summary of ATLAS Pythia 8 tunes", tech. rep. ATL-PHYS-PUB-2012-003, CERN, 2012.

[36]   M. L. Mangano et al., "ALPGEN, a generator for hard multiparton processes in hadronic collisions", *JHEP* 0307 (2003) 001, arXiv: `hep-ph/0206293 [hep-ph]`.

[37]   "New ATLAS event generator tunes to 2010 data", tech. rep. ATL-PHYS-PUB-2011-008, CERN, 2011.

[38]   P. M. Nadolsky et al., "Implications of CTEQ global analysis for collider observables", *Phys. Rev.* D78 (2008) 013004, DOI: `10.1103/PhysRevD.78.013004`, arXiv: `0802.0007 [hep-ph]`.

[39]   R. Thorne et al., "Status of MRST/MSTW PDF sets" (2009), arXiv: `0907.2387 [hep-ph]`.

[40]   Top Working Group, *HFOR Tool (online)*, URL: `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/HforTool`.

[41]   "Measurement of the b-tag efficiency in a sample of jets containing Muons with Data from the ATLAS Detector", tech. rep. ATLAS-CONF-2012-043, CERN, 2012.

[42]   "Commissioning of the ATLAS high-performance *b*-tagging algorithms in 7 TeV collision data", tech. rep. ATLAS-CONF-2011-102, CERN, 2011.

[43]   M. Feindt and U. Kerzel, "The NeuroBayes neural network package", *Nucl. Instrum. Meth.* A559 (2006) 190–194, DOI: `10.1016/j.nima.2005.11.166`.

[44]   ATLAS Collaboration, "Measurement of the *t*-channel single top-quark production cross section in *pp* collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector", *Phys. Lett.* B717 (2012) 330–350, DOI: `10.1016/j.physletb.2012.09.031`, arXiv: `1205.3130 [hep-ex]`.

[45]   Atlas Top Group, *Bill Tool (online)*, URL: `https://twiki.cern.ch/twiki/pub/AtlasProtected/BillTool`.

[46]   M Stein, K Kroeninger and A Quadt, "Kinematic Fit of ttbar Events in the Semi-Leptonic Decay Channel with a chi2-Method at the ATLAS Experiment", tech. rep. ATL-COM-PHYS-2008-117, CERN, 2008.

[47]   J Erdmann et al., "Kinematic fitting of ttbar-events using a likelihood approach: The KLFitter package", tech. rep. ATL-COM-PHYS-2009-551, CERN, 2009.

[48]   "Jet energy scale and its systematic uncertainty in proton-proton collisions at $\sqrt{s}$=7 TeV in ATLAS 2010 data", tech. rep. ATLAS-CONF-2011-032, CERN, 2011.

List of Figures

List of Tables