Search for the associated production of a W boson and a top quark with the ATLAS detector at 7 TeV

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

von
Jan André Stillings
aus
Ottweiler

Bonn, 2015

Dieser Forschungsbericht wurde als Dissertation von der Mathematisch-Naturwissenschaftlichen Fakultät der Universität Bonn angenommen und ist auf dem Hochschulschriftenserver der ULB Bonn `http://hss.ulb.uni-bonn.de/diss_online` elektronisch publiziert.

1. Gutachter:	Prof. Dr. Ian C. Brock
2. Gutachter:	Prof. Dr. Jochen C. Dingfelder
Tag der Promotion:	05.03.2015
Erscheinungsjahr:	2015

For Noah

Contents

CHAPTER 1

Introduction

The Large Hadron Collider (LHC) started to collide protons with a previously unreachable centre-of-mass energy of $\sqrt{s} = 7$ TeV at the end of March 2010. This marked the starting point of the LHC Run 1, which includes three very successful years of data-taking by the ATLAS detector, one of the two large multi-purpose particle detectors at the LHC. Run 1 culminated in the discovery of the Higgs boson in mid 2012, which was one of the main goals driving the design process of the LHC and its experiments.

The Standard Model describes three of the four fundamental forces in nature (electromagnetic, weak and strong) and the elementary particles from which all visible matter in the universe is built. Since the first postulation of the foundations of the electroweak theory in the early 1960s, the Standard Model has been tested to high precision over a wide range of energies. The heaviest particle in the Standard Model is the top quark with a mass of 173 GeV. It has been observed at the Tevatron collider experiments CDF and DØ first, where many of its properties were also measured. The high collision energy of the protons at the LHC allowed its existence to be confirmed and the precision of the knowledge of its characteristics to be further improved. Top quarks can be produced in pairs via the strong interaction and individually via the weak force. Both production mechanisms occur with a large rate at the LHC, where the weak-force production reaches about one third of the top-quark pair production cross-section. The individual generation, referred to as single top-quark production, can be subdivided into three different channels: the t-channel, the s-channel and the associated production of a W boson and a top quark, which is abbreviated as Wt-channel.

The single top-quark sector is of particular interest, as top quarks are very short-lived and do not hadronise, i.e. form bound states with other quarks, before they decay. This allows direct access to the production vertex, which includes a W boson, a bottom quark and the top quark, using the reconstructed top-quark decay products. As the top quark is the heaviest quark, physics manifestations beyond the Standard Model, like additional fermion generations or alternative couplings, can be probed measuring the production cross-section of single top quarks. While the t-channel and the s-channel dominate the production processes at the Tevatron and have been isolated there, the start of the LHC opened the first opportunity to also measure the production in the Wt-channel as one of the undiscovered production processes predicted by the Standard Model. The decay products of the latter include two W bosons and a bottom quark, due to the dominant decay mode of the top quark. W bosons decay either leptonically, into a charged lepton and the corresponding neutrino, or hadronically, into two quarks. As the production cross-section for hadronic final states dominates by orders of magnitude at the LHC, charged leptons are very good to trigger on and events with one or two leptons have a higher potential for the isolation of

rare signals. While the Wt-channel production could be measured by the ATLAS collaboration already, the search for the single charged lepton final state is the topic of this thesis.

The aim of this analysis is to develop and optimize a multivariate technique to access this challenging physics channel. In addition to the detailed studies based on a neural network approach a second more sophisticated two-step method has been exploited and tested. The latest statistical tools have been used to validate the output.

This thesis is organised as follows: a short introduction into theoretical basics of the Standard Model, focussing on top quarks is given in Chapter 2 followed by a description of the LHC and the ATLAS experiment, which measured the collision data for this analysis in Chapter 3. Chapter 4 explains the reconstruction of physical objects from the raw data including their required base properties, the calibration procedures and the reachable measurement precision, followed by a description of the simulation of signal and background processes in Chapter 5. Chapter 6 contains details about the used collision data and the LHC run conditions over the recording period. It also explains the data-driven estimation of backgrounds which are difficult to model. The primary isolation of the signal from all production processes is described in Chapter 7 and Chapter 8 discusses further characteristic variables which can be used to refine it, leading to the application of the NeuroBayes software for optimal variable selection and combination into a single discriminator by an artificial neural network training. In Chapter 9 the results of these neural network trainings are shown in detail. They comprise the evaluation of the separation power, the discriminating power and influence of a variety of studied input variables and the potential of an alternative method using the combination of two neural networks to reach an even better separation. A maximum likelihood fit is applied to the discriminators to extract the Wt-channel, as explained in Chapter 10. The inclusion of systematic uncertainties in the measurements and the estimation of their significances are also explained therein, followed by the introduction of an optimisation procedure for the neural network training, which improves the overall power of the method. The thesis concludes with a summary and review of the analysis results in Chapter 11.

Considerable effort has been invested into the development of software for the analysis that has been later extended in collaboration with other members of the ATLAS collaboration in Bonn and other institutions. Appendix D sketches the technical solutions for general data analysis problems in high-energy physics analyses and discusses sociological and management aspects of a joint software development project in the field.

CHAPTER 2

Theoretical Basics

High energy particle physics is the research field which searches for elementary particles and investigates their properties and interactions. So far the Standard Model is the accepted and, to a great extent, verified theory in the field and a basic introduction of it will be given in the following. The ATLAS detector, together with the other experiments at the Large Hadron Collider (LHC), has probed this theory to previously not reachable energies and found one of its missing but predicted particles: the Higgs boson [1, 2]. The aim of this chapter is to set the scene for the following analysis and no attempt is made to cover the mathematical concepts of the Standard Model in detail. The discovery of the top quark, its properties, its production at the LHC and also the types of its possible decays will be discussed, focussing on the single top-quark production in association with a W boson, which is the target channel of this analysis.

A good overview of the current status of the Standard Model just before the discovery of the Higgs boson can be found in [3] and introductory textbooks are also available in a good variety, e.g. [4] or [5]. The introduction of the Standard Model follows the one given in [6].

2.1 The Standard Model

The Standard Model is a gauge theory describing the fundamental interactions between the elementary particles except for gravity[1]. The underlying gauge symmetry has the form

$$\mathrm{SU}(3)_C \times \mathrm{SU}(2)_{I_W} \times \mathrm{U}(1)_Y.$$

This symmetry unifies two basic theories: the Glashow-Salam-Weinberg (GSW) theory for the weak and electromagnetic interactions and quantum chromodynamics (QCD) for strong interactions. QCD is based on the $\mathrm{SU}(3)_C$ symmetry group. The corresponding conserved quantum number is called colour charge or just colour (C). It can have three different values called red, green and blue. The strong force is mediated by gauge bosons (spin 1), the gluons. Due to the non-abelian nature of $\mathrm{SU}(3)_C$, gluons interact with each other and hence carry colour. Each gluon carries both colour and anticolour resulting in eight different color states. Additionally gluons are postulated to be massless and have no electric charge.

[1] While the missing integration of the gravitational force is unfortunate and subject to large efforts in theoretical physics, its importance at the microscopic scales of high energy particle physics is negligible compared to the other three.

In the electroweak sector, the GSW theory uses the non-abelian $SU(2)_{I_W} \times U(1)_Y$ gauge group. Quantum numbers for this group are the weak isospin I_W and the hypercharge Y. In the electroweak case we have four gauge bosons mediating the forces, namely an uncharged photon, two charged W bosons and one uncharged Z boson. Unlike in QCD, $W^{\pm}$ and Z^0 bosons are found to have mass. To keep the theory gauge invariant the symmetry is spontaneously broken by introducing an additional scalar field, the Higgs field. This field is responsible for mass generation of fermions and the electroweak gauge bosons and also yields a Higgs boson. After breaking the overall symmetry, the theory remains invariant under $U(1)_{EM}$ transformations which describe the electromagnetic interactions. Therefore the photon is massless and the electromagnetic charge, Q, is the corresponding conserved quantum number. Photons do not carry electric charge as $U(1)_{EM}$ is an abelian group.

Matter particles in the Standard Model are exclusively fermions (spin 1/2). They appear in both left-handed and right-handed states. Only neutrinos are supposed to be left-handed only in the Standard Model[2]. The left-handed fermions can be grouped into doublets according to their weak isospins:

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L$$

$$\begin{pmatrix} u \\ d \end{pmatrix}_L \quad \begin{pmatrix} c \\ s \end{pmatrix}_L \quad \begin{pmatrix} t \\ b \end{pmatrix}_L$$

These doublets are also called generations. Right-handed fermions only form isosinglets. Doublets in the upper row are leptons, in the lower row the quarks are organised. Each of the listed fermions also has an antiparticle with opposite charge and different hypercharge.

The electron (e) is the lightest charged lepton and was the first to be discovered by J.J. Thomson in 1897 [7]. Thomson concluded from experiments with cathode rays that they consist of single particles and already could determine the charge over mass ratio of the electrons. The electron has an electric charge $Q = -1\mathrm{e}$, like all charged leptons, and a mass of 0.5 MeV. Its antiparticle is called positron. Muons (μ) are rather exact copies of electrons despite their larger mass of 105.7 MeV. This mass difference was the basis for the discovery by Anderson and Neddermeyer in their extensive research of cosmic ray tracks [8]. Contrary to their lighter partners, muons are not stable and decay weakly into electrons and neutrinos with a lifetime of 2.2 µs. The third and heaviest type of charged leptons are tau leptons (τ) or just taus. They were discovered by Martin Perl in 1974 with the Mark I detector at the SLAC e^+e^- collider [9–11]. Taus are the heaviest charged leptons with a mass of 1776.8 MeV and a lifetime of 290.6×10^{-15} s. Taus mostly decay to neutrinos, kaons, pions and other mesons but a fraction of about 35 % also decays via three-body decays to electrons or muons and neutrinos, which were the discovery channels at Mark I. All the previous mentioned charged leptons interact electromagnetically and weakly with other particles. They carry no colour charge and thus are not influenced by the strong force.

Mentioned briefly in the possible decay modes above is the second class of leptons called neutrinos. They occur in three different states named after their charged leptonic partners: electron neutrinos (ν_e), muon neutrinos (ν_μ) and tau neutrinos (ν_τ). Neutrinos carry neither electromagnetic charge nor colour and only interact weakly with other particles which makes their detection hard. The existence of neutrinos was first postulated by Wolfgang Pauli in 1930 to explain the non-discrete spectrum of beta decays. Experimentally the neutrino was only accessible after the development of nuclear reactors. Reines and

[2] As experiments show small neutrino masses, the Standard Model has to be expanded. The current two models are to introduce Dirac neutrinos or a new type of particles called Majorana neutrinos.

Cowan finally found the electron neutrino in 1956[3] [14]. The Standard Model assumes all neutrinos to be massless which newer experiments contradict[4].

The second big group of massive particles are the quarks. Like neutrinos, their first appearance happened in theory. M. Gell-Mann and G. Zweig independently postulated the first quark model in 1964 [17–19] to explain the large and growing number of found hadrons. The experimental discovery of quarks started with the observation of the proton substructure in deep inelastic scattering experiments at the Stanford Linear Accelerator (SLAC) in the late 1960s, which was acknowledged with a Nobel Prize to J. Friedman, H.W. Kendall, and R.E. Taylor in 1990 [20–22]. While initially only consisting of three quarks, namely up, down and strange quarks, the theoretical model soon had to be expanded. In 1974 a fourth quark, the charm quark, was found by B. Richter and S. Ting independently detecting the J/ψ resonance [23, 24]. The third generation's bottom-quark discovery followed in 1977 by a L.M. Lederman leading the E288 experiment at Fermilab through a resonance of dimuon final states [25]. The completion of the third generation saw many years of efforts, which were only rewarded with the discovery of the top quark in 1995 [26, 27]. Quarks are the only particles that couple to all three interactions as they carry electric charge, colour and weak hypercharge. The properties of the quarks are given in Table 2.1.

Quark	Charge	Mass	Flavour
Up	$+\frac{2}{3}$	$2.3^{+0.7}_{-0.5}$ MeV	$I_z = +\frac{1}{2}$
Down	$-\frac{1}{3}$	$4.8^{+0.5}_{-0.3}$ MeV	$I_z = -\frac{1}{2}$
Charm	$+\frac{2}{3}$	1.275 ± 0.025 GeV	Charm = +1
Strange	$-\frac{1}{3}$	95 ± 5 MeV	Strangeness = -1
Top	$+\frac{2}{3}$	$173.21 \pm 0.51 \pm 0.71$ GeV	Top = +1
Bottom	$-\frac{1}{3}$	4.18 ± 0.03 GeV	Bottom = -1

Table 2.1: Quark properties from [28].

A special conserved quantum number of quarks is given in the table: the flavour. Up and down quarks are characterised by the third component of the isospin (I_z) while the other quark types have none ($I_z = 0$) and therefore their quark flavour is given. Similar quantum numbers also exist for leptons: the leptonic family number and the lepton number. The lepton number counts the leptons coming from an interaction vertex positively while antileptons are counted negatively. The leptonic family number is counted in the same way but only considering leptons of the same family. Leptonic family numbers and lepton numbers are also conserved in the Standard Model if we consider zero neutrino masses. Leptons in one family thus can not be directly changed into a lepton of a different family. In case of neutrino oscillations this principle is violated. For the quarks the situation is different. In QCD the flavour is conserved like in the leptonic case, though for the electroweak theory we have a breaking of this symmetry. In the theory the transition probability between quark types is given by the Cabibbo-

[3] The muon neutrino was found by L.M. Lederman, J. Steinberger and M. Schwartz at the AGS experiment in Brookhaven [12]. The tau neutrino has only been detected in 2000 with the DONUT detector at Fermilab [13].
[4] Neutrino experiments have shown neutrino oscillations which implies that neutrinos have a small mass [15, 16].

Kobayashi-Maskawa (CKM) quark mixing matrix:

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

The CKM matrix connects the mass eigenstates of the quarks (d, s, b) to the corresponding weak eigenstates (d', s', b'), which are different in GSW. With the current best fit values one can see that transitions between the generations are suppressed while transitions between quarks inside a family are favoured:

$$\begin{pmatrix} 0.97427 \pm 0.00014 & 0.22536 \pm 0.00061 & 0.00355 \pm 0.00015 \\ 0.22522 \pm 0.00061 & 0.97343 \pm 0.00015 & 0.0414 \pm 0.0012 \\ 0.00886^{+0.00033}_{-0.00032} & 0.0405^{+0.0011}_{-0.0012} & 0.99914 \pm 0.00005 \end{pmatrix}$$

More details about the CKM matrix elements can be found in [28] and [29].

The largest problem about the Standard Model are the numerical parameters which were shown so far. These parameters are not included in the theory, but have to be chosen and hence have to be measured by experiments. In total, the Standard Model has 18 free parameters: nine fermion masses, three CKM mixing angles and a complex phase, one coupling constant for each of the interactions (α, α_s and G_F), the mass of the Z^0 boson and the mass of the Higgs boson. An additional seven parameters have to be added taking neutrino masses into account. All these parameters have been measured to high precision as the uncertainties on the lepton masses and the CKM matrix elements suggest. It is the task of the experiments to search for deviations and inconsistencies in the Standard Model which can hint at new physics beyond this remarkably successful model.

2.2 Modelling proton collisions

The collision of protons can be of elastic or inelastic nature. In the elastic case, the protons remain intact and the change of the system can be characterized by the scattering angle exclusively, which depends on the momentum transfer between the two protons. When colliding protons with high energies, not the protons themselves but their point-like subcomponents, named partons, interact with each other as protons are compound objects. This parton model has been proposed by R. Feynman [30] and proven to be a very good description for the processes occurring at high energies.

Following this model, the partons carry a fraction, x, of the total momentum of the mother proton. The distribution of the longitudinal parton momenta, x_i, inside the proton i is given by the parton distribution functions (PDF), $f_i(x_i, \mu^2)$. They show a dependence on the momentum fraction, x_i, and the square of the transferred momentum, μ^2, which is the energy scale for an event. This is depicted in the left part of the schematic in Figure 2.1. The partons were later identified as the quarks which the proton consists of: three valence quarks, two up quarks and a down quark, gluons, as mediators of the strong force which hold the proton together, and sea quarks, which are created by splitting of the gluons into quark and antiquark pairs and annihilate with each other most of the time.

To measure how often a specific production process occurs, particle physics analyses measure the cross-section of the process. This variable corresponds to an effective area around a target particle which has to be hit to trigger this process and is hence measured in units of the area, m^2. As cross-sections in particle physics are usually very small, a special unit called *barn*, b, has been introduced, which is defined as

$$1\,\mathrm{b} = 1 \times 10^{-28}\,\mathrm{m}^2.$$

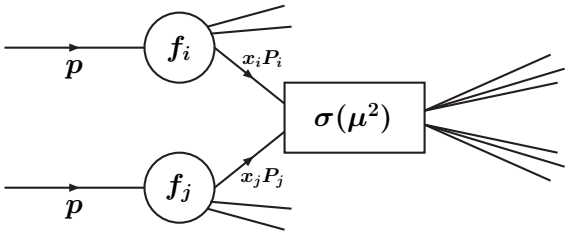


Figure 2.1: Graphical representation of the factorisation theorem for proton collisions: f_i and f_j represent the proton distribution functions (PDFs) and σ the cross-section of the hard scattering process.

Evaluating the production process of interest where two partons interact with each other at very short distance it can be assumed that the strong coupling of the partons is weak. This allows the process to be computed analytically using a perturbative expansion in orders of the strong coupling constant, α_s. In practice, this is done by evaluation of all Feynman diagrams of the process for a specific order of the perturbation, i.e. exponential order of α_s. Leading-order (LO) calculations are often single graphs for a process, while the next-to-leading order (NLO) includes already a large number of graphs with radiative corrections, loops, etc. The result of this calculation is called the hard process of the interaction and shown on the right side of Figure 2.1.

The combination of the two parts allows the full cross-section of the process $pp \to X$ to be calculated according to the factorisation theorem [31]. The hard process depends on the initial energy of the interacting partons, which is in turn dependent on the centre-of-mass energy of the proton collision. One big problem of the PDFs is that they cannot be calculated analytically with pertubative QCD and thus have to be determined from experiments like deep-inelastic scattering. The PDFs used at LHC were measured in fixed target experiments and especially at the Tevatron and HERA colliders [32]. Finally, both factors depend on the factorisation scale, μ, which is used to define which parts of the interaction are specified in the hard process and which part is absorbed into the PDF. As the PDFs are generally determined at a lower factorisation scale than needed for the TeV energies at the most powerful hadron colliders, the PDFs are scaled to a higher energy scale by using the DGLAP equations[5] [33]. In Figure 2.2, two example parametrisations are shown from the most recent HERAPDF2.0 set [34]. These plots nicely show the distributions of valence quarks, sea quarks and gluons in the proton depending on x. At high x, most of the proton momentum is carried by the three valence quarks while at low values x gluons largely dominate, followed by the sea quarks. Valence quarks hardly play a role any more, especially keeping in mind that the gluon and sea-quark distributions are reduced by a factor of 20 in the plots.

2.3 Event generation

The mechanisms described in the previous section, are not only used to calculate predictions for the cross-section of a certain production process, but also to simulate collision events. As this requires randomisation of particle properties according to the theoretical predictions, Monte Carlo methods have been proven to be a very successful instrument, especially with the vast improvements in computing power of the last decades. The software tools are also called *Monte Carlo generators*. The goal of such a generator is to simulate the final state particles which can be detected using the final experiment out of the initial proton collision. This is done in several steps, which are sketched here:

[5] Short for Dokshitzer-Gribov-Lipatov-Altarelli-Parisi QCD evolution equations.

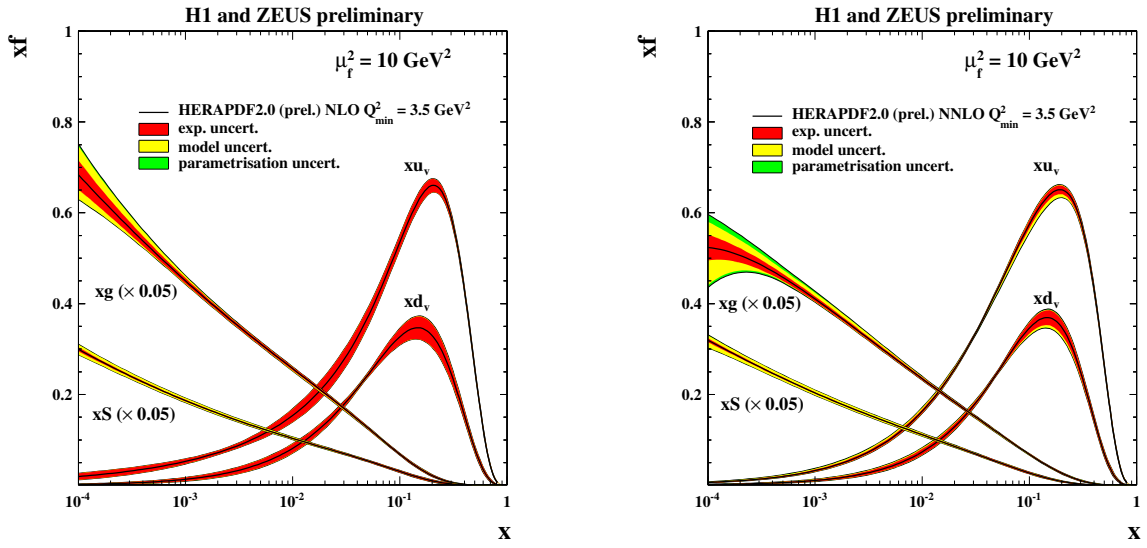


Figure 2.2: The parton distribution functions from HERAPDF2.0 (prel.) at NLO (left) and NNLO (right) at the factorisation scale $\mu^2 = 10\,\mathrm{GeV}^2$. xu_v and xd_v are the distributions for valence quarks, xS for sea quarks and xg for gluons. The gluon and sea-quark distributions are scaled down by a factor of 20. The experimental, model and parametrisation uncertainties are shown separately [34].

- *Parton distribution functions:* as explained above, the PDFs describe the contents of the two colliding protons and define which partons interact as well as their kinematics based on the randomised factors x and the previously set factorisation scale μ.

- *Hard scattering process:* given the two interacting partons and their kinematics from the PDFs, the hard scattering process is calculated. Depending on the Monte Carlo generator, this step is taken using perturbative QCD calculations as detailed above with leading order or next-to-leading order precision. Most processes can be simulated using the higher precision of the latter and theoretical modules are under development to reach the next-to-next-to-leading order (NNLO). The outcome of the first two steps of the calculation are the immediate final-state particles of the desired process. This is mostly not a physically observable final state as it consists of single particles which carry colour charge.

- *Parton showering:* particles which carry electric or colour charge are prone to spontaneous emission of photons and gluons. The parton showering algorithms simulate these emissions as well as gluon splitting before and after the hard scattering, which is referred to as *initial-state radiation* (ISR) and *final-state radiation* (FSR). This radiation is partially included into NLO calculations of the hard scattering process but also takes higher order corrections into account. It is a very delicate topic not to double the generation of emissions once by the hard-scattering process and once by the the parton shower. The according technique to prevent it is called *parton shower matching*.

- *Hadronisation:* quarks and gluons, as colour charged objects, cannot be observed directly, but always form colour-neutral compound states of two quarks, mesons, or three quarks, baryons. The creation of the hadronic final state of the generated process is called *hadronisation* and is a very important part of the event generation, as only hadronic final states can be measured in

the detectors. It cannot be calculated analytically using perturbative QCD but phenemenological models exist which are implemented into the Monte Carlo generators. A good introduction into these models can be found in [3] and in the documentation of the used generator.

- *Particle decay:* both leptons and hadrons which are produced in the hard process and the hadronisation often have small lifetimes. The Monte Carlo generators also simulate the decay of such particles correctly. As the decay of a particle can follow different chains, the proper branching ratios must be taken into account.

- *Underlying event:* given the high energy of proton collisions the event is not just limited to the hard process. The *underlying event* contains the parts of the collision which are not included into the calculations so above, e.g. multi-parton interactions and parton showers from ISR and FSR.

Monte Carlo simulations of collision events have become a very powerful and fundamental tool in particle physics. The output of the event generation are the four-vectors of all particles which are created as a result of a collision. Events can be generated in huge amounts, which are usually larger than the measured data and are only limited by the available computing power. The samples of simulated events are used in a large variety of tasks, such as the determination of detector efficiencies and acceptance, comparison of the measured data with theoretical predictions and the interpretation of the measurements. Before large experiments are even built, Monte Carlo generators are already used to evaluate the performance of the detectors, to develop and test analysis techniques and to study the attainable precision of physics results. More details on the Monte Carlo generators used in this analysis can be found in Chapter 5.

2.4 The top quark

2.4.1 Discovery of the top quark

In 1973 M. Kobayashi and T. Maskawa presented an extension to the electroweak theory that included a third generation of quarks[6] [35]. The first step to verify this theory was taken with the discovery of the bottom quark in 1977 at Fermilab [25]. Over the following years more and more physicists were convinced that a sixth quark was missing in the model and many experiments narrowed the constraints on the particle's mass. A first direct measurement was finally accomplished by the CDF and DØ collaborations at Fermilab in 1995 [26, 27]. Due to the measurements of the properties by the Tevatron experiments, the LHC experiments could already identify candidate events for top-quark production shortly after the launch of the experiment and delivered first measurements shortly afterwards [36, 37].

2.4.2 Top-quark properties

Given the large mass of the top quark of about 173 GeV, the Tevatron collider was the first machine that could produce top quarks allowing indirect constraints of the top-quark's properties only. Since the first candidate events, the CDF and DØ collaborations have made big efforts to improve the knowledge about the top-quark's main characteristics. With the availability of the Large Hadron Collider, the ATLAS and CMS collaborations joined the effort and published a large number of measurements, both at 7 and 8 TeV collision energies. The current status of the measurements will be summarised briefly in the following.

[6] This extension made it possible to explain the CP symmetry violation in kaon decays which would not occur in the GSW model for four quarks.

Mass The top quark is the heaviest known elementary particle and its mass is one of the important parameters of the Standard Model. All four experiments which can measure top quarks have combined [38] the results of their most precise measurements into a world average of the top-quark mass:

$$M(t) = 173.34 \pm 0.27\,(\text{stat.}) \pm 0.71\,(\text{syst.})\,\text{GeV}.$$

The combined error of 0.76 GeV corresponds to a measurement precision of about 4 ‰, which is much better than all other quarks. This precision can only be reached because of the short lifetime and the resulting special decay properties compared to other quarks which will be highlighted below. An overview of the measurements entering the combination is shown in Figure 2.3. It is visible that each experiment

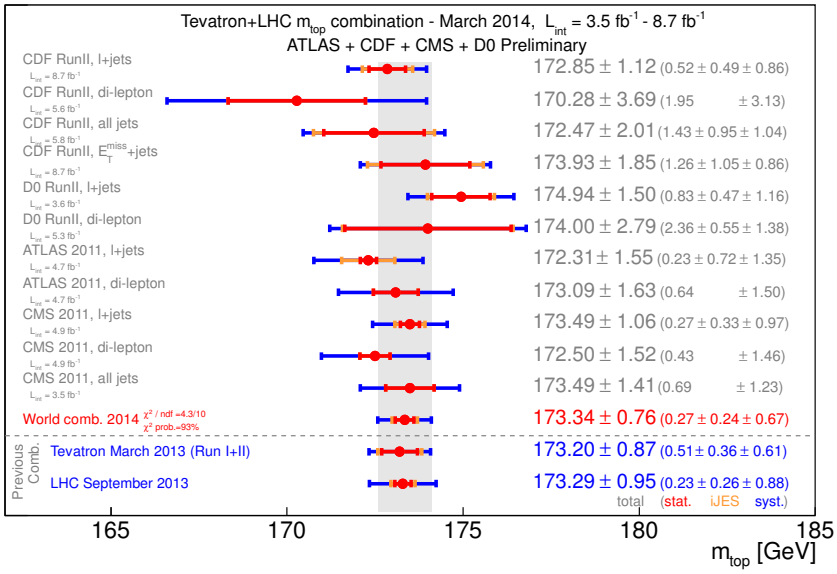


Figure 2.3: Summary of the input measurements and resulting world average mass of the top quark. For each measurement, the total uncertainty, the statistical and the iJES contributions (when applicable), as well as the sum of the remaining uncertainties are reported separately. iJES is the part of the jet energy scale uncertainty which results from in situ calibration procedures. The grey vertical band reflects the total uncertainty on the combined $M(t)$ value [38].

contributed measurements of at least two final states. The last two lines show the single combinations by collider. Tevatron's combination is limited by statistics while the LHC experiments reach a similar precision with a dominating systematic uncertainty.

Due to its large mass, the top quark plays an important role in the Standard Model. One prominent example is higher-order corrections to gauge boson propagators due to fermion loops and higher-order vertex corrections. The top-quark mass is large enough to make these corrections measurable even below the energy threshold for top-quark production at the LEP e^+e^- collider [39]. Another fact is the strong coupling to the Higgs particle, which makes the top quark an excellent probe for the theory of the electroweak symmetry breaking.

Decay width and lifetime In the Standard Model the lifetime of the top quark is predicted to be around $\tau \sim 4 \times 10^{-25}$ s, a very small value compared to other particles. This corresponds to a travelled distance of $c\,\tau \sim 1.5 \times 10^{-16}$ m in the detector, which cannot be measured directly. What can be mea-

sured instead is the decay width of the top quark. The most precise measurement has been performed by the CDF collaboration at the Tevatron collider using the full available dataset of run II [40]. A resulting limit of $1.10 < \Gamma(t) < 4.05\,\text{GeV}$ has been obtained at the 68 % confidence level for a top-quark mass of 172.5 GeV which corresponds to a lifetime limit of $1.6 \times 10^{-25} < \tau(t) < 6.0 \times 10^{-25}$ s. This result is well compatible with the prediction of the Standard Model. The short lifetime has a very important consequence: the top quark does not hadronise or form bound states before decaying. This has direct implications on the decay products of the top quark. As it does not hadronise, the top quark can be seen as quasi-free and hands its spin information over to the decay products. A measurement of these properties has been carried out by all top-quark producing experiments and are a test for the couplings active at the top decay vertex [41, 42].

Electric charge Like all other up-type quarks in the Standard Model, the top quark is predicted to have a charge of $Q = +2/3$ e. The electric charges of the other quarks have been successfully determined at e^+e^- colliders. Since the so far built e^+e^- machines operated below the production threshold of the top quark, the charge has yet to be measured. The most precise measurement at hadron colliders is reported by the ATLAS experiment [43]. The result of the measured charge in units of the elementary charge is $0.64 \pm 0.02\,(\text{stat.}) \pm 0.08\,(\text{syst.})$. A model which proposed a heavy quark with an electric charge of $Q = -4/3$ e has been probed at other experiments. The ATLAS measurement finally ruled it out with a strong significance of eight standard deviations.

2.4.3 Top-quark production at the Large Hadron Collider

Top quarks at the Large Hadron Collider can be created either in strong-force or in weak-force interactions. The first type of production leads to the creation of pairs of top and antitop quarks and his hence named top quark pair production or $t\bar{t}$ production. The process can be induced either by quarks, as shown in the Feynman diagram in Figure 2.4(a) or by gluons, as shown exemplified in Figure 2.4(b). A

(a) $qq \to t\bar{t}$ (b) $gg \to t\bar{t}$

Figure 2.4: Examples for Feynman diagrams of the quark-induced (a) and gluon-induced (b) production of top quark and antitop-quark pairs at the leading order or perturbative QCD.

small calculation shows which one plays the dominant role at the Large Hadron Collider. To produce a $t\bar{t}$ pair at rest, at least twice the top-quark mass is needed as centre-of-mass energy, $\sqrt{s}$, in the collision: $\sqrt{s} \geq 2M(t)$. Calculating the energy in the hard process, W, and assuming $x_1 \approx x_2 \equiv x$ we can derive

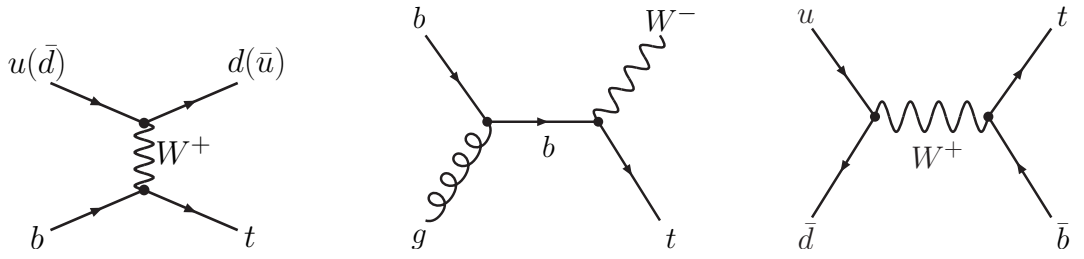
$$W = \sqrt{(x_1 P_1 + x_2 P_2)^2} = x\sqrt{(P_1 + P_2)^2} = x\sqrt{s}$$

and approximate x with $x \approx \frac{2M(t)}{\sqrt{s}} = 0.05$ at 7 TeV proton collision energy and a top-quark mass of 173 GeV. This shows that x gets lower with increasing collision energy and the gluon PDF dominates the proton structure in the creation of top-quark pairs. The cases where x_i and x_j are asymmetric also

contribute to the dominating process as they provide incoming low-x gluons. For the larger x at Tevatron the valence-quark PDFs are in the same order and a bit higher than the gluon PDF and quark-antiquark annihilation is the dominant process.

The cross section for $t\bar{t}$ production at the LHC for 7 TeV has been calculated with next-to-leading order (NLO) precision including next-to-leading logarithmic (NLL) soft-gluon resummation for a top-quark mass of 172.5 GeV. A cross section of 177^{+10}_{-11} pb has been predicted using the method detailed in [44].

The second type of production is the creation of single top quarks in weak interactions. The three production channels defined at leading order are the t-channel, associated production of a top quark and a W boson (Wt-channel) and the s-channel production. Examples for Feynman graphs for each of these channels are shown in Figure 2.5.



(a) t-channel (b) Wt-channel (c) s-channel

Figure 2.5: Examples for Feynman diagrams of single top-quark production for each of the three production channels: t-channel (a), Wt-channel (b) and s-channel (c).

t-channel and s-channel production

In the t-channel production a b quark from a gluon-splitting in the proton fuses with a virtual W boson to form a top quark. The second b quark created in the process is usually omitted because it cannot be measured by the detector as it escapes to the forward region. The production cross-section is estimated to be $64.6^{+2.6}_{-1.7}$ pb [45] at $\sqrt{s} = 7$ TeV for a top-quark mass of $M(t) = 172.5$ GeV. Hence, the t-channel production is the dominant process of single top-quark creation. It has been discovered at the Tevatron in combination with the s-channel in 2009 [46, 47]. In the latter case, a quark and an antiquark of different types interact and form a time-like W boson which in turn decays to a single top quark and a $\bar{b}$ quark. Its production cross-section in proton collisions at 7 TeV is predicted to be $4.63^{+0.20}_{-0.18}$ pb [48], which is the smallest of the three production channels. At the LHC, the s-channel plays a much smaller role, as it is suppressed compared to the Wt-channel due to the small antiquark PDF compared to the Tevatron, where antiprotons were part of the collision. This is also the reason why the s-channel production could only be observed at Tevatron so far [49]. The comparably large cross-section of the t-channel, in contrast, has triggered a rich program of measurements at the ATLAS and CMS experiments. The cross-section could be measured with good precision at a collision energy of 7 TeV in both experiments [50, 51], where ATLAS also measured the top-quark and antitop-quark cross-sections separately and determined their ratio. Further analyses have also been published using the larger available luminosity of the 2012 run of the LHC at $\sqrt{s} = 8$ TeV, which is not topic of this thesis.

Wt-channel production

The channel that this analysis searches for is the associated production of a W boson and a top quark, named Wt-channel, or also tW production in more recent publications. A b quark originating from a gluon splitting in one of the colliding protons interacts with a gluon of the other proton and emits a W boson while transforming into a top quark. While it plays no role at the Tevatron due to its small production cross-section, it is the second most important channel for top-quark production mediated by the weak force at the LHC. Its predicted cross-section reaches $15.7 \pm 1.2\,\text{pb}$ [52] for proton collisions at a centre-of-mass energy of $\sqrt{s} = 7\,\text{TeV}$. To understand the peculiarities of this channel, the decay of top quarks need to be understood. Therefore, the further discussion is deferred to Section 2.5.

2.4.4 Top-quark decay

As discussed in Section 2.4.2 the top quarks have an average lifetime of

$$\tau \sim 4 \times 10^{-25}\,\text{s}$$

and thus do not travel far enough to get detected even by the high-resolution detectors built for the LHC experiments. To study its properties, it is therefore needed to measure the decay products, which can be used to infer the properties of their parent particles. Precise information about the decay of the top quark is thus needed.

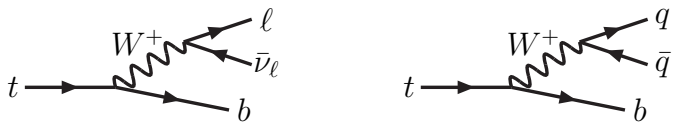


Figure 2.6: Feynman diagrams for leptonic (left) and hadronic (right) top-quark decays.

Unlike other quarks, the top quark does not persist long enough to form a bound state and decays directly[7]. It decays nearly exclusively to a W boson and a b quark via a weak decay. This decay is called Cabibbo-favoured as the corresponding CKM matrix element is large compared to the ones of the other possible decays.

$$R = \frac{B(t \to Wb)}{B(t \to Wq)} = \frac{|V_{tb}|^2}{|V_{td}|^2 + |V_{ts}|^2 + |V_{tb}|^2} \approx 1$$

The resulting W boson and b quark are not stable particles either. The b quark hadronises into a B meson which mostly decays weakly into a D meson and finally into stable particles with a jet signature, which is a collimated shower of hadrons, electrons and photons. The decay into a D meson is again Cabbibo-favoured as the CKM matrix element V_{ub} is much smaller than V_{cb}. The W boson on the other hand decays in different channels into stable and unstable particles. It can either decay leptonically into a lepton (e, μ or τ) and its corresponding neutrino or hadronically into a quark and an antiquark. The corresponding Feynman diagrams are shown in Figure 2.6. Due to kinematical reasons, the quark-antiquark combinations in the hadronic decay are only composed of light quarks ($u\bar{d}$ and $c\bar{s}$ for W^+).

The two decay classes define the possible final states of top-quark production using simple combinatorics. While the single top-quark production in the t-channel and s-channel only involves one W boson, which results from the decay of the mother top quark, two W bosons are present in the Wt-channel and the pair production of top quarks. In the latter two cases we distinguish into three different final states, which are usually analysed exclusively by the experiments' collaborations:

[7] The b quark for example forms bound states with lower mass quarks to form a B hadron before decaying.

- **all-hadronic final state:** both W bosons decay hadronically into light-quark jets. While this final state has a large branching fraction it experimentally problematic as it is only composed of jets. Such events are very hard to separate from background production and also difficult in terms of the reconstruction of the top quarks due to the combinatorial problem of jet assignments to the source particles.
- **lepton+jets final state:** one W boson decays hadronically while the other decays leptonically. This final state offers the advantage of a single lepton with high momentum which can be used for the isolation of events against background. At the same time, the leptonic decay also reduces the combinatorics in the jet assignment of the reconstruction, especially with the additional identification of b-quark jets.
- **dilepton final state:** both W bosons decay leptonically into leptons and neutrinos. Although the final state has the lowest branching fraction it offers the best separation against background events as the two charged leptons can be identified very well. The problem of this final state is that the top quarks cannot be fully reconstructed as the two neutrinos escape the detectors undetected.

2.5 Single top-quark production in association with a W boson

2.5.1 Theoretical definition at next-to-leading order

As introduced in Section 2.4.3, the Wt-channel is the second most important contribution to single top-quark production at the LHC. This analysis aims at the isolation and cross-section measurement of the Wt-channel in the lepton+jets final state, which consists of a lepton and a neutrino, two light-quark jets and a b-quark jet. This final state is the same as the lepton+jets final state of top-quark pair production, except for one additional b quark. This fact already indicates that there could be overlaps between the two when looking at the contributing Feynman graphs which include higher-order corrections of QCD. At leading order, only the two graphs shown in Figure 2.7 contribute to the cross-section calculations. In

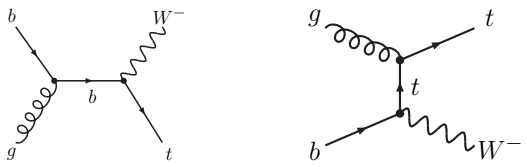


Figure 2.7: Feynman diagrams of the single top-quark production in the Wt-channel at leading order.

this case, the Wt-channel is well-defined and there is no overlap with the top-quark pair production. At next-to-leading order, the situation is different. Some examples of contributing Feynman diagrams are shown in Figure 2.8. These graphs can be also seen as leading-order production of top-quark pairs where

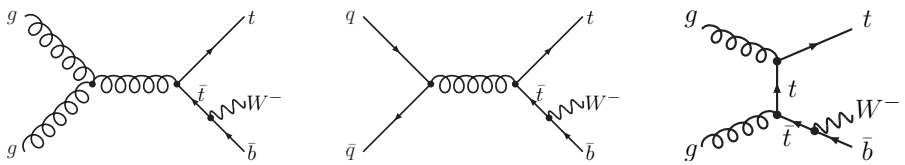


Figure 2.8: Examples of Feynman diagrams of the single top-quark production in the Wt-channel at next-to-leading order which include the production of top-quark pairs including the weak decay of one of the top quarks.

the antitop-quark decays into a W boson and a $\bar{b}$ quark if the invariant mass of the Wb-system gets close

to the top-quark mass. In that case, the intermediate top-quark propagator would be a real top quark. This, of course, raises some doubts about the clean definition of the Wt-channel at higher orders. The problem has been discussed extensively in the theory community and several possible variations have been proposed to circumvent the overlap [53–56]. The solution applied in this analysis was developed in [57] and implemented into the used event generator. It relies on two strategies:

- **Diagram removal (DR):** in this approach, all graphs containing a second top quark which can become real, like the ones in Figure 2.8, are removed.
- **Diagram subtraction (DS):** in the second approach a subtraction term is implemented which locally cancels the $t\bar{t}$ contribution when the second top-quark propagator becomes real.

Together with cuts which enhance the Wt-channel against the top-quark production, these two strategies ensure a valid definition at next-to-leading order. The proposed veto against a second b quark, which is further discussed in [58] is also applied in this analysis.

2.5.2 Experimental measurements

As the Wt-channel is largely suppressed in proton-antiproton collisions, the Tevatron experiments could not isolate it and previous colliders lacked the centre-of-mass energy to produce single top quarks. First attempts to measure the channel at the LHC were already taken early in 2011 and preliminary results were shown based on $0.7\,\text{fb}^{-1}$ by the ATLAS experiment [59] and based on $2.1\,\text{fb}^{-1}$ by the CMS collaboration [60]. Both analyses focussed on the final state with two charged leptons, which also was the target of the following ones. Analysing $2.05\,\text{fb}^{-1}$ of collision data, ATLAS presented first evidence for the existence of the Wt-channel with a significance of 3.3 standard deviations [61, 62]. The analysis measured a cross-section of $\sigma_{Wt} = 16.8 \pm 2.9\,(\text{stat.}) \pm 4.9(\text{syst.})\,\text{pb}$. The CMS collaboration also found evidence for the Wt-channel using the full $4.9\,\text{fb}^{-1}$ data of the 2011 LHC run [63]. The analysis could isolate a signal with $4.0\,\sigma$ significance and measured a cross-section of $16^{+5}_{-4}\,\text{pb}$. Both measurements are compatible with the theoretical cross-section prediction based on the Standard Model.

The observation of the Wt-channel with a signal significance of more than $5\,\sigma$ could only be achieved using the larger set of collision data from 2012 at a collision energy of $\sqrt{s} = 8\,\text{TeV}$. Analyses were published by both collaborations [64, 65] and a combination has been produced [66]. The results of these analyses and the combination are shown in Figure 2.9 while Figure 2.10 shows an overview of all cross-section measurements from the two largest LHC experiments. The latter shows the achieved precision in the first running period of the collider but also shows that there is still potential for higher precision measurements and the isolation of the s-channel.

2.5.3 Background processes

The presented analysis selects events with a measured event topology that corresponds to the expected final state of a lepton, a neutrino, two light-quark jets and a b-quark jet as described in Section 2.5. As the applied selection criteria can be fulfilled by events originating from different production processes, these also have to be taken account in the analysis. Such events are called *background events* and the corresponding production processes are organized into background channels according to their real gauge boson or heavy quark multiplicity. An overview of the cross-section of the main production channels in proton collision at energies in the TeV-range is shown in Figure 2.11. It illustrates nicely that several channels have cross-sections which are orders of magnitudes larger than the single top-quark production cross-sections. Hence, background events can have a large impact on the analysis and have to be simulated using Monte Carlo techniques as described later in Chapter 5. Studying the

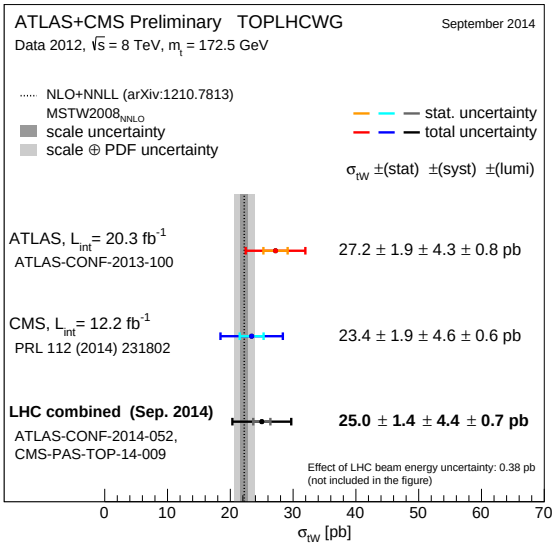


Figure 2.9: Cross-section measurements of the *Wt*-channel by the ATLAS and CMS collaborations and their combination at $\sqrt{s} = 8\,\text{TeV}$. The dashed line shows the theoretical prediction and its uncertainties as gray bands [66].

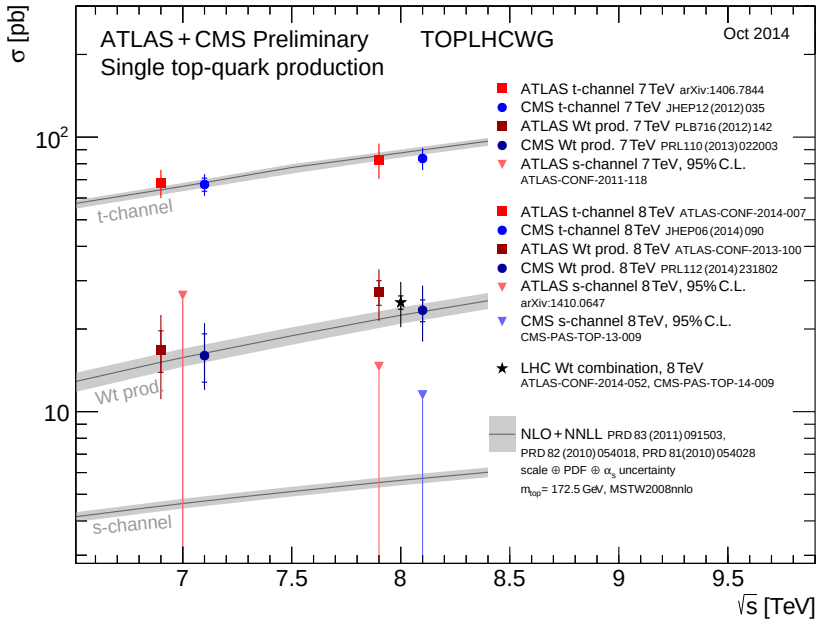


Figure 2.10: Summary of LHC measurements of the single top-quark production cross-sections as a function of $\sqrt{s}$ compared to a theoretical prediction. Only upper limits are given for the *s*-channel production [67].

properties of simulated background events can also help to reduce their contribution. A brief overview of the background channels to the analysed final state will be given in the following, including a short explanation and discussion on the possible suppression.

W+jets production The production of single W bosons in association with jets with different multiplicities has the second-largest cross-section of the studied background channels. Additionally it has a very high potential to result in a similar final state as the signal process. The produced W boson is also present in the signal process and can decay leptonically leading to a charged lepton and a neutrino with similar kinematics. Additionally, the accompanying jets may lead to the same event topology. This production channel is further separated according to the flavour of the jets that are produced into $Wc\bar{c}$, $Wb\bar{b}$, Wc and W+light-quark jets. The first three form the W+heavy flavour category (W+HF) while the last counts to the W+light flavour category (W+LF) in the analysis. As the production of jets during hadronisation of the quarks and gluons is a non-perturbative effect, it cannot be calculated analytically but has to be modelled. The uncertainty on the prediction of the cross-sections is thus large and rises with the number of jets that are produced. The suppression of the W+jets background is one of the main challenges of the analysis but it benefits from selection of three jets as the cross-section of this background falls steeply with increasing jet multiplicity.

Z+jets production According to Figure 2.11 the production of Z bosons already yields a considerably lower cross-section. Additionally, the Z boson mostly decays into two leptons in the events which enter the analysis and is thus largely suppressed by the veto on the second lepton. Only events where leptons are misreconstructed as jets can pass the event selection. Another good veto is the requirement for missing transverse momentum resulting of the neutrino, which is not present in large amounts in Z+jets events. The cross-section shows the same fall with increasing jet multiplicity as the W+jets background and therefore this channel only plays a minor role.

WW, ZZ and WZ production The production probability for two weak bosons per event is already smaller than the probability to produce top-quark events. Looking at the event topologies, the creation of two Z bosons is very different from the signal and, as is the case for Z+jets, not expected to have a large impact. In case of one W and one Z boson, the final state can be similar to W+jets if the latter decays hadronically. The production of two W bosons is similar to the signal but doesn't show the expected b-quark jet. In general, the three channels can be well separated and have a small contribution due to the lower cross-sections.

Top-quark pair production If a top and an antitop quark are produced in an event, the final state can be very similar to the associated production of a W boson and a top quark. This has been discussed already in Sections 2.4.3 and 2.5. Therefore, this background is the most problematic one, especially with a multiplicity of three jets per event. The primary way to suppress this background is the discrimination against a second b-quark jet, which is technically difficult. Other possibilities use event properties and the kinematics of reconstructed objects. This background's cross-section is also more than a factor of 10 larger than the signal cross-section, which results in a large event rate.

Single top-quark production Naturally, also the production of single top quarks in the t-channel and the s-channel have to be considered as backgrounds. s-channel production has a three times smaller cross-section than the signal and thus has minor impact. Additionally, the final state is expected to have two b-quark jets and no light-quark jets, which allows easy discrimination against signal events. On

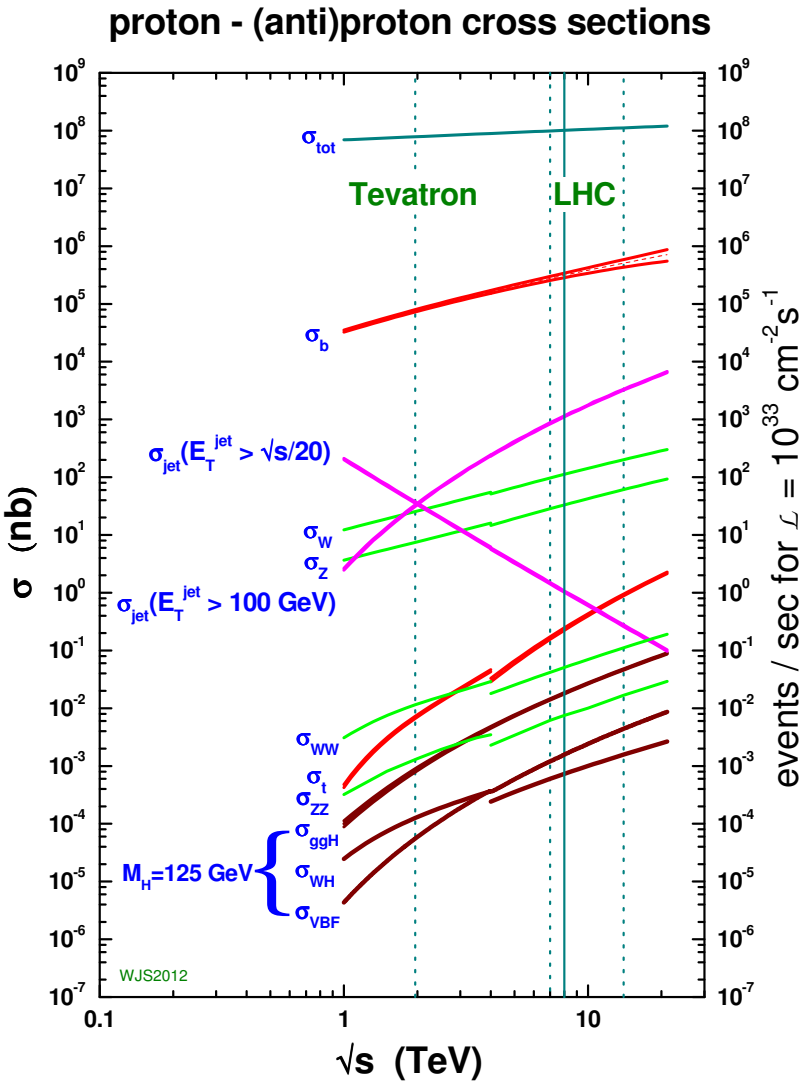


Figure 2.11: Cross-sections of production processes in proton-proton and proton-antiproton collisions in dependence of the collision energy $\sqrt{s}$. On the second vertical axis, the cross-sections are translated into the event rate at an instantaneous luminosity of $10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. The dashed vertical lines correspond to the 1.92 TeV collision energy at Tevatron, 7 TeV at the LHC, and the planned 13 TeV of the run 2. The solid line highlights the 8 TeV energy of the 2012 LHC run. The discontinuity of the cross-sections at 4 TeV marks the boundary between proton-proton and proton-antiproton collisions [68].

one hand, the production in the t-channel shows more similarity to the signal as only one b-quark jet is expected. On the other hand, only one W boson is present and at leading order, only one light-quark jet is expected which is mostly radiated into a more forward region of the detector, which is not covered by this analysis. Hence, the channel can be well suppressed and its cross-section is only a factor of four larger.

Multijet events The majority of the collisions in the LHC produce neither top quarks nor heavy weak bosons, but only result in different multiplicities of the jets in the detector. The cross-section of these so-called multijet events is close to σ_{tot} shown in Figure 2.11 and more than ten orders of magnitude larger than the signal cross-section. While in general multijet events are largely cancelled by the requirement of a single isolated charged lepton, the production rate is so large that events with jets only can still pass all cut requirements. Due to technical reasons during the reconstruction of particles in the detectors, jets can be mis-reconstructed as leptons and finally fake the expected signal topology. The production of these so-called multijet events cannot be simulated easily. Hence techniques have been developed which use the data to estimate this background. This is discussed in Section 6.2.

2.6 Kinematic variables for proton collisions

To describe, measure and calculate particle attributes in hadron collisions it is important to consider specific features of these events as explained in Section 2.4.3. As the colliding quarks and gluons carry a longitudinal momentum fraction, x_i, of the proton, their initial momentum is not known[8] thus applying a Lorentz boost of unknown size to the event parallel to the beam axis. This affects all outgoing momenta, $\vec{p}$. To acquire a variable independent from this boost in the z direction hadron collider experiments usually refer to the transverse momentum

$$p_{\text{T}} = \sqrt{p_x^2 + p_y^2}.$$

The same constraint is valid for the missing energy of an event. As the incoming partons are expected to have approximately no transverse momentum, the corresponding sum of all outgoing particles has to vanish as well due to momentum conservation. As the detector is not able to trace some particles[9] it is possible to calculate this missing transverse momentum, $E_{\text{T}}^{\text{miss}}$, and to use it for event selection.

Positions inside the detector are usually given in cylindrical coordinates (r, θ, ϕ) with r as the radial distance from the beam axis and ϕ as the azimuthal angle which describes the direction perpendicular to the beam axis. Instead of the polar angle, θ, between the flight direction of the particle and the beam axis, the third coordinate is given in terms of the pseudorapidity.

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

This variable has the advantage that differences in η are invariant under Lorentz boosts that especially occur in hadron collider events. A positive side effect of this representation is the steep rise of η towards small angles, i.e. towards the direction of the beam axis. The area around small values of $|\eta|$ contains the highest density of large transverse momentum objects, which are of central interest in most analyses. The finer granularity of η in this area allows an easier separation of reconstructed objects as it leads to flatter distributions.

[8] In e^+e^- collisions, the initial momentum is a known variable and useful for further calculations.
[9] Neutrinos pass the detector mostly without interaction.

After defining η it is convenient to give distances in the η-ϕ plane. The distances are then defined as

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

CHAPTER 3

The ATLAS Experiment at the LHC

This analysis has been performed within the ATLAS collaboration. ATLAS is one of the main experiments at the Large Hadron Collider (LHC) at CERN[10]. In this chapter the LHC and the ATLAS experiment are briefly described especially in context of the expected measurement performance.

3.1 The Large Hadron Collider

The Large Hadron Collider is a circular superconducting accelerator and collider built in the former LEP[11] tunnel [69, 70] at the Swiss-French border near the city of Geneva. The accelerator has a circumference of about 27 km and is located 50 to 175 m below the ground. It consists of eight arcs connected by eight straight sections of 528 m length, in four of which the beams cross and experiments are located. One straight section is used for the two RF systems[12] which accelerate the particles. Two interaction sections host the multi-purpose detectors ATLAS [71] and CMS [72]. The design luminosity at these two interaction points is $10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. The LHCb experiment [73] concentrates on bottom quark physics investigating quark mixing processes and searching for rare decays. In addition to protons, the LHC can also collide lead ions. A dedicated experiment called ALICE [74] studies lead-lead as well as proton-lead collisions.

Additional to these four large experiments there are three smaller ones which serve very special purposes but do not occupy dedicated straight sections. TOTEM[13] [75] measures the total cross section, elastic proton scattering and diffractive processes close to the CMS experiment and LHCf [76] is built to measure high energy neutral pions in the forward directions of the ATLAS detector for astroparticle physics purposes. The third small experiment, called MoEDAL[14] [77], searches for a magnetic monopole as well as other exotic particles and is built around the collision point of LHCb.

In the LHC ring protons or ions are accelerated in two independent beam-pipes with opposite magnetic dipole fields. To capture the beam from the pre-accelerator chain, accelerate it and store it at the projected peak energy of 7 TeV per proton, an RF system in the LHC of frequency 400.8 MHz is

[10] European Organisation for Nuclear Research; name derived from Conseil Européen pour la Recherche Nucléaire
[11] Large Electron Positron Collider (operation: 1989 to 2000)
[12] The radio frequency (RF) system converts high voltage pulses from the power supplies into radio frequency waves which are transported to the accelerator and used in the cavities to accelerate particles.
[13] Total Elastic and Diffractive Cross Section Measurement
[14] Monopole and Exotics Detector At the LHC

installed. The accelerator is designed store up to 2 808 bunches of up to 1×10^{11} protons for each proton beam with a bunch length of 1.06 ns and a minimum bunch spacing of 25 ns at maximum energy. Both the acceleration cavities and the guidance magnets of LHC use superconducting technologies. The dipoles are cooled down to a temperature of 1.9 K using superfluid helium to achieve a maximum central field strength of 8 to 8.5 T which is needed to bend the beams.

The proton beams are pre-accelerated by a chain of reused accelerators that were upgraded to meet the requirements of the LHC. An overview of this chain is given in Figure 3.1. The protons are produced in a duoplasmatron from hydrogen gas and injected into the Linac2 which accelerates the protons up to 50 MeV. In the following chain they are accelerated by the Proton Synchrotron Booster (PSB) to 1.5 GeV, by the Proton Synchrotron (PS) up to 25 GeV and finally by the Super Proton Synchrotron (SPS) to 450 GeV injection energy for the LHC ring. A detailed description of the LHC can be found in the LHC design report [78–80].

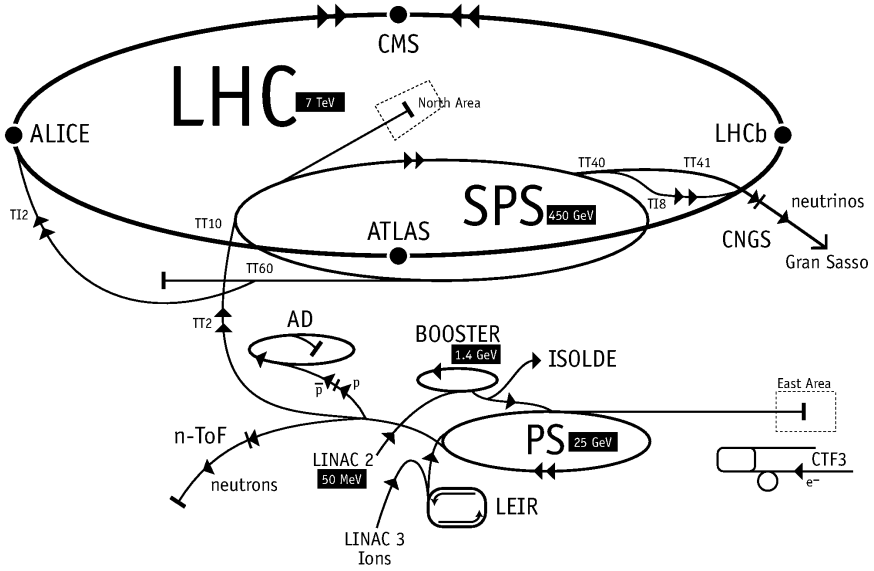


Figure 3.1: Illustration of the CERN accelerator complex including the LHC and its accelerator chain [81].

3.2 The ATLAS detector

"A Toroidal LHC ApparatuS" (ATLAS) is one of the two multi-purpose detectors at the LHC. Figure 3.2 shows an overview of the detector and its main components. The dimensions are about 44 m in length and 25 m in height and width. The main systems of the detector are the Inner Detector, the calorimeters, the muon and the magnet system. The following is a short overview of the important detector subsystems used for this analysis. Details are taken from the ATLAS detector paper [71].

3.2.1 Inner Detector

Surrounding the interaction point of ATLAS the Inner Detector (ID) is responsible for track reconstruction, momentum measurement, charge determination and vertex finding. The detector is surrounded by a solenoid magnet which generates a field of about 2 T to ensure sufficient bending of high momentum particle tracks. A pseudorapidity range of $|\eta| < 2.5$ is covered by the detector system. It consists of three

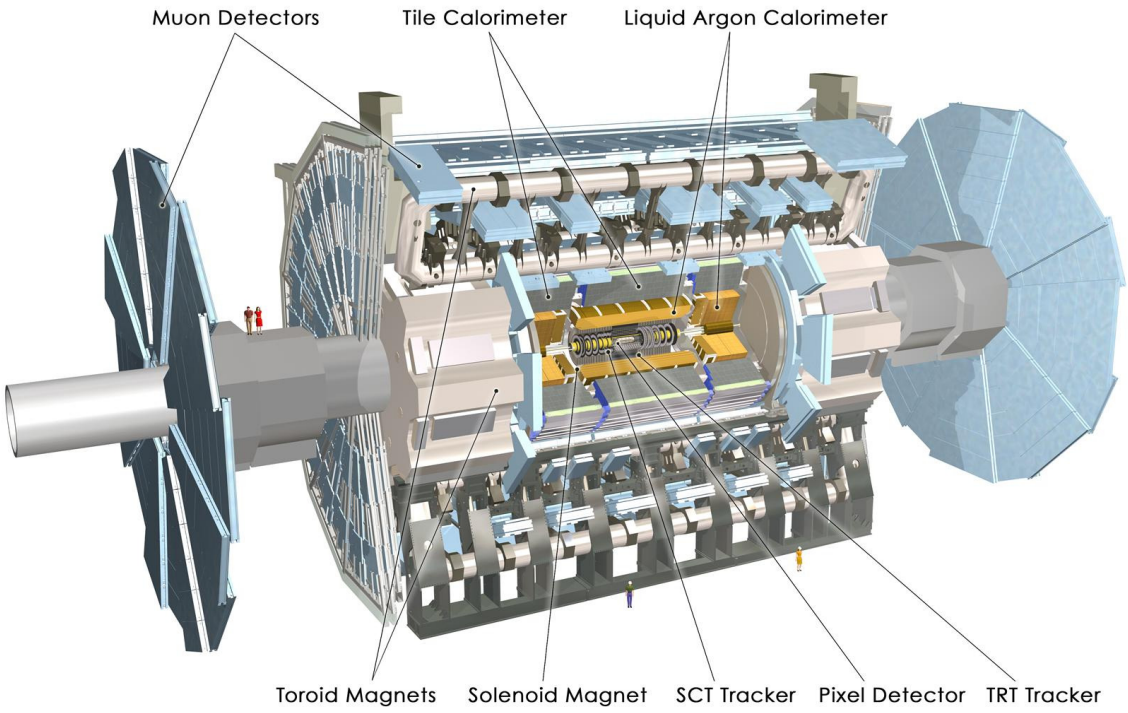


Figure 3.2: Computer-generated image of the ATLAS detector and its subsystems [82].

main components: the Pixel Detector, the Semiconductor Tracker (SCT) and the Transition Radiation Tracker (TRT). Figure 3.3 shows the overall setup of the ID and its subcomponents. The pseudorapidity-coverage of the different components and their dimensions can be read from the drawing in Figure 3.4.

The Pixel Detector is the component closest to the interaction point and the most precise tracking device in the ATLAS detector. It is built from silicon pixel sensors with a pixel size of $50 \times 400\ \mu m^2$. The sensors are placed on staves in three layers around the interaction point in the central region and on three disks in each of the forward regions. The innermost layer is only 50.5 mm away from the centre of the detector. Together the layers detect at least three points per track with very high intrinsic accuracy of 10 µm in $R - \varphi$ and 115 µm in z for the barrel and R for the disks in the forward directions.

The SCT is built around the Pixel Detector and uses silicon strip sensor technology. Each module consists of two sensors with 768 readout strips each with a constant 80 µm pitch that are rotated by 40 mrad against each other. The modules are combined into four layers mounted in the barrel region and nine disks on each side in the forward regions. The achieved intrinsic resolution is 17 µm in $R - \varphi$ and 580 µm in z/R. Together with the Pixel Detector the SCT provides an average of seven hits per track.

To increase the number of measured track positions substantially, the silicon detectors are enclosed by the TRT made from straw tubes in the pseudorapidity range of $|\eta| < 2.0$. These tubes have a diameter of 4 mm and are enclosed by 35 µm thick multi-layer films which act as cathodes supplied with high voltage. They are filled with a gas mixture of 70 % Xe, 27 % CO_2 and 3 % O_2. The signal is captured by a grounded anode wire in the middle and read out at one end[15]. In the barrel region the straws are mounted parallel to the beam pipe and have a length of 144 cm with a total of 73 straw planes. In the endcap regions the straw tubes are arranged radially in wheels with a length of 37 cm building up 160

[15] Anode readouts are located at both ends for the longer modules in the central region.

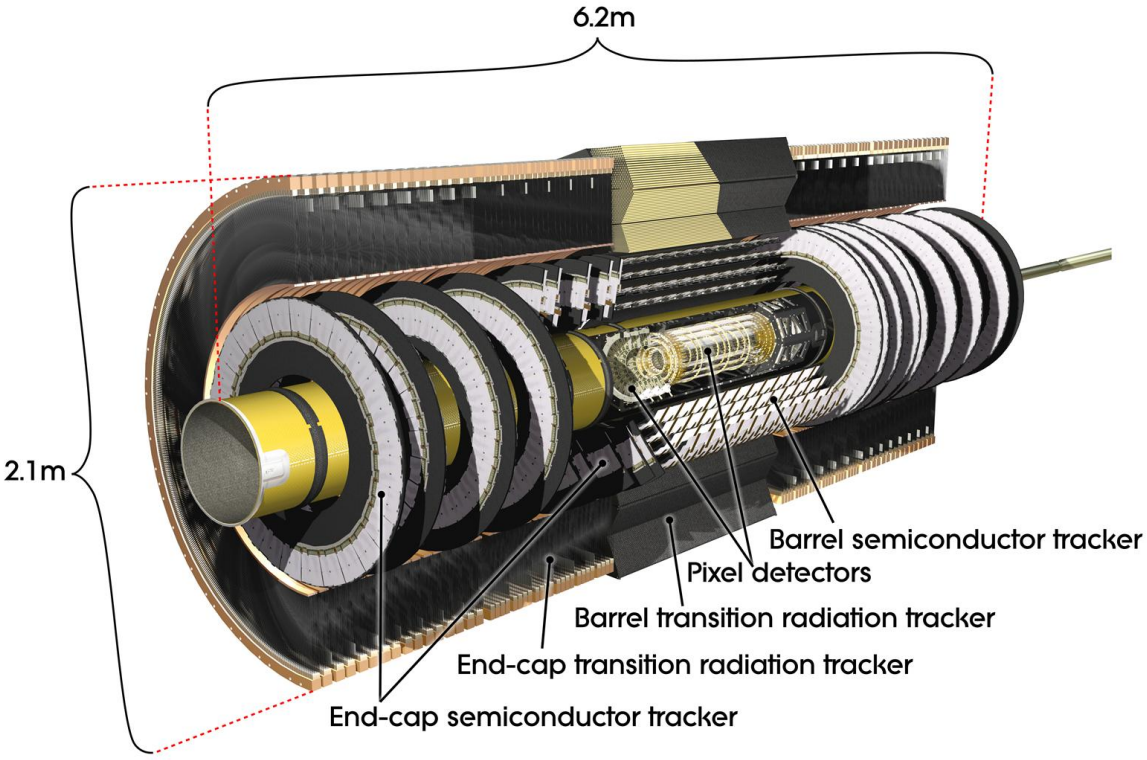


Figure 3.3: Computer-generated image of the ATLAS Inner Detector and its subsystems [82].

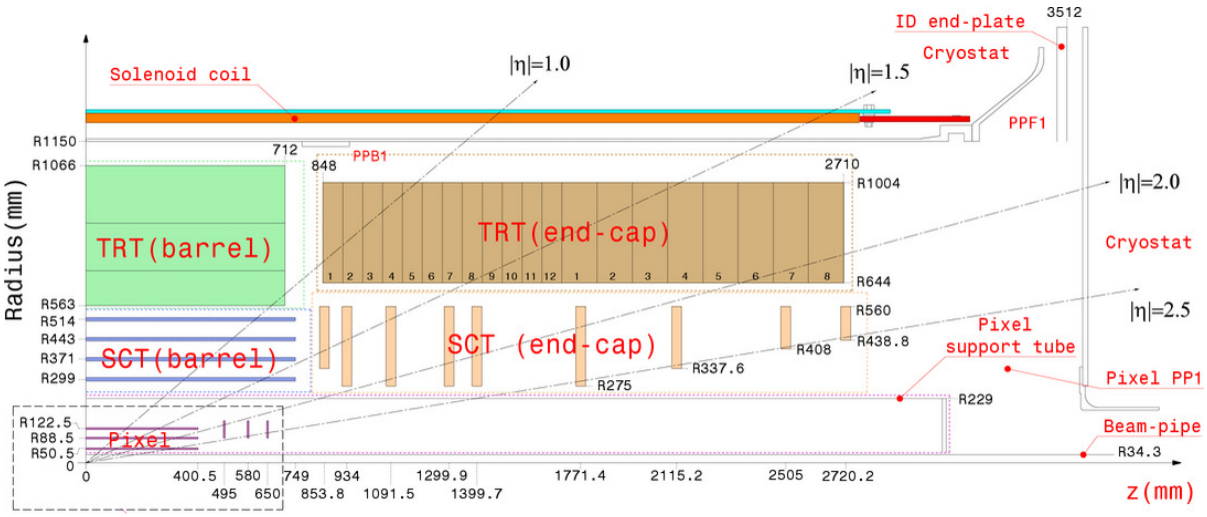


Figure 3.4: Technical drawing of the Inner Detector components' dimensions [83].

straw planes. Overall the TRT reaches an intrinsic accuracy of 130 µm in $R - \varphi$. It adds an average of 36 additional points per track. Apart from the improvement to the Inner Detector tracking, the TRT can also improve the particle identification, in particular of electrons. Plastic fibres (barrel) and foils (end-caps) are placed as radiators between the straws to increase the emittance of photons by transition radiation. These photons are absorbed in the gas and lead to larger signal amplitudes. Employing two different thresholds, the front-end electronics of the TRT can identify such photons. As mainly electrons generate transition radiation, this technique can be used to distinguish them from hadrons.

3.2.2 Calorimeter

To measure particle energies and complement the particle identification the calorimeter plays a major role in modern particle physics detectors. The ATLAS calorimeter is divided into two devices: the electromagnetic calorimeter (EM) that is optimized to register showers created by electrons, positrons and photons and the hadron calorimeter that is optimized for hadronic showers. Figure 3.5 shows a computer-generated overview of the main calorimeter parts. In ATLAS both calorimeter types are sampling calorimeters where passive material layers, which are responsible for most of the shower creation, alternate with layers of active material used for the measurement. Their depth is designed to capture the full electromagnetic and hadronic showers. The electromagnetic calorimeter covers more than 22 radiation lengths (X_0) in the central region and more than 24 X_0 in the forward regions. Hadronic showers are covered up to approximately 9.7 interaction lengths (λ) in the central region and 10 λ in the forward regions. Possible punch-throughs into the muon system are reduced by an additional 1.3 λ of support material outside the calorimeters.

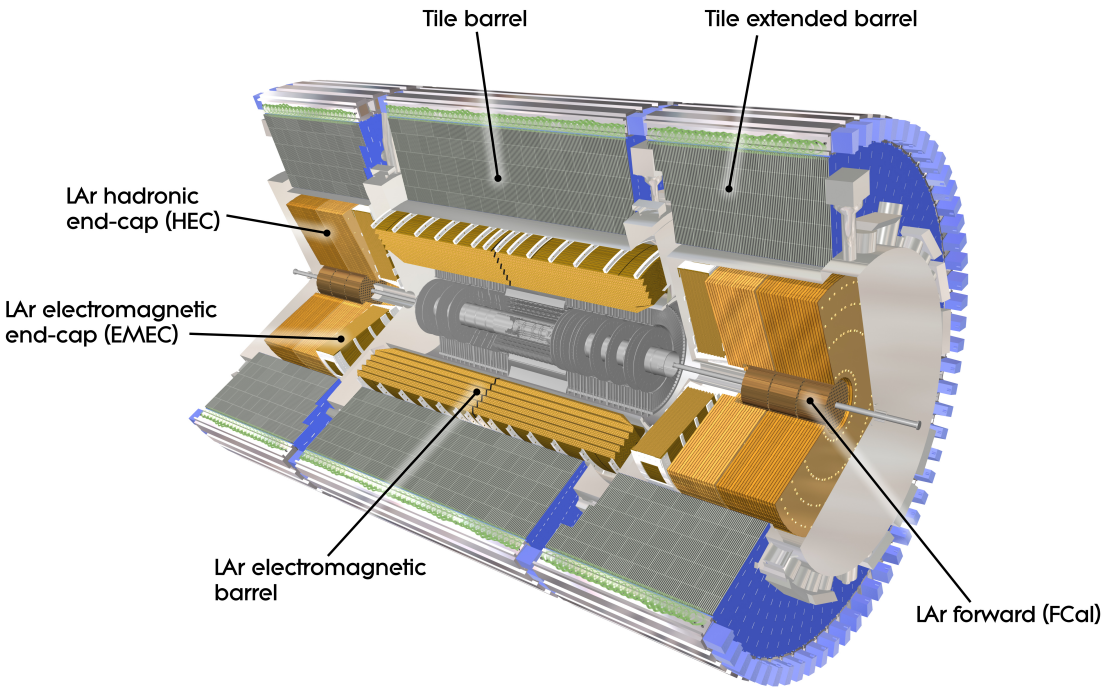


Figure 3.5: Computer-generated image of the ATLAS calorimeter systems [82].

Electromagnetic calorimeter

The EM calorimeter consists of a barrel part ($|\eta| < 1.475$) and two wheel sections in the forward directions ($1.375 < |\eta| < 3.2$). All the parts use liquid argon (LAr) as active detector material and accordion shaped lead plates as absorbers which provides full φ-coverage without any cracks. The readout is implemented with three layers of copper in between two absorbers. A common vacuum vessel contains both the solenoid magnet of the Inner Detector and the barrel of the electromagnetic calorimeter to minimize material in front of the latter. Two barrel parts are separated by a 4 mm gap at $z = 0$ while the end-caps consist of an inner and an outer wheel at both sides. Those components which cover the region of precision measurements ($|\eta| < 2.5$) have three layers and a higher granularity compared to the two layers in the forward regions. In addition to these, the region of $|\eta| < 1.8$ is also covered by a single LAr layer to correct for energy lost in front of the calorimeters (presampler). The calorimeter's granularity ranges from 0.0032 to 0.1 in $\Delta\eta$ and 0.025 to 0.1 in $\Delta\varphi$. Its energy resolution depends on $|\eta|$ as well as on the particle type. For electrons it is expected to reach [83]

$$\begin{aligned} \frac{\sigma(E)}{E} &= \frac{9.3\%}{\sqrt{E\,[\mathrm{GeV}]}} \oplus 0.56\,\% \qquad \text{for } |\eta| = 0.3 \text{ and} \\ \frac{\sigma(E)}{E} &= \frac{19.4\%}{\sqrt{E\,[\mathrm{GeV}]}} \oplus 0.43\,\% \qquad \text{for } |\eta| = 1.65. \end{aligned}$$

Hadron calorimeter

The hadron calorimeter is composed of three parts: the tile, the LAr hadron end-cap and the forward calorimeter. The tile calorimeter has a barrel part ($|\eta| < 1.0$) and two extensions which cover the range of $0.8 < |\eta| < 1.7$. Both consist of 3 layers with a cell granularity of 0.1 in both $\Delta\eta$ and $\Delta\varphi$ where the outermost layer cells are coarser with 0.2 in $\Delta\eta$. Steel is used as absorber and scintillating tiles, which are read out on two sides separately, are used as active material . The LAr calorimeters in the end-cap region are wheel-shaped and divided by copper plates, while liquid argon is again used as active material. They share the LAr cryostats with the EM calorimeter wheels and the forward calorimeter and cover a range of $1.5 < |\eta| < 3.2$. The two wheels of two layers hence overlap with the other two hadronic calorimeter parts, the tile calorimeter at low η and the forward calorimeter at high η. The forward calorimeter has a compound design made of copper plate for electromagnetic measurements and two modules made of tungsten. It is entirely immersed in LAr as active medium in the gaps of the metal web and the readout electrodes. In total the hadron calorimeter covers a pseudorapidity range of $|\eta| < 4.9$.

The resolution of the hadron calorimeter is expected to reach [83] typical values of about

$$\frac{\sigma(E)}{E} = \frac{60\%}{\sqrt{E}}\,[\mathrm{GeV}].$$

The missing transverse energy is expected to be measured with an accuracy of about

$$\sigma(E_{\mathrm{T}}^{\mathrm{miss}}) = (0.53 - 0.57) \times \sqrt{\sum (E_{\mathrm{T}} = 20 - 2000\,\mathrm{GeV})}.$$

3.2.3 Muon system

In order to measure the high momentum muons produced in LHC proton-proton collisions, strong magnets are required to bend their trajectories. The muon detector system of ATLAS is thus based on a

combination of three air-core superconducting magnet systems, two end-caps and one barrel toroid. The design goal is to measure muons with momenta up to 1 TeV with a resolution below 10 %. The barrel part covers a range of $|\eta| < 1.4$ and is built of eight racetrack-shaped coils enclosed in vacuum vessels with 25.3 m length and a diameter of 20.1 m. The barrel field is designed to provide a bending power of 1.5 to 5.5 T m. The end-cap toroids are likewise built out of eight coils each which are mounted together inside a large cryostat. They reach a bending power of 1 to 7.5 T m in the covered region of $1.6 < |\eta| < 2.7$. The transition regions are covered by superposition of both fields with lower bending power.

The precision momentum measurement of the muon tracks is accomplished by three cylindrical layers of drift chambers in the barrel region and three layers of chambers installed in the end-cap regions perpendicular to the beam. Most chambers are Monitored Drift Tubes[16] (MDTs) which have a diameter of 3 cm and are filled with 93 % Ar and 7 % CO_2 gas at 3 bar pressure. The muons passing through the straws ionise the gas and the resulting electrons are collected on a tungsten-rhenium wire at high voltage. The MDTs can achieve a resolution of 60 to 80 µm. In the high-η forward regions in front of the end-cap magnets Cathode-Strip Chambers (CSCs) are used. CSCs are multiwire proportional chambers with segmented cathode strips for the readout. The wires are spanned parallel to the central wire which is oriented in radial direction and the cathode stripes are built perpendicular to the wires on one side and parallel to the wires on the other side if a chamber. They achieve a resolution of 40 µm in the bending plane and 5 mm in the transverse plane. The lower transverse resolution is caused by the position of the readout system. With the help of an optical alignment system and dedicated algorithms it is possible to achieve a position resolution of 30 µm for a muon track.

The detector also includes a muon trigger system that covers a range of $|\eta| < 2.4$ and is built from Resistive Plate Chambers (RPC's) in the barrel and Thin Gap Chambers (TGC) in the end-cap regions. The trigger identifies bunch-crossings, provides p_T thresholds as well as position information in the direction orthogonal to the one measured by the precision-tracking system.

The overall muon p_T resolution reaches 4 % over a wide range of p_T (15 to 200 GeV) and differs over the η range varying between 3 and 4 % for $|\eta| < 1$ and about 9 % at $|\eta| \approx 1.5$.

3.2.4 Luminosity detectors

Additionally to the systems described above, two smaller detector subsystems are used for luminosity measurements in ATLAS. The LUCID[17] detector consists of sixteen aluminium tubes filled with C_4F_{10} gas mounted around the beam pipe at 17 m from the centre of the ATLAS detector. Charged particles from collision events create photons which are detected by photomultiplier tubes covering a range of $5.6 < |\eta| < 6.0$. While the Beam Conditions Monitor (BCM) has been mostly designed to trigger beam aborts in cases of augmented radiation in the Inner Detector, it can also be used for luminosity measurements complementary to LUCID. The BCM uses four stations of diamond sensors mounted to the Pixel Detector's support structure about 20 mm outside of the beam pipe at a distance of 184 cm from the interaction point. Both LUCID and BCM are fast detector systems which can provide a luminosity measurement for each bunch in the LHC filling scheme of 25 ns bunch spacing without dead time.

3.2.5 Trigger system

During the 2011 run of the LHC, 1380 bunches of the 2808 possible ones spread over the ring were collided with a minimum bunch spacing of 50 ns reaching a maximum instantaneous luminosity of up

[16] The MDTs are mounted on and between the toroid coils as well as in the forward wheels.
[17] Luminosity measurement using Cerenkov Integrating Detector

to $3.7 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$[84]. This filling scheme corresponds to an event rate of 20 MHz. Considering an average event size of 1.5 MB this would result in a throughput of 30 TB/s which cannot be handled by current computing equipment. To circumvent this problem a three-level trigger system is integrated into ATLAS which reduces the data amount based on different selections. The three consecutive levels of the trigger are called Level-1 (L1), Level-2 (L2) and the event filter (EF), where the two latter together form the high-level trigger (HLT).

The L1 trigger consists of the calorimeter trigger (L1Calo), the muon trigger and the Central Trigger Processor (CTP) as the core trigger component of ATLAS. All these components are implemented as custom electronics and thus not easily reconfigurable. L1Calo uses lower granularity information (larger than 0.1×0.1 in $\Delta\eta \times \Delta\varphi$) from the whole calorimeter as inputs and applies selections based on candidate signatures for electrons, photons, jets and hadronically decaying tau leptons as well as missing and total transverse energy. The muon trigger uses data from the trigger chambers mentioned in Section 3.2.3. Both triggers identify areas with relevant signatures and mark them as Regions-of-Interest (RoI). The CTP combines the information from the two single triggers and calculates final L1 trigger decisions for all the defined triggers. It is also possible to configure prescale factors to only read fractions of events which pass a trigger. This way the L1 trigger reduces the initial event rate within a maximum decision latency below 2.5 µs down to about 60 kHz. It is limited at 60 kHz by dead time from the detectors.

Together with the clock signal from the LHC the CTP distributes the final L1 trigger decision back to the front-end electronics of each detector component (Trigger, Timing and Control (TTC) partition) and the RoI information to the L2 trigger. The L2 trigger farm reads the event data within the RoIs from the read-out system buffers of all detector components (RoI fragments) and calculates the selection result for all L2 items. With the increased granularity and precision of the RoI fragments, the L2 reduces the event rate down to 5 kHz with a maximum processing time of 40 ms. The maximum event rate is limited by the detector read-out system (6 kHz) and the L2 CPU (7 kHz).

Based on L2 decisions, the full event information is read out from the read-out system by the Event Builder and finally passed to the EF which performs a final selection and triggers the final storage of the event data. The EF decision calculation takes about 5 s per event and reduces the event rate to 400 Hz on average, which corresponds to a data rate of about 600 MB/s. These rates are mostly capped by the CERN computing facility CPU capacity. The HLT is fully implemented on industrial standard computing components and thus allows later adaptation to analysis needs.

A typical distribution of the trigger output event rates for all three triggers is shown in Figure 3.6. The distributions show a general falling behaviour which is explained by the drop of instantaneous luminosity over one LHC run. The rising steps reflect the adjustment of the trigger prescales during the run to maximize the amount of recorded events.

Based on the EF trigger decisions, events are written to different data streams. An overview of the rates of the main data streams for physics analyses over the 2011 run period is shown in Figure 3.7. The two important streams for this analysis are “Muons”, which contains events passing any muon trigger, and “Egamma”, which holds events with either electron or photon signatures.

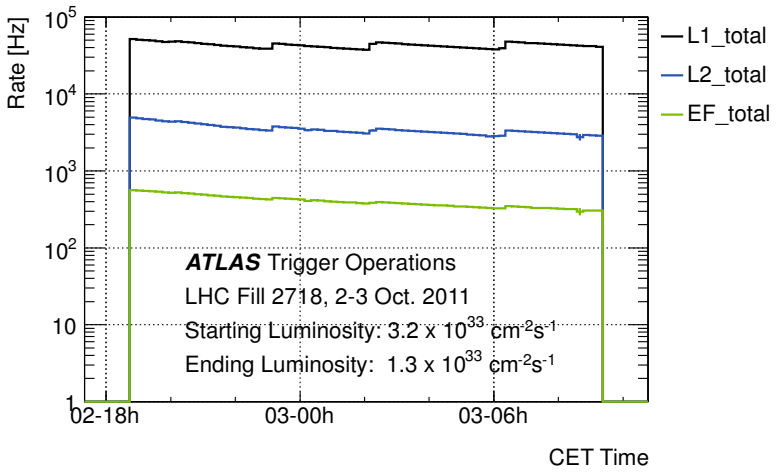


Figure 3.6: Data trigger output and recording rate of ATLAS at a luminosity of 3.2×10^{32} cm^{-2} s^{-1} (Internal: LHC Fill 2178) [85].

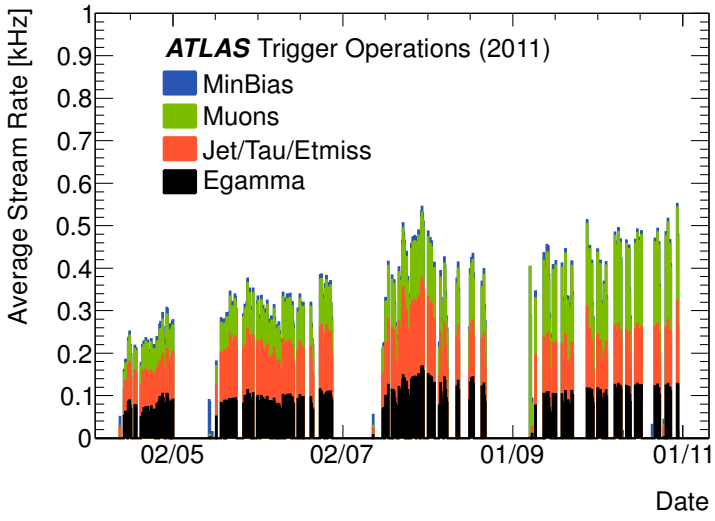


Figure 3.7: Event Filter stream recording rates from 2011. Filtered for LHC stable beams and ATLAS ready [85].

CHAPTER 4

Particle Reconstruction

To analyse the data from the ATLAS detector it is necessary to map it to physical observables like particles and their properties. This step is executed centrally in ATLAS by the reconstruction algorithms. The algorithms use the detector signals as inputs to identify particle types and their properties, such as momentum, energy, mass and lifetime. Not all particles can be measured directly by the detector for different reasons like very short lifetimes, weak interaction with material or no electric charge. In such cases, algorithms use approximations or indirect methods. This chapter describes the identification and reconstruction of particles which are important for the analysis and shows briefly the achieved performance. It also summarises the additional selection requirements on object level used in this analysis. At the beginning of the chapter, the basic technique of tracking is described in Section 4.1. The rest of the chapter covers the reconstruction of electrons (Section 4.2), muons (Section 4.3) and jets (Section 4.4) and the cuts applied to them.

4.1 Track and vertex reconstruction

4.1.1 Tracking

Only a brief overview of the track reconstruction techniques used in ATLAS will be given here, based on the detailed description of the underlying software system in [86] and the performance study of the 2011 run [87]. "Track reconstruction deals with all aspects of determining the trajectories of the particles and estimating their kinematic properties" [3]. The main tracking device of ATLAS is the Inner Detector with its three sub-detectors Pixel, SCT and TRT. To improve the performance of the measurement of muon kinematics, the Muon System improves the tracking of muons especially in the high momentum regime with its strong magnetic field and the longer lever arm. The modular technical implementation of the tracking procedures in ATLAS allows inputs from both systems to be processed with the same algorithms.

Track-fitting algorithms have to find hits in the tracking detectors that belong to one track and fit the track parametrization to those hits. Three-dimensional space points are determined from the hits in the silicon trackers to server as track seeds. Space points from the Pixel detector are the centres of clusters from the raw data. The SCT strips do not allow a measurement of three coordinates. Therefore the intersection of two strips on each side of a module is calculated based on a beam spot constraint and used as a space point. Three space points from the two detectors compatible with a track form a seed.

In a second stage an inside-out sequence is applied to reconstruct the full track. A combinatorial Kalman filter [88, 89] is used to fit the clusters belonging to the seeds' space points and also to combine more hits into the track. Once tracks from the silicon detectors are formed, a scoring system, based on e.g. the number of hits, the number of missing hits (holes) and the detector parts, is employed to reduce the number of fake tracks and to resolve ambiguities. Taking into account the candidate's ranking, the Kalman filter extrapolates the track into the TRT and adds TRT hits to the track fit. The TRT in general is able to improve the momentum measurement as it adds a large number of hits elongating the measured track. If a lower score of the track candidate is reached by this procedure, the TRT hits are only associated to the track as outliers and do not contribute to the fit.

After the inside-out sequence, space points can remain unused. In this case drift circles from TRT straws are used to create seeds for a Kalman filter applied from the outside to the inside of the detector. This procedure, named *back-tracking*, can help to recover ambiguities and also improves the reconstruction of tracks from particles that were not created in the primary collision process as well as long-lived particles and particles from conversions. As mentioned above, the same algorithms can also be applied to the data of the muon system, extending the tracks through the whole detector and adding segments from the muon detectors to create combined tracks.

4.1.2 Vertexing

With the help of precisely measured tracks it is possible to search for their origin in the detector, which is called a vertex. Vertices can be produced at the collision points of protons in the ATLAS detector, originate from unstable particle decays or result from particle-matter interactions. Vertices from collision points are called *primary vertices* while the latter two cases are usually referred to as *secondary vertices*. The search algorithms for primary vertices use an iterative procedure to associate tracks to vertex points. The algorithm and its performance is described in [90]. In LHC collisions at high luminosity a multitude of primary vertices are found as a result of pileup. They can be used to identify objects which do not result from the primary collision of interest and treat them accordingly.

The reconstruction of secondary vertices plays a major role in the identification of particles. They offer information about the lifetime of a particle and allows their decay products to be tagged. A very important application is the tagging of quark flavour of jets, which is described in Section 4.5.

4.1.3 Alignment

To achieve the highest possible resolutions of track and vertex reconstruction, the exact positions of the detector elements must be known. Despite highest engineering precision during the assembly, geometrical deviations in module positioning and deformations cannot be prevented. Hardware solutions are built into the muon spectrometer[18] and the SCT[19]. Software alignment techniques are used to compensate for small effects to reach the ultimate tracking accuracy. They make use of cosmic ray events as well as collision tracks. Details of these complicated algorithms and the resulting performance can be found in [92, 93].

4.2 Electron reconstruction

The reconstruction of electrons is subdivided into several steps: calorimeter cluster measurement, track-cluster matching and finally the electron identification. This section will give a brief overview of the

[18] Three optical systems are used in the Muon System to align most chambers [71].
[19] Frequency-scanning interferometry is used to determine changes of the SCT module positions [91].

process following the information given in [94]. Several different algorithms are used to reconstruct electrons in ATLAS, but only the reconstruction of standard central electrons ($|\eta| < 2.47$) will be detailed here as only these objects are used in the further analysis.

4.2.1 Calorimeter-based reconstruction

To measure electron energies in the central region of the detector, reconstructed tracks from the Inner Detector are matched to energy clusters in the EM calorimeter. In a first step calorimeter towers are created, summing up the measured energy of all cells in a certain piece of the detector which is defined by the granularity (as given in Section 3.2.2). Afterwards a sliding window algorithm is applied to identify towers with a measured transverse energy of $E_{\mathrm{T}} > 2.5\,\mathrm{GeV}$ in a window of size 3×5 in units of tower sizes in $\eta \times \phi$ which serve as seeds for electron clusters. All reconstructed tracks from the Inner Detector are extrapolated into the middle layer of the EM calorimeter and matched to the cluster seeds if they overlap within a distance of 0.05 in $\Delta\eta$ and $\Delta\phi$. Bremsstrahlung losses can lead to a larger bending of the track during the traversal of the solenoid field. To preserve the matching efficiency the window in the bending direction is enlarged to $\Delta\phi < 0.1$. In case of multiple matched tracks to one cluster seed, only the closest track in ΔR is kept.

To add energy that is possibly deposited outside of the sliding window, the cluster energies are re-calculated using the towers in an enlarged range of 3×7 in the EM barrel region and 5×5 in the end-caps. The cluster energy is then determined based on a detailed simulation of the energy deposit in the calorimeter material as well as the expected deposits outside the active detector region. The energy resolution depends on a very accurate description by the Monte Carlo simulation both of the calorimeter material and the material that electrons pass before reaching it. Figure 4.1 shows this material as a function of the pseudorapidity, where peaks show up at the boundaries of the subsystems.

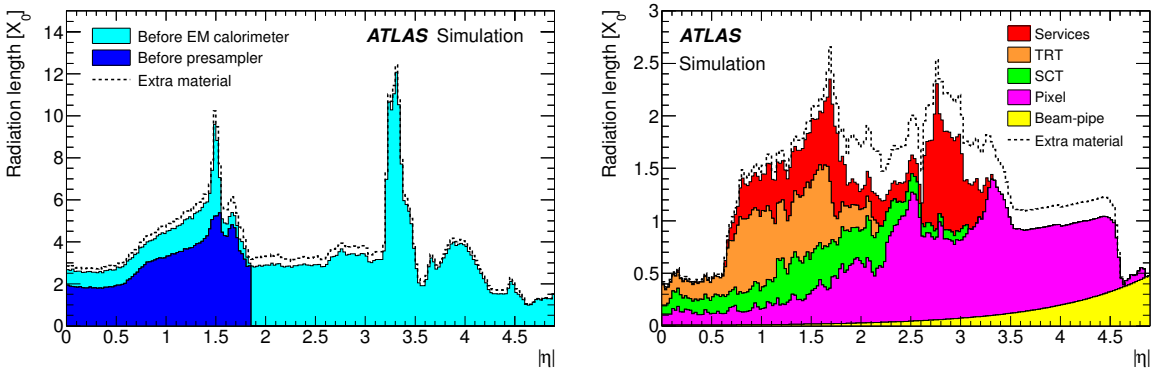


Figure 4.1: Amount of material, in units of radiation length X_0, traversed by a particle as a function of η: (left) material in front of the presampler detector and the EM calorimeter, and (right) material up to the ID boundaries. The contributions of the different detector elements, including the services and thermal enclosures are shown separately by filled colour areas. The extra material used for systematic studies is indicated by dashed lines. The primary vertex position has been smeared along the beam line [94]

.

The electron candidate is created using the energy measured from the calorimeter and the directions (η and ϕ) from the associated inner detector track. Transverse energies are calculated with

$$E_{\mathrm{T}} = \frac{E^{\mathrm{cluster}}}{\cosh(\eta^{\mathrm{track}})}.$$

The trigger and reconstruction efficiencies are measured in $Z \to ee$ decays using a tag-and-probe method in both collision data and simulated events. Differences are corrected by the application of scale factors to the event based on the electron's properties.

4.2.2 Electron identification

True electrons are reconstructed with a very high efficiency with above method but the algorithms also reconstruct background electrons which result mainly from photon conversions and Dalitz decays, as well as electron candidates from jets with an electron-faking signature in the detector. For the purpose of this analysis all of the three classes are background electrons, which are further suppressed after the reconstruction step using cuts on calorimeter and tracking variables. Three sets of cuts, sorted by increasing rejection of fake electrons but decreasing reconstruction efficiency, are defined as standard for central electrons.

Loose Cuts are applied on shower shape variables of the EM middle layer and variables sensitive to energy leaks into the hadron calorimeter. The electron candidates are also required to be in the central region of the detector ($|\eta| < 2.47$).

Medium Shower shape variables of the first EM calorimeter layer, track quality requirements and a more precise track-cluster matching are used to increase the rejection of fake electrons.

Tight Additional track quality requirements, the closest match of tracks to clusters, the TRT electron identification capabilities and the ratio of the cluster energy and the track momentum are used to provide the cleanest sample of real electrons. Background electrons from photon conversions are rejected by a vetoing candidates, which also correspond to a reconstructed photon.

The values of the cuts using above variables are stored inside a menu. Both values and variable composition have evolved over the time of the LHC data taking. They have been adjusted to the higher pile-up conditions which arise from the increase of luminosity. The current menu is distinguished from the previous by attaching "++" to the names of the standard set of cuts as listed above. For the analysis, electrons selected using the "Tight++" criteria are used.

An additional important discriminating property of the electron candidate against background electrons is the isolation. Misreconstructed jets, leptons from semileptonic heavy quark decays and electrons originating from pile-up events often appear in conjunction with other objects, like jets, that deposit energy in the calorimeter close to the electron candidate's cluster. To suppress such electron candidates, the transverse energy in a cone around the candidate is calculated with

$$E_{\mathrm{T}}^{\mathrm{cone}}(R_0) = \left(\sum_{\substack{\mathrm{cells} \\ \Delta R < R_0}} E_{\mathrm{T}}^{\mathrm{cell}} \right) - E_{\mathrm{T}}(5 \times 7 \text{ cells}),$$

where R_0 is the cone's radius. The subtrahend accounts for the energy of the electron candidate leaving $E_{\mathrm{T}}^{\mathrm{cone}}$ as the energy around it. A correction is applied to account for leakage of deposited energy of the electron into the cone and for additional energy from objects originating from pile-up events. A similar

variable can be constructed using tracking information to identify close-by tracks:

$$p_{\mathrm{T}}^{\mathrm{cone}} = \sum_{\substack{\text{good tracks} \neq e \\ \Delta R < R_0}} p_{\mathrm{T}}^{\mathrm{track}}.$$

Isolated leptons of interest can be accompanied by low-quality tracks which would diminish the effect of the separator and create a dependence on the number of pile-up events. A selection of good tracks in the calculation of the variable thus recovers its effect.

In this analysis cone radii of $R_0 = 0.2$ for $E_{\mathrm{T}}^{\mathrm{cone}}$ and $R_0 = 0.3$ for $p_{\mathrm{T}}^{\mathrm{cone}}$ are chosen. The cut values on the deposited energies in these cones are not fixed but dynamic to achieve a selection efficiency of 90 % over the full range of electron E_{T} and η. The efficiency is derived using a tag and probe method on $Z \to e^+e^-$ events. To account for efficiency differences in collision and Monte Carlo events, scale factors are applied to the simulated events.

4.2.3 Energy scale and resolution

Decays of Z bosons and J/Ψ mesons are used to measure the energy scale of electrons. Additional studies exploit the ratio of the calorimeter energy to the track momentum in decays of W bosons. According results are used to correct the electron energy in collision data. The energy scale uncertainty varies from 0.3 to 1.6 %. To measure the energy resolution, the mass spectrum of the two electrons from Z decays has been fitted in collision data and been compared to Monte Carlo simulation. The energy resolution is parametrized as

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

Due to limited statistics, the first term (sampling term) has been fixed to the simulated resolution with a 10 % uncertainty. The noise term, proportional to b, has also been fixed to the one from simulation justified by adjustments of the Monte Carlo based on calibration data. Only the constant third term becomes dominant at higher energies and has been extracted from the fits. Its value is 1.2 % in the calorimeter barrel and 1.8 % in the end-cap region with $|\eta| < 2.47$. In the analysis, electron energies in the Monte Carlo events are smeared to match the measured resolution in collision data. Additional calibration corrections are applied to electrons in the transition regions of the calorimeter ($1.42 < |\eta| < 1.55$) to account for the detector inefficiency.

4.2.4 Electron selection

The final electrons used in the analysis are required to have a transverse momentum $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and to be in the central region of the detector ($|\eta| < 2.47$). Additionally, electrons are removed from the inefficient transition region of the calorimeter ($1.37 < |\eta| < 1.52$). For both cuts, the pseudorapidity of the electron's calorimeter cluster is used. As the ATLAS reconstruction can identify the same signal as two different particle types, creating ambiguities, an overlap procedure has to be used to disentangle these. After applying the overlap removal as described in Section 4.4.2 electrons are again checked for overlap with jets and removed if a jet is found in the vicinity ($\Delta R < 0.4$).

4.3 Muon reconstruction

Unlike for the detection of electrons, the muon energy cannot be measured by the calorimeter system as they behave like minimum ionizing particles and lose only little energy passing through matter. Muon

properties can be determined using the tracking capabilities of the Inner Detector and the muon system. The reconstruction of muons used in this analysis uses one of the available algorithm chains in ATLAS, named MuID or Chain 2. The overview given hereafter follows the algorithm description in [95] and the more recent identification notes [96, 97].

In a first step, a pattern recognition is used to identify hits to be used as seeds for track segments in each chamber of the muon system. Tracks are then constructed from the found segments starting from the outermost layers of the muon system to the inner ones until they reach the innermost stations. The found tracks are finally refitted over the full range including all nearby hits into the fit. In a final step, inner detector tracks compatible with the muon standalone ones are searched for and a global track refit is executed. To improve the efficiency of the muon finding, another algorithm (MuGirl) uses Inner Detector tracks as seeds and reconstructs muons searching for segments and tracks, reconstructed by the muon system only, in the extrapolation region of the Inner Detector track. An overlap of the candidates found by the two algorithms is accounted for.

At analysis level, additional requirements have to be fulfilled by the created muon candidates:

- Only combined muons, which are built from Inner Detector tracks and muon system information, are used.
- The muon candidate must be found in the central detector region ($|\eta| < 2.5$) and its transverse momentum has to be larger than 20 GeV.
- The muon track has to match the following track quality criteria:
  - at least one hit in the innermost layer of the pixel if the crossed pixel was active;
  - number of pixel hits + number of crossed dead pixel sensors > 1;
  - number of SCT hits + number of crossed dead SCT sensors > 5;
  - number of crossed active pixels and SCT modules which did not register a hit < 2;
  - in the TRT barrel region ($|\eta| < 1.9$), the total number of TRT track points (TRT hits and outliers) should be larger than five and the fraction of outliers should not exceed 90 %;
  - in the TRT end-caps, the requirement on the fraction of outliers is only applied for more than five track points, while higher fractions are also allowed for tracks with less than five track points;
- The muon track needs to be compatible with the primary vertex ($|z_0^{PV}| < 2$ mm).

To suppress muons from secondary decays, isolation criteria are required for muons as well. As an efficiency-dependent cut procedure was not available for muons yet, fixed cuts are used:

- $E_{\mathrm{T}}^{\mathrm{cone}} < 2.5$ GeV in a radius $R_0 = 0.3$ and
- $p_{\mathrm{T}}^{\mathrm{cone}} < 4.0$ GeV in a radius $R_0 = 0.2$.

In the same way as for electrons and jets, overlap of muons with other objects has to be handled. High-p_{T} muons can be created in jets from semileptonic decays of hadrons. Muons are therefore removed if a jet selected as described in Section 4.4 is found in the vicinity of a muon ($\Delta R(\mu, \mathrm{jet}) < 0.4$).

4.3.1 Muon reconstruction efficiencies and momentum resolution

Reconstruction and isolation efficiency measurements have been made using a tag-and-probe method on $Z \to \mu\mu$ events and results can be found in [97]. The overall muon efficiency for MuID muons is measured to be above 96 %. To correct for the small efficiency differences, scale factors are extracted for the reconstruction, the trigger and the isolation efficiencies in the analysis.

The resolution of the transverse muon momentum has been evaluated in [98]. As muons are combined using Inner Detector tracks and tracks reconstructed in the muon system, the overall resolution depends on the weighting of the two parts. Therefore, resolutions are measured separately for the Inner Detector and the muon system and as a function of the muon's transverse momentum. Measurements are additionally carried out separately in four regions of pseudorapidity. The results show resolutions of the muon system ranging from 4 to 9 % for 25 GeV muons to 6 to 17 % for muons with 200 GeV. The resolution in the Inner Detector is better for muons with low transverse momentum than in the muon system, ranging from 2 to 5 % at 25 GeV in the regions of $|\eta| < 2$ to 10 to 28 % at 200 GeV muon transverse momentum. The resolution in the region $|\eta| > 2.0$ degrades rapidly which results from worse coverage, e.g. by the TRT. In general, the precision is dominated by the Inner Detector measurement at low transverse momenta, while at high p_{T} it benefits from the muon system's distance to the interaction point and the strong toroid field which bends the muon track. The referenced analysis shows that the measured momentum resolution is still less precise than the 4 to 10 % projected in the design but the used algorithms, calibration and alignment has also been improved since. In the presented analysis the momentum scale of muon in Monte-Carlo data is adjusted and the momenta are smeared according to the results of above study which was updated to the 2011 data but has not been published.

4.4 Jet reconstruction

Quarks with high momentum created in the proton collisions hadronise leaving a collimated shower of hadrons, electrons and photons in the calorimeters. Such structures are called jets and allow the initial properties of the particle to be deduced. The reconstruction and energy calibration of jets in ATLAS is described in [83, 99, 100] and summarized in the following.

In a first step, calorimeter cells are grouped together. Cells with an energy deposit well above the background noise (electronic and pile-up related) are used as seeds and neighbouring cells are added to form a cluster if their energy deposit exceeds a certain threshold. The final topological cluster is then formed adding all nearest-neighbour cells together. The four-vector of such a cluster is defined by the energy sum of all grouped cells and its direction is defined by the barycentre of the energies with respect to the centre of the ATLAS coordinate system. If multiple local energy maxima are found within one cluster, the cluster can be subdivided.

After processing all cells with the above algorithm and finding possible clusters, the anti-k_{T} algorithm [101] is used to merge clusters into jets. It defines a distance between two clusters i and j as

$$d_{ij} = \min\left(\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2}\right)\frac{\Delta_{ij}^2}{R^2}$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ with the rapidity y and the distance parameter R, which is set to 0.4 in this analysis. For an object i the distance to the beam is defined as $d_{iB} = 1/p_{T,i}^2$. The algorithm searches for the smallest distance d_{ij} between two clusters and merges them if $i \neq j$ until the smallest distance found corresponds to d_{iB}. In that case the final merged object i is added to the list of jets and

all clusters associated with it are removed. The algorithm continues to find objects to merge until no further topological clusters remain.

4.4.1 Jet energy calibration

The reconstructed objects are jets calibrated at the electromagnetic (EM) energy scale, which scales the energy deposited by electromagnetic showers correctly to electron and photon energies. At this scale, a correction is applied to the jets that accounts for pile-up effects as described in [102]. An additional correction is applied to the jet's direction assuming it to originate from the primary vertex position. To bring the jet energy to the particle level, it is calibrated using factors which are derived from simulation data. The corresponding correction factors are parametrized in bins of p_{T} and η. Figure 4.2 shows the inverse of these correction factors as a function of the pseudorapidity.

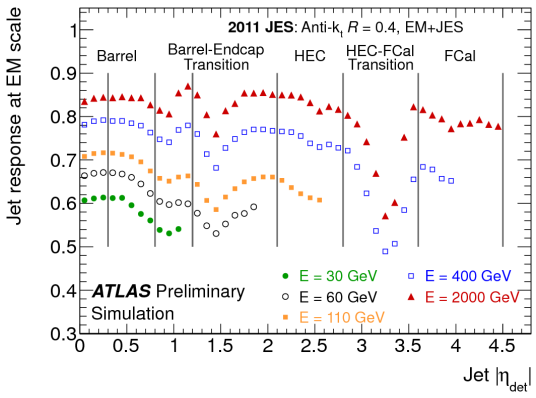


Figure 4.2: Average energy of jets calibrated at EM scale with respect to the truth jet energy ($E^{\mathrm{EM}}_{\mathrm{jet}}/E^{\mathrm{truth}}_{\mathrm{jet}}$) as function of η_{det} of the jet before applying the primary vertex correction for different jet energies E. The main regions of the calorimeter are indicated. Results are based on an inclusive jet sample generated with the Pythia generator [99].

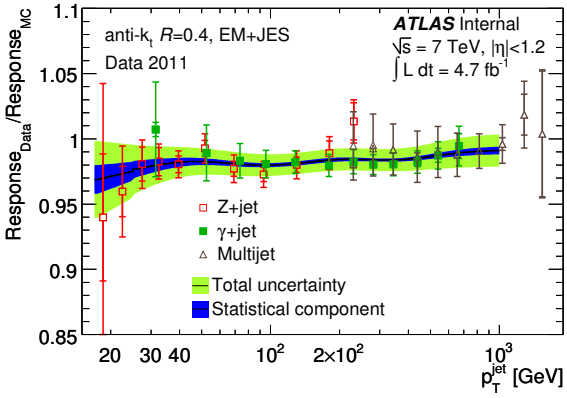


Figure 4.3: Ratio of the average jet response, $\langle p_{\mathrm{T}}^{\mathrm{jet}}/p_{\mathrm{T}}^{\mathrm{ref}}\rangle$, measured in collision data to that measured in simulation for jets within $|\eta| < 1.2$ as a function of p_{T}. The ratios are shown separately for the three *in situ* techniques used in the combined calibration. The error bars indicated the statistical and total uncertainties (statistical and systematic uncertainties added in quadrature). The error bands indicate the statistical and total uncertainties on the combination [99].

A final correction of the jet energies is based on differences observed in data/simulation comparisons. These final corrections are determined using three different *in situ* methods:

- In a first step, the p_{T} balance between central and forward jets in dijet events is used to correct the average p_{T} of jets in the forward region ($0.8 \leq |\eta| \leq 4.5$) using central jets ($\eta \leq 0.8$).
- In a second step, jets in the pseudorapidity region $|\eta| < 1.2$ are corrected for their jet energy scale based on corrections deduced from balancing photons and Z bosons against jets in the same region. The produced correction factors are also used to correct jets in the more forward regions ($\eta \geq 1.2$).
- After applying the two first steps, systems of recoiling low-p_{T} jets against high-p_{T} jets are used to calibrate very high energy jets.

The effect of the combination of these methods and its uncertainty can be seen on the ratio of responses of jets in data and simulation in Figure 4.3. The combined correction is applied to data events.

4.4.2 Jet selection

Only central jets with $|\eta| < 2.5$ are used in this analysis. Additionally a transverse momentum of $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ is required.

The jet reconstruction described in Section 4.4 still identifies jet candidates that stem from background sources like calorimeter noise, beam-gas and beam-halo effects in the LHC accelerator and energy deposits from cosmic muons. To eliminate this background, typical properties of such jet candidates have been found and default selections have been defined for analyses using jets. This analysis uses the "loose" selection as defined in [103] and tests the event for jet candidates from background sources, which are referred to as *bad jets*. Events containing such objects are disregarded.

The analysis uses final state particles which correspond to the hard process of Wt production only. As described in Chapter 2, the large luminosity of the LHC can lead to multiple proton-proton interaction in one bunch-crossing. These additional collisions are dominated by jets and form another background source. To remove these unwanted real jets from the event, the jet vertex fraction (JVF) is used. It is defined as the fraction of the track-momentum of a jet associated to a vertex. In a first step, all tracks are associated to a jet within a distance of $\Delta R \leq 0.4$. The JVF for each vertex j is then calculated as

$$JVF(\mathrm{jet}, \mathrm{vtx}_j) = \frac{\sum_k p_{\mathrm{T}}(\mathrm{trk}_k, \mathrm{vtx}_j)}{\sum_n \sum_l p_{\mathrm{T}}(\mathrm{trk}_l, \mathrm{vtx}_n)},$$

where k is the sum over all tracks associated to vertex j, n is the number of primary vertex candidates in the event and k the sum over all tracks associated to the respective vertex. After the identification of the main primary vertex, a cut is applied on the JVF of each jet, rejecting all jets with $|JVF| < 0.75$. The absolute value is used here as the algorithm defaults to -1 if no tracks are associated to the jet and those jets are usually not from pile-up vertices.

To improve the modelling of the data by Monte Carlo simulation, all simulated jets are rescaled using efficiency and inefficiency scale factors. For this purpose, four factors have been derived using tag & probe methods on events with $Z \to \mu\mu$ and $Z \to ee$ decays back-to-back with a central jet. A truth matching procedure using a maximum distance of $\Delta R < 0.4$ is applied to determine if the efficiency or inefficiency scale factor is to be applied combined with the information if the reconstructed jet has passed the JVF requirement.

The ATLAS reconstruction software does not directly decide if a reconstructed object is a jet or an electron. As the detector signals of such objects can look very similar, they can be identified by the electron and the jet reconstruction algorithms at the same time. One reason are high-p_{T} electron energy deposits leaking into the hadronic calorimeter. To prevent such ambiguities, the overlap of jets and electrons is checked. If one or more jets are found within a distance of $\Delta R(e, \mathrm{jet}) < 0.4$ from the electron, the jet closest to the electron is removed.

4.4.3 Jet energy resolution

The resolution on the jet energy has been measured using different *in situ* techniques as described in [104] using events with exactly two jets. The combined resolution for jets in the rapidity region $|y| < 0.8$ is given in fractions of the transverse momentum resolution ($\sigma(p_{\mathrm{T}})/p_{\mathrm{T}}$) for different bins of mean p_{T} of the dijet system and ranges from 18 % in the region below 40 GeV down to 6 % at more than 300 GeV. As jet energies are used in this analysis to reconstruct W bosons and top quarks from their decay products, an accurate measurement of energy-related quantities is essential. The resolution also has an impact on the accuracy of the missing transverse momentum measurement as described in Section 4.6.

4.5 Flavour of jets

The identification of jets containing b-hadrons, which is called b-tagging, is very important for the reconstruction of top quarks and other particles in the event. Top quarks decay to b quarks and a W boson which in turn decays to two light quarks. Disentangling the quark flavour of the jets improves the top quark as well as the W boson reconstruction. The usage of the correct combination of final state objects for particle reconstruction depends directly on the efficiency of the tagging algorithms. The information can also be used in the event selection to reduce the background contamination.

4.5.1 b-tagging algorithms

This analysis uses the MV1 tagger, which combines three dedicated b-tagging algorithms into one weight. The combination variable for MV1 is the output of a neural network which uses the weights of the three algorithms JetFitterCombNN, IP3D and SV1 together with the transverse momentum and pseudorapidity of the jet as input variables and is trained for a maximal separation between jets originating from b-quark hadronisation and light-quark jets [105]. The three algorithms each exploit unique features of b-quark jets which will be shortly detailed in the following based on [83, 105, 106].

SV1 After the production in e.g. top-quark decays, b quarks can bind to other quarks forming hadrons. These b hadrons have lifetimes around 1.6 ps [28] which corresponds to a mean decay length of $\beta\gamma c\tau =$ 480 µm. Their decay vertices can therefore be identified and the properties of these secondary vertices are exploited. The algorithm scans all track pairs associated to the jet for such vertices and inspects the following properties:

- invariant mass of all tracks associated to the vertex;
- energy fraction of tracks from the jet associated to the vertex;
- number of found two-track vertices.

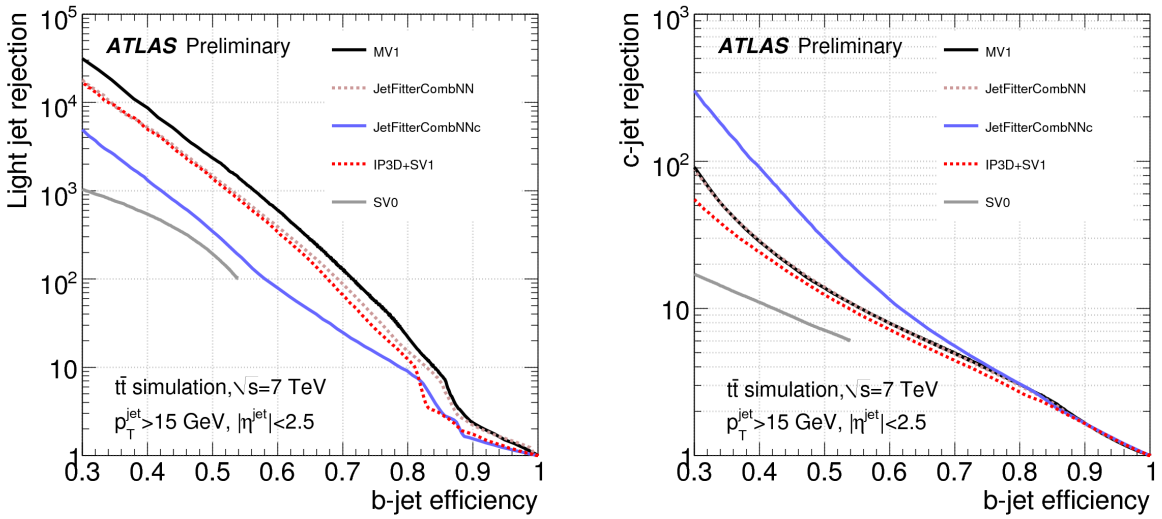
Likelihood ratios are defined for these variables based on the comparison to reference probability distributions for b-quark jets and light-quark jets that are extracted from simulations. The two first variables are merged into a two-dimensional distribution and combined with the third variable adding the logarithms of the likelihood.

IP3D The impact parameter of a track is defined as the distance of closest approach of the track to the primary vertex of the collision event. Tracks associated to a b-quark jet are expected to have a significantly large impact parameter. The significance is defined as the impact parameter divided by its uncertainty. The IP3D tagger uses a two-dimensional reference distribution of the signed transverse impact parameter significance and the longitudinal impact parameter significance from simulations to create a likelihood.

JetFitterCombNN A Kalman filter reconstructs the flight path of the b hadron based on the secondary vertices of the b-hadron decay and the subsequent decay vertex of a c hadron assuming that they all lie on one line. It also reconstructs the position of the decay vertices on the flight path. A neural network is trained to create a single classifier using properties of the flight path and vertices as inputs, such as the vertex masses, momenta, the flight-length significance and track multiplicities of the decay vertices.

4.5.2 Performance of tagging algorithms

The performance of b-quark jet taggers can be expressed with two characteristic variables. The efficiency is defined as the fraction of b-quark jets of all b-quark jets in the sample tagged correctly by the tagger. One can define a light-quark efficiency and also a charm-quark efficiency accordingly. Instead of the efficiency of light quarks and charm quarks, the rejection is used to underline non-b-quarks as background. The rejection is defined as the inverse of the efficiency. The tagging efficiency and background rejection can be determined from simulation samples using a matching of reconstructed jets to the corresponding quarks in the simulation which fixes the flavour of the reconstructed jet. In Figure 4.4 the rejections are shown in dependence of the efficiency of the tagging algorithms mentioned in Section 4.5.1. The MV1 tagging algorithm reaches the best performance of the shown taggers in terms of the light quark jet rejection and also offers a decent rejection of charm-quark jets. A better rejection is only achieved by a tagger which is optimised for charm quark rejection (MV1c).



(a) light-quark jet rejection

(b) c-quark jet rejection

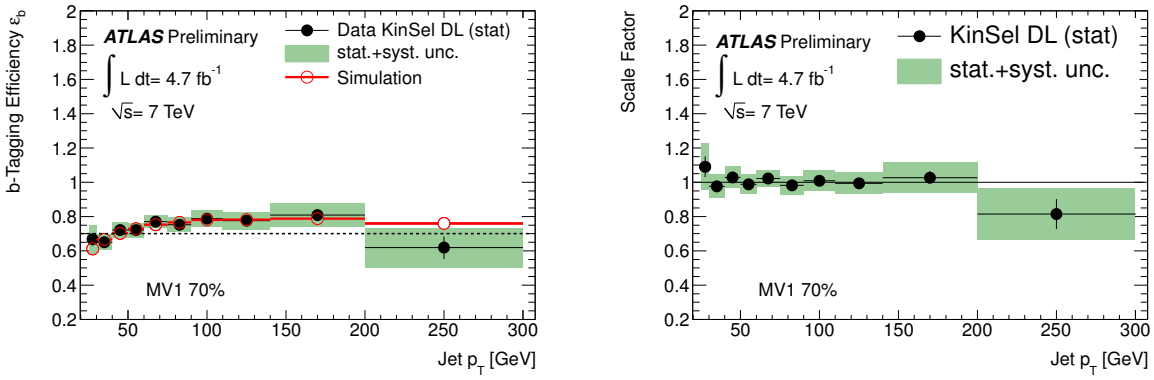
Figure 4.4: Light-quark jet rejection (a) and c-quark jet rejection (b) as functions of the b-tagging efficiency for different b-tagging algorithms based on simulated $t\bar{t}$ events [105].

In the following analysis a working point with 70 % efficiency is chosen to tag jets as b-quark jets. At this efficiency, the tagger reaches rejection factors of 150 for light-quark jets, 14 for tau leptons and 5 for charm-quark jets [107]. The tagged jets can be used in the event selection and to improve the reconstruction of particles in the later analysis.

4.5.3 Calibration of the MV1 tagging algorithm

The tagging algorithm performance extracted from simulation samples does not precisely reflect the performance of the detector. Both efficiency and mis-tag rate therefore have to be measured in data to calibrate the tagging algorithm. The result of the calibration is a p_{T}-dependent scale factor, which is applied to the jets in simulated events in the analysis to model the performance in data. Three different calibration methods are combined to derive these scale factors. Two methods exploit semileptonic b-quark decays that are enhanced in a dijet Monte Carlo simulation sample: $p_{\mathrm{T}}^{\mathrm{rel}}$ and System8. A description of both methods and the derived efficiencies can be found in [108]. The third method uses

events with top-quark pair production to calibrate the tagging algorithms. This sample is enhanced with b-quark jets as they always appear as top-quark decay products. Only events with two leptons in the final state are used by the method to have a statistically independent sample. The method and its results are detailed in [109] and an example for the efficiency and the scale factors for the methods is shown in Figure 4.5.



(a) efficiency

(b) scale factors

Figure 4.5: Example for the measured b-tagging efficiency in data compared to that in simulation (a) and the resulting scale factors (b) for the MV1 algorithm at 70% b-tagging efficiency measured by the dilepton $t\bar{t}$ analysis. The error bars show the statistical uncertainties while the green band indicates the total uncertainty. [109].

Two complementary methods are used to measure the rate of misidentified b-quark jets from data. One method exploits the invariant mass of the tracks associated to a secondary vertex and the other method, called negative tag method, uses the negative decay length of secondary vertices and negative impact parameters of the tracks associated to a jet to enhance wrongly tagged jets. Both methods are described in [110]. The rates were measured again using 4.7 fb^{-1} of 2011 data with the same methods for the MV1 b-quark tagger.

4.6 Missing transverse momentum reconstruction

Some weakly-interacting particles, like neutrinos, escape undetected from the ATLAS detector. As they leave no signal, a reconstruction is only possible using indirect methods. One measure is the missing energy in an event. In a hadron collider, the initial momenta of the partons involved in the hard interaction in the direction of the beam axis are only known on a statistical basis, predicted by the PDF. In the direction perpendicular to this axis, the initial momentum is known to vanish. This allows the missing transverse momentum ($E_{\mathrm{T}}^{\mathrm{miss}}$) to be reconstructed, which is obtained from the negative vector sum of the transverse momenta of all particles in the event. A detailed description of the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction in ATLAS can be found in [111, 112].

The $E_{\mathrm{T}}^{\mathrm{miss}}$ for top-quark analyses is defined in the following way:

$$-E_{x,y}^{\mathrm{miss}} = E_{x,y}^{\mathrm{RefEl}} + E_{x,y}^{\mathrm{RefJet}} + E_{x,y}^{\mathrm{RefSoftJet}} + E_{x,y}^{\mathrm{RefMu}} + E_{x,y}^{\mathrm{CellOut}},$$

where each term on the right side of the equation corresponds to one source of transverse momentum. The magnitude is defined as

$$E_{\mathrm{T}}^{\mathrm{miss}} = \sqrt{(E_x^{\mathrm{miss}})^2 + (E_y^{\mathrm{miss}})^2}$$

and the azimuthal direction as

$$\phi^{\text{miss}} = \arctan(E_y^{\text{miss}}/E_x^{\text{miss}}).$$

The single components are matched as closely as possible to the object definitions from top-quark analyses and the calorimeter clusters are associated to the objects in the order they appear in Equation 4.6. The calibration is applied depending on the matched object. Below a few comments are given on the different terms.

- $\boldsymbol{E}_{x,y}^{\textbf{RefEl}}$: electrons are defined using the Tight++ quality requirements with $p_{\text{T}} > 10\,\text{GeV}$. All corrections described in Section 4.2 are applied, except for out-of-cluster corrections.
- $\boldsymbol{E}_{x,y}^{\textbf{RefJet}}$: Topological clusters assigned to jets with $p_{\text{T}} > 20\,\text{GeV}$ enter in this term. They are calibrated to the hadronic energy scale.
- $\boldsymbol{E}_{x,y}^{\textbf{RefSoftJet}}$: All clusters assigned to jets with $7\,\text{GeV} < p_{\text{T}} < 20\,\text{GeV}$ enter here, calibrated to electromagnetic scale.
- $\boldsymbol{E}_{x,y}^{\textbf{RefMu}}$: In the $E_{\text{T}}^{\text{miss}}$ calculation, the p_{T} of muons reconstructed by the MuID algorithm in the range of the muon spectrometer ($|\eta| < 2.7$) is used. Isolated muons are required to have no jets nearby ($\Delta R = 0.3$). If the muon is isolated, the combined muon enters into the muon term, which includes the muon's energy loss in the calorimeter by definition. The energy loss in the calorimeter of non-isolated muons is covered by a special muon calorimeter term and only the muon spectrometer measurement is added to the muon term to avoid double-counting of the calorimeter energy.
- $\boldsymbol{E}_{x,y}^{\textbf{CellOut}}$: All topological clusters not assigned to any of the objects described before are added to the CellOut term.

4.6.1 Resolution of missing transverse momentum

The performance of the missing transverse momentum reconstruction has been studied using $4.2\,\text{fb}^{-1}$ of 2011 data in [111] using leptonic decays of W and Z bosons. The resolution depends on the scalar sum of all transverse energy in the calorimeter ($\sum E_{\text{T}}$) and varies from 5 GeV for $\sum E_{\text{T}} = 50\,\text{GeV}$ to about 19 GeV at $\sum E_{\text{T}} = 700\,\text{GeV}$. The distribution can be parametrized as $\sigma = k\,\sqrt{\sum E_{\text{T}}}$ with values of around $0.7\,\text{GeV}^{1/2}$ for the fit parameter k.

4.7 Luminosity measurement

The search for new processes and cross-section measurements in general rely on an accurate determination of the delivered luminosity that a dataset used in analysis corresponds to. The uncertainty on the luminosity measurement also translates directly into a systematic error on the measured cross-section and also affects many other measured variables.

In this analysis, the luminosity is used to determine the expected number of events from signal and background processes. The measurement and calibration of the luminosity in ATLAS is explained in [113] and summarized shortly in the following. To determine the luminosity, different measurements of subdetector components are combined, which have orthogonal acceptance regions, pile-up response and sensitivity to other backgrounds. The measurements extract an observed average number of interactions

per bunch crossing, μ_{vis}, which is related to the luminosity by

$$\mathcal{L} = \frac{\mu_{\text{vis}} n_b f_r}{\sigma_{\text{vis}}},$$

where n_b is the number of crossing bunches per revolution and f_r denotes the revolution frequency. The observed variables are connected to the real ones by the efficiency, ϵ, of the algorithm and the detectors,

$$\sigma_{\text{vis}} = \epsilon \sigma_{\text{inel}} \qquad \text{and} \qquad \mu_{\text{vis}} = \epsilon \mu,$$

with the cross-section of inelastic proton collisions σ_{inel}.

To determine the absolute luminosity, the methods need to be calibrated. As reference, the luminosity can be measured from beam parameters from

$$\mathcal{L} = \frac{n_b f_r n_1 n_2}{2\pi \sigma_x \sigma_y},$$

where n_1 and n_2 are the number of protons in the bunches of each of the beams and σ_x and σ_y are the horizontal and vertical convolved beam widths. The beam widths can be measured in van der Meer scans [114, 115] which separate the beams in fixed steps at the interaction points of the collider. The number of protons in a bunch is measured separately upstream in the accelerator chain.

The integrated luminosity of the used dataset in this analyses corresponds to $4.7\,\text{fb}^{-1}$ and the uncertainty is estimated to be $\pm 1.8\,\%$.

CHAPTER 5

Monte Carlo Simulation

Theory predictions will be used in this analysis to search for the production of a single top quark accompanied by a W boson. While Chapter 2 described the current status of the Standard Model and how the theoretical knowledge can be used to simulate events, this chapter describes the used tools and also the simulation of the detector signals that can later be compared to the measured data. By comparing measured events to simulated events it is possible to understand e.g. the composition of the taken data in terms of physics processes. The simulation process can be subdivided into several basic steps: the event generation following the theoretical process, the application of radiative corrections, simulation of the ATLAS detector response and finally the conversion into detector signals (digitisation). Following this process, the result will be the same as raw data produced by the detector. The common ATLAS software framework Athena [116] integrates all the steps listed above and also applies the reconstruction methods, described in Chapter 4 to both simulated and measured events in the same way. The whole production chain is visualized in Figure 5.1 and detailed in the following.

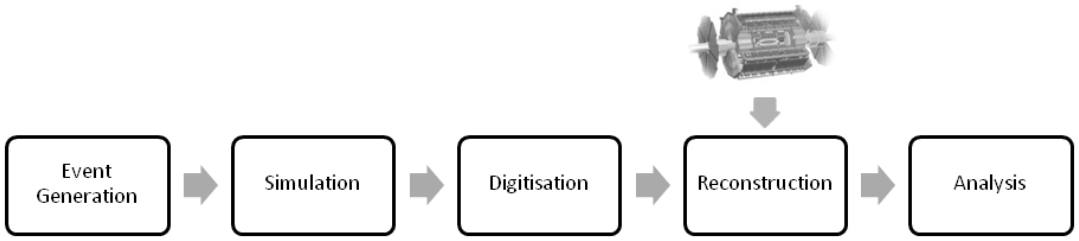


Figure 5.1: Flowchart depicting the process of Monte Carlo simulation and reconstruction in ATLAS. The same software is used for recorded data from the reconstruction step onwards [6].

5.1 Event generation

The first step in the simulation of events is the event generation. Monte Carlo generator packages are used in this step to simulate events for dedicated production channels. Depending on the program, leading-order or next-to-leading-order calculations are used along with randomization procedures to create particle four-vectors. These generators are interfaced to parton shower generators which apply the radiative corrections and simulate the hadronic showers using the four-vectors provided by the former

to obtain the final state and also add underlying event information. For many physics processes used in this analysis, a variety of generator and parton shower tools can be used. This allows the effect of different event generation models on the analysis to be studied.

The main generator used in this analysis is the next-to-leading-order generator MC@NLO [117] in version 4.01. The parton shower simulation is done with the leading-order generator HERWIG [118] version 6.520 which is also used as event generator for some samples. Additionally, POWHEG BOX [119–121] version 1.0, the multileg generator ALPGEN [122] version 2.13 and the leading-order generator AcerMC [123] version 3.8 are employed. Parton showers are also simulated with PYTHIA [124] in version 6.425, which is additionally employed in the generation of minimum-bias pile-up events as described in Section 5.3. Additional tools help to model very specific physics effects. Multiple parton scattering is simulated by the generator JIMMY [125] version 4.31, PHOTOS [126] handles QED bremsstrahlung corrections in the cascading decays and tau lepton decays are simulated by TAUOLA [126].

As described in Chapter 2, the distribution of partons in the proton cannot be calculated analytically and is therefore parametrized by the parton distribution functions, which are determined in combined fits of several proton physics experiments. The PDF sets used in this analysis are CT10 [127] for the NLO generators MC@NLO and POWHEG. The leading-order generators AcerMC, HERWIG and PYTHIA use the MRST LO** PDF set [128] and ALPGEN is used with the CTEQ6L1 PDF set [129].

A full overview of the samples used in the analysis and the respective generator combination can be found in Section 5.4 and detailed in Appendix A.

5.2 Detector simulation

The full set of four-vectors of all final-state particles generated by the Monte Carlo generators still describes an idealistic state. In LHC collisions, the created particles traverse several layers of detector material and are deflected by strong magnetic fields. The influence of the ATLAS detector on the particles, as described in Chapter 3, is simulated using the GEANT4 software toolkit [130]. A detailed geometry description of the full ATLAS detector with all its subcomponents is provided to GEANT4, which in turn calculates the impact positions of particles on the detector components and the corresponding energy deposits. The tool simulates the passage through matter based on the particle type and the provided material information.

5.2.1 Fast detector simulation - ATLFAST2

Due to the high particle multiplicity resulting from proton collisions, the simulation of the material interaction is very computing intensive. Even though a large infrastructure has been created to process the data recorded by the LHC experiments as well as Monte Carlo simulation and data analysis, the sample sizes do play a considerable role. An alternative simulation has been established to be able to produce also samples for systematic studies in the needed sizes. This tool has been named ATLFAST2 which stands for a fast simulation. The technique is to use simpler descriptions of the detector where the impact on the result is small and to use the full simulation described above where it is needed.

The interactions in the Inner Detector are simulated using the full simulation with Geant4. As described in [131], the simulation of the interactions and shower development in the calorimeter, uses about 90 % of the simulation computing time. A parametrization of the shower development of electrons, photons and pions derived from the full simulation is used for simplification. The parametrization for pions is also applied to hadrons in general and a rough geometry of the calorimeter further improves

computing performance. The detector response for muons is again simulated using the full description of the Inner Detector, the calorimeter and the muon system.

The samples simulated with ATLFAST2 were validated against the reference samples created with the full simulation and particle properties are adjusted in the few situations where differences have been observed.

5.3 Digitization and pileup simulation

After simulation of the energy deposits of the particles in the detector as described in the previous Section, the digitization step uses the hits provided by Geant4 and transforms them into detector responses, like voltages and times. The response depends on the detector subsystem and has the same format as the data actually measured by the corresponding system. This step allows the same reconstruction algorithms to be run, which are described in Chapter 4, on the raw files resulting from data-taking or Monte Carlo simulation.

The highly squeezed proton bunches in the LHC ring not only provide a very high instantaneous luminosity, but also lead to unwanted effects such as pileup. Together with the hard process, which physics analyses are mostly interested in, this effect leads to additional particles in the event that have to be modelled in the simulation. Special samples have been generated using the PYTHIA 6 Monte Carlo generator which models secondary interactions (minimum bias events) and are added to the simulated hits of the hard process in the digitization step. Other sources of background particles include beam gas, beam halos and also cavern background particles as described in [132]. Particles from these sources can also be included in the digitization process. A reweighting procedure is used to match the distributions of the mean number of primary vertex candidates in simulation to the one measured in ATLAS data.

5.4 Samples used in this analysis

In the following, the Monte Carlo samples used in this analysis for the different production channels are listed. All used samples were generated in an ATLAS-wide centralized effort by the ATLAS simulation group as part of the MC11c production campaign. The combination of simulation tools for the main samples and the samples for systematic studies are given. More information on the studies of systematic uncertainties which result from the generators is given in Chapter 10.2. A complete list of the used samples and their statistics can be found in Appendix A. All samples are reweighted to the measured luminosity of the used data sample using the theoretically predicted cross-sections. Unless given otherwise, the relative error of the prediction is used as uncertainty for a given Monte Carlo dataset.

Single top-quark production in the Wt-channel The default generator for the signal channel samples of this analysis is MC@NLO, which can only be interfaced to HERWIG and JIMMY. The ATLAS AUET2 tune [133] is applied to improve the description of multiple parton interactions. The diagram-removal method described in Chapter 2 has been applied, but the same sample was also generated using the diagram-subtraction method and can be used for comparison. Additional simulated Wt events are available generated with POWHEG combined with PYTHIA using the Perugia2011C tune [134], and generated by the leading-order generator AcerMC with the PYTHIA parton showering model.

Single top-quark production in the t-channel In contrast to the other single top-quark production channels, the t-channel production is modelled by the AcerMC generator at leading order only, as the available NLO generators create additional unphysical jets. The parton shower is simulated by PYTHIA

applying the Perugia2011C tune. Only top-quark decays with leptons in the final state are simulated. The sample is divided into three subsamples for each lepton type in the final state (electron, muon and tau).

Single top-quark production in the *s*-channel Events with single top-quark production in the *s*-channel are generated with the same generators as the Wt signal samples. The MC@NLO generator combined with HERWIG is also used as default here. Only events with a lepton in the final state have been generated. The fully hadronic final state has not been taken into account as it is suppressed by the trigger selection and the event selection.

Top-quark pair production All the generator combinations used for the single top-quark production samples are also used to generate samples for $t\bar{t}$ production. The MC@NLO generator combined with HERWIG is again used as default here, matching the choice for two of the single top-quark production channels. The samples are split into two subsamples containing fully hadronic final-state events and final states with at least one lepton. Fully hadronic decays are not taken into account in this analysis as the acceptance for such events is negligible.

W+jets The generator ALPGEN in combination with HERWIG is used to create events with W-boson production. The samples only contain events where the W boson decays into a lepton and a neutrino and are subdivided according to the number of additional partons in the event ranging from zero to five. They are further separated regarding the lepton type (electron, muon or tau). Separate samples are available modelling the production of W bosons in association with heavy-flavour quarks. $W + b\bar{b}$ production and $W + c\bar{c}$ production samples are modelled with zero up to three additional partons and the Wc production samples contain zero to four additional jets as well as phase space cuts ($p_{\mathrm{T}}(c\text{-jet}) > 10\,\mathrm{GeV}$ and $\Delta R(c, \mathrm{jet}) > 0.7$, the distance between the c-quark jet and the closest light-quark jet). The samples are subdivided according to the number of additional partons and are not separated according to the lepton type. All samples are reweighted according to their respective cross-section to describe the W+jets production.

As only loose phase space cuts on the heavy quarks are applied, the heavy flavour production samples have considerable overlap with the light flavour samples which can contain heavy flavour quarks from gluon splittings. To avoid double-counting in the combination of these samples, the *heavy flavour overlap removal tool* is available in the ATLAS software which reclassifies the events according to truth information. The output of the tool is used in the analysis to sort events into W-boson production with heavy flavour or light flavour. The actual flavour composition is determined in-situ in the analysis.

The normalization of the number of events for the ALPGEN samples is derived from the Monte Carlo sample and renormalized to the measured data luminosity, just like for the other channels. As this prediction is based on the leading-order generator, an additional scale factor of 1.2 is applied to the $W + b\bar{b}$, $W + c\bar{c}$ and W+ light jets samples, based on a next-to-leading order calculation using FEWZ [135]. A scale factor of 1.52 is applied to the $W + c$ samples, which has been extracted from next-to-leading order calculations using MCFM [136]. The uncertainty on the normalization is based on the theoretically predicted one, which is 4 % for inclusive W production for events with no extra jets. For each additional jet an uncertainty of 24 % is added in quadrature, which is derived from studies of the ratio of $W + (n + 1)$-jet production to $W + n$-jet production.

Z+jets Events with Z-boson production are also generated with the ALPGEN generator and the HERWIG parton shower model. In the analysis samples containing Z bosons are used with up to five light

flavour partons. They are separated according to the lepton type of the Z-boson decay and the number of partons. Additional samples model the production of Z bosons, b-quark pairs and up to three additional partons. The same heavy flavour overlap removal tool as in the case of W-boson production is used to avoid double-counting when combining to the two classes of samples.

The normalization procedure for Z+jets events is the same as for W+jets events, as the same generator is used. The uncertainty on the normalization is also calculated in the same way.

Diboson production The simulation of events with the production of W-boson pairs, Z-boson pairs and WZ is done with the HERWIG leading-order generator. They are commonly named diboson samples. The samples are filtered to contain one lepton with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ and $|\eta| < 2.8$.

CHAPTER 6

ATLAS Collision Data

The analysis uses LHC collision data recorded by the ATLAS detector. The conditions of the accelerator and the detector during the data-taking period are summarized in the first part of this chapter. The second part of this chapter details the data-driven methods which are used to determine the background impact of multijet production, as not all background channels of the single top-quark production in the Wt-channel can be well described by simulation data.

6.1 Collision data used in the analysis

In this analysis a dataset corresponding to an integrated luminosity of $4.7\,\text{fb}^{-1}$ collected by the ATLAS detector at the LHC in 2011 is used, which amounts to the full data usable for physics analyses for that year. The luminosity is a measure of the number of proton collisions and hence depends on the operation parameters of the LHC collider during the year which are detailed in the following.

6.1.1 Start of the LHC and run conditions in 2011

The official start of the LHC in the year 2008 was interrupted by a severe accident in the accelerator after just a few days. A faulty superconducting connection between magnets lead to a magnet quench which in turn evoked a helium explosion. As a consequence, several magnets were destroyed and displaced. After finishing repairs, the first proton collisions at a centre-of-mass energy of 450 GeV could be recorded by the LHC experiments in late 2009. The year 2010 marked the first year of continuous LHC operation. An integrated luminosity of $45\,\text{pb}^{-1}$ of proton collision data was collected by the ATLAS detector at a collision energy of $\sqrt{s} = 7\,\text{TeV}$. During the year, extensive machine development took place. The number of bunches per LHC beam was in-

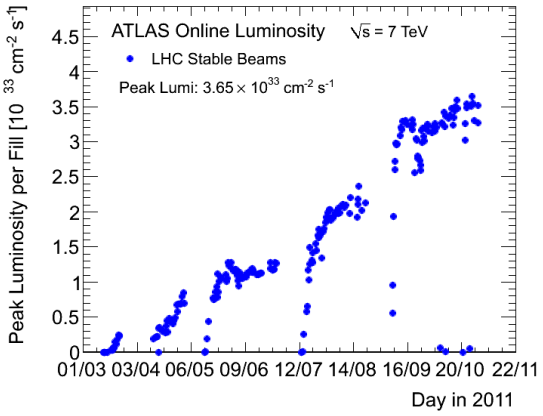


Figure 6.1: The maximum instantaneous luminosity versus day delivered to ATLAS in 2011. Only the peak luminosity during stable beam periods is shown [137].

creased from only one up to 348 and the instantaneous luminosity increased from $1 \times 10^{27}\,\mathrm{cm^{-2}s^{-1}}$ up to $2 \times 10^{32}\,\mathrm{cm^{-2}s^{-1}}$.

After the first successful run, the goals set for 2011 were to increase the instantaneous luminosity even further and to deliver a large dataset of collision events to the experiments for physics analyses, recording over the whole year. Through a combination of machine development and data-taking, these expectations could be reached. The maximum instantaneous luminosity in 2011 was $3.65 \times 10^{33}\,\mathrm{cm^{-2}s^{-1}}$ with 1854 colliding bunch-pairs and an integrated luminosity of more than $5\,\mathrm{fb^{-1}}$ has been delivered to each of the large experiments, ATLAS and CMS. The distribution of the day-dependent peak luminosity shown in Figure 6.1 gives a good impression of the successful increase of both number of bunches in the LHC as well as stronger beam focusing. Only data that is taken during the stable beam conditions is counted towards the integrated luminosity. In this phase, no adjustments are made to the beam orbit to ensure safe operation of all detector components.

6.1.2 ATLAS 2011 data-taking period

The experiment's aim is to record the collisions delivered by the LHC as efficiently as possible. This requires all subcomponents of the detector to be fully operational during data taking, which is ensured by the data quality group. The overall recorded luminosity is mostly reduced due to the detector start-up time. Some detector components close to the beam are enabled only after stable beam is declared. Figure 6.2 shows a comparison of the total integrated luminosity evolution for all data delivered by the LHC, the data recorded by the ATLAS detector and the data used in physics analysis for the year 2011. Physics analyses have demands on the data quality of individual detector components as well as subsystems and event variables. Again the distribution clearly shows the effect of increased instantaneous luminosity over the year as seen from Figure 6.1.

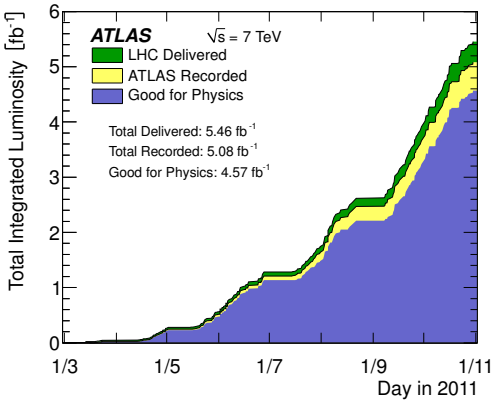


Figure 6.2: Cumulative luminosity versus time delivered to (green), recorded by ATLAS (yellow), and certified to be good quality data (blue) during stable beams for pp collisions at 7 TeV centre-of-mass energy in 2011 [137].

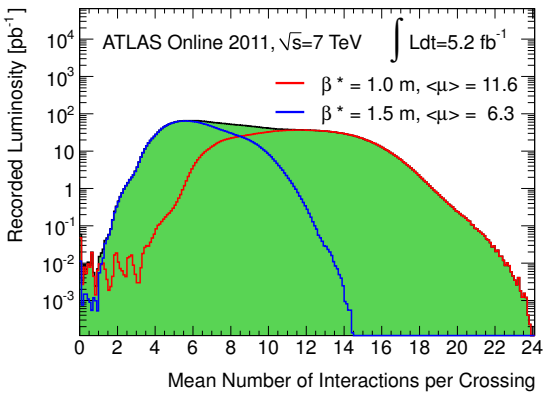


Figure 6.3: Luminosity-weighted distribution of the mean number of interactions per bunch-crossing for 2011. The plot is shown for data taken before and after the September Technical Stop where the β^* was reduced from 1.5 to 1.0 m [113, 137].

A negative effect of the stronger beam focus is increased pile-up. Figure 6.3 shows the amount of data taken with a certain mean number of interactions per bunch-crossing. The allotment to the two different focus configurations is also shown. Clearly the stronger beam focus, lower β^*, which is a measure for the transverse size of the beam at the interaction point, favours larger pile-up which can influence reconstruction and analyses.

The analysis uses single lepton final states, where leptons refer to muons and electrons. Therefore, only events from the ATLAS data streams *Egamma* and *Muons* are used as described in Section 3.2.5 for the corresponding electron and muon channels. Due to the luminosity enhancement over the year, the trigger definitions had to be changed over the course of the year. The analysis always uses the unprescaled lepton trigger with the lowest possible transverse momentum requirement as detailed in Chapter 7. Data taken in 2011 are grouped into trigger periods ranging from A to M out of which only the periods B to M are used in this analysis. The trigger periods are used to reweight the pile-up distribution of the simulated data to match the one observed in collision data and to apply the correct trigger requirements.

To account for problematic detector conditions during a data-taking run a list of good runs and the corresponding luminosity blocks is provided by the data quality experts to the physics analysis groups. A luminosity block is hence the smallest collection of events for which the luminosity is determined and corresponds to about two minutes of data-taking.

6.2 Data-driven background estimations

Most channels which contribute to the overall background contamination can be modelled using Monte Carlo event generators as described in Chapter 5. The generation of events for some production channels can be complicated or even technically impossible. The reasons are e.g. a low selection efficiency or the need to have a good description of the tails of distributions. To model a channel well and allow analysis conclusions later, a large number of events compared to the number of expected events for this channel in data is required. If this is not the case, the few generated events get large weights assigned to them, which leads to non-smooth distributions and therefore an imprecise modelling. The only technical solution would be to generate a huge number of events which can become problematic for channels with large cross-sections and low selection efficiencies – many events would be generated just to be cut away in the selection process and play no role.

In this analysis, only the background from events with fake electrons and muons is modelled using the data-driven techniques discussed in the following. Events of that category are described with the term multijet events, as no leptons are expected to be produced in the main production process directly (primary leptons). The produced leptons mainly stem from misidentification or from semileptonic quark decays. Two different models are used: the jet-electron model is used to obtain events with fake electrons and the matrix method is applied in the muon channel. Both models were studied in the framework of the ATLAS top group and applied to several analyses involving top quarks.

6.2.1 The jet-electron model

As calorimeter-based objects, reconstructed electrons can be mimicked by jets with properties that are similar to the electron definition. This is exploited to select events from collision data which can be used to model the multijet background in the electron channel. The normalization of the multijet contribution is extracted using a binned likelihood fit of the missing transverse momentum distribution.

In a first step, events containing an electron-like jet have to be selected applying the same selection as later in the analysis. Events from the JetTauEtmiss data stream are selected requiring a jet trigger with up to 20 GeV transverse energy threshold. Different triggers were combined for the full data sample to achieve the needed statistics. The electron-like jet in the event is required to have a transverse energy of $E_{\mathrm{T}} > 25\,\mathrm{GeV}$, at least three tracks associated to it and it has to be reconstructed in the pseudorapidity range defined for the electron, including the calorimeter transition region cut. As electrons deposit most energy into the electromagnetic calorimeter, the jets are additionally required to have a fraction of 80 to

95 % of their energy deposited into that part of the calorimeter. The jet passing these requirements is then used as an electron and the event is vetoed in the case that it contains another electron as defined in the electron reconstruction in Section 4.2, except for the isolation criteria. A random charge is finally assigned to this jet-electron and the standard overlap removal is applied.

Using the full event selection, without a cut on the missing transverse momentum, significantly enhances the background-contribution, especially the multijet event fraction. A binned likelihood fit is applied to the complete distribution of the missing transverse momentum and the fraction obtained from the fit result is used subsequently for the normalization of the multijet background. The fit is performed separately for events with b-tag requirement and without, accounting for the different background composition in the two regimes.

6.2.2 Matrix method

The estimation of events with fake muons is performed using a matrix method. This method defines a region enriched with events containing fake muons and predicts the number of such events in the analysis region. The prediction relies on the assumption that the probability of finding a real or fake muon, which is selected in the fake-enriched region, is the same in the signal region. These probabilities are defined as efficiencies and can be derived using different approaches as explained in this section. The fake-enhanced region in this case is defined by a looser definition of the muon object than the default selection given in Section 4.3. A loose muon candidate lacks the requirement of isolation compared to the default, which is called a tight muon. It is possible to categorize muons into real ones, in the sense of primary muons, and fake muons:

$$N = N_{\text{real}} + N_{\text{fake}}.$$

The same relation also holds for number of selected loose and number of selected tight muons separately:

$$N^{\text{tight}} = N^{\text{tight}}_{\text{real}} + N^{\text{tight}}_{\text{fake}} \quad \text{and} \quad N^{\text{loose}} = N^{\text{loose}}_{\text{real}} + N^{\text{loose}}_{\text{fake}}.$$

The efficiencies can then be defined as

$$\epsilon_{\text{real}} = \frac{N^{\text{tight}}_{\text{real}}}{N^{\text{loose}}_{\text{real}}} \quad \text{and} \quad \epsilon_{\text{fake}} = \frac{N^{\text{tight}}_{\text{fake}}}{N^{\text{loose}}_{\text{fake}}}.$$

Using these, it is possible to express the relation for tight muons in terms of the number of loose muons as

$$N = \epsilon_{\text{real}} N^{\text{loose}}_{\text{real}} + \epsilon_{\text{fake}} N^{\text{loose}}_{\text{fake}}.$$

The figure needed for analysis is the number of fake leptons passing the tight selection. Assuming that the efficiencies are considerably differing from each other, this number can be calculated from the following relation:

$$N^{\text{tight}}_{\text{fake}} = \frac{\epsilon_{\text{fake}}}{\epsilon_{\text{real}} - \epsilon_{\text{fake}}} \left(\epsilon_{\text{real}} N^{\text{loose}} - N^{\text{tight}} \right).$$

The numbers of loose and tight muons can be counted applying the different object definitions, while the real and fake efficiencies need to be determined.

The real efficiency (ϵ_{real}) has been determined using simulation data and cross-checked with a tag-and-probe method on Z-boson production events ($Z \to \mu\mu$). The extraction of the fake efficiency has been more elaborate. Here, the fact is exploited that most multijet events that appear as background for top-quark production result from heavy-flavour quark jets. Due to the lifetime of the heavy quark, the muon that results from the semileptonic b-quark decay is expected to have a larger impact parameter

with respect to the primary vertex of the event compared to prompt muons. The impact parameter of tracks is also used in b-tagging of jets as explained in Section 4.5. In this case, the signed impact parameter, d_0^{sign}, is used to define a background enriched region – all events where the muon has large d_0^{sign}. An efficiency function, depending on the cut value d_0^{cut} is then constructed:

$$\epsilon(d_0^{\text{cut}}) = \frac{\sum_{d_0^{\text{sign}} > d_0^{\text{cut}}} N^{\text{tight}}}{\sum_{d_0^{\text{sign}} > d_0^{\text{cut}}} N^{\text{loose}}}.$$

The full distribution is fitted with the heuristically determined function

$$f(d_0^{\text{cut}}) = a\, \mathrm{e}^{-b\,(d_0^{\text{cut}})^2} + c\, d_0^{\text{cut}} + d,$$

where the Gaussian part parametrizes the prompt-muon events in the region with a low cut value and the linear part models the multijet contribution. The extrapolation of the linear function to the point $d_0^{\text{cut}} = 0$ (factor d in the parametrization) corresponds to the fake efficiency.

With the knowledge of the fake and real efficiencies it is then possible to calculate the number of fake muons from data. The actual distribution of the multijet background is generated from collision data using a weight factor for every event which is derived from the rate calculation and differs for events with a loose (w^{loose}) and a tight muons (w^{tight}):

$$w^{\text{loose}} = \frac{\epsilon_{\text{fake}}}{\epsilon_{\text{real}} - \epsilon_{\text{fake}}} (\epsilon_{\text{fake}} - 1) \quad \text{and} \quad w^{\text{loose}} = \frac{\epsilon_{\text{fake}}}{\epsilon_{\text{real}} - \epsilon_{\text{fake}}} \epsilon_{\text{real}}.$$

These weights are calculated using η-dependent efficiencies. The efficiencies are also extracted separately for events with b-tag requirement and without. The method has been developed using simulation data and the final fake efficiency has been extracted using the full 2011 dataset. Several cross-checks, like application of the method in different pseudorapidity-regions and data-taking periods as well as closure tests have been performed. The resulting distributions are shown in the next chapter.

CHAPTER 7

Selection of Wt Channel Candidate Events

The reconstructed physical objects described in Chapter 4 are used to define an event selection which follows the expected decay topology of single top-quark production in the Wt-channel. This event selection is applied to simulated and measured data in the same way. The comparison of distributions of the two data sets allows one to judge the quality of the simulation. A good description of the collision data by the simulation is both a prerequisite for any theoretical interpretation and for the application of further analysis tools, like the multivariate analysis using an artificial neural network presented here. In this chapter, the basic preselection will be detailed in Section 7.1 followed by a stricter filtering of events using b-tagging information in Section 7.2. At the end of the chapter, an overview of the selection performance is given comparing collision and simulation data.

7.1 Preselection

A base selection, which is shared among several top-quark analyses evaluating data taken in the same period of time in the ATLAS collaboration, is applied to all data. As the decay topology of the Wt-channel is very similar to that of the top-quark pair production, the preselection follows the one from $t\bar{t}$ analyses. This decision allows techniques, like the multijet fake estimates, as well as efficiency, resolution and energy scale corrections, to be used across different analyses. They remain valid as long as the selection does not deviate too far.

All collision events are required to be included in the good run list (GRL) provided by the ATLAS data quality experts. As mentioned before, different triggers have to be used in the selection due to the increasing instantaneous luminosity of the LHC. The trigger cuts are also used to divide the analysis into an electron and a muon channel. In each channel, the lowest unprescaled lepton event filter trigger is always used for each run period. In the muon channel, the trigger `EF_mu18` is required for periods B to I and `EF_mu18_medium` for all the following periods both in collision data and simulation samples. The number in the trigger name always corresponds to the minimum lepton transverse momentum which is required on trigger level, while the appendix `medium` denotes an isolation requirement. In the electron channel the situation is slightly more elaborate, as the trigger menu had to be changed more often. `EF_e20_medium` is used for periods B-J, `EF_e22_medium` is required for period K and periods L-M require `EF_e22vh_medium1` or `EF_e45_medium1`. For the simulation samples equivalent to periods L to M, `EF_e22_medium1` is used, as `EF_e22vh_medium1` is not available in the simulation trigger menu. The efficiencies have been confirmed to be the same for these triggers and the analysis cut of

$p_{\mathrm{T}}(e) > 25\,\mathrm{GeV}$ is applied in the high efficiency plateau region for both.

After the above trigger and data quality cuts, the following selection is applied using the offline object definitions given in Chapter 4. A veto is applied on non-collision background events: only events which include a primary vertex with at least four associated tracks are accepted. Exactly one central electron or muon with $p_T(l) > 25\,\mathrm{GeV}$ is required, depending on the lepton channel defined by the trigger cut. A second electron is vetoed with the same definition while no muons with a $p_T(\mu) > 20\,\mathrm{GeV}$ are to be found in the event. This lower transverse momentum condition enhances the muon veto creating a better suppression of background events including Z-boson decays with only small impact on the single top-quark selection efficiency. Finally, the selected lepton must match the one found by the trigger system within a distance of $\Delta R < 0.15$.

As a neutrino is produced in the decay of one of the W bosons in signal events, a missing transverse energy of $E_T^{\mathrm{miss}} > 30\,\mathrm{GeV}$ in the electron and $E_T^{\mathrm{miss}} > 25\,\mathrm{GeV}$ in the muon channel is demanded. To suppress the number of multijet events a two-dimensional cut is applied in the muon channel: $m_T(l\nu) + E_T^{\mathrm{miss}} > 60\,\mathrm{GeV}$, where $m_T(l\nu)$ is the transverse mass of the leptonically decaying W boson and is calculated using the reconstructed lepton four-vector and E_T^{miss}. For particles with negligible masses, it is defined as

$$m_T^2(l_1 l_2) = 2E_T(l_1)E_T(l_2)(1 - \cos\phi),$$

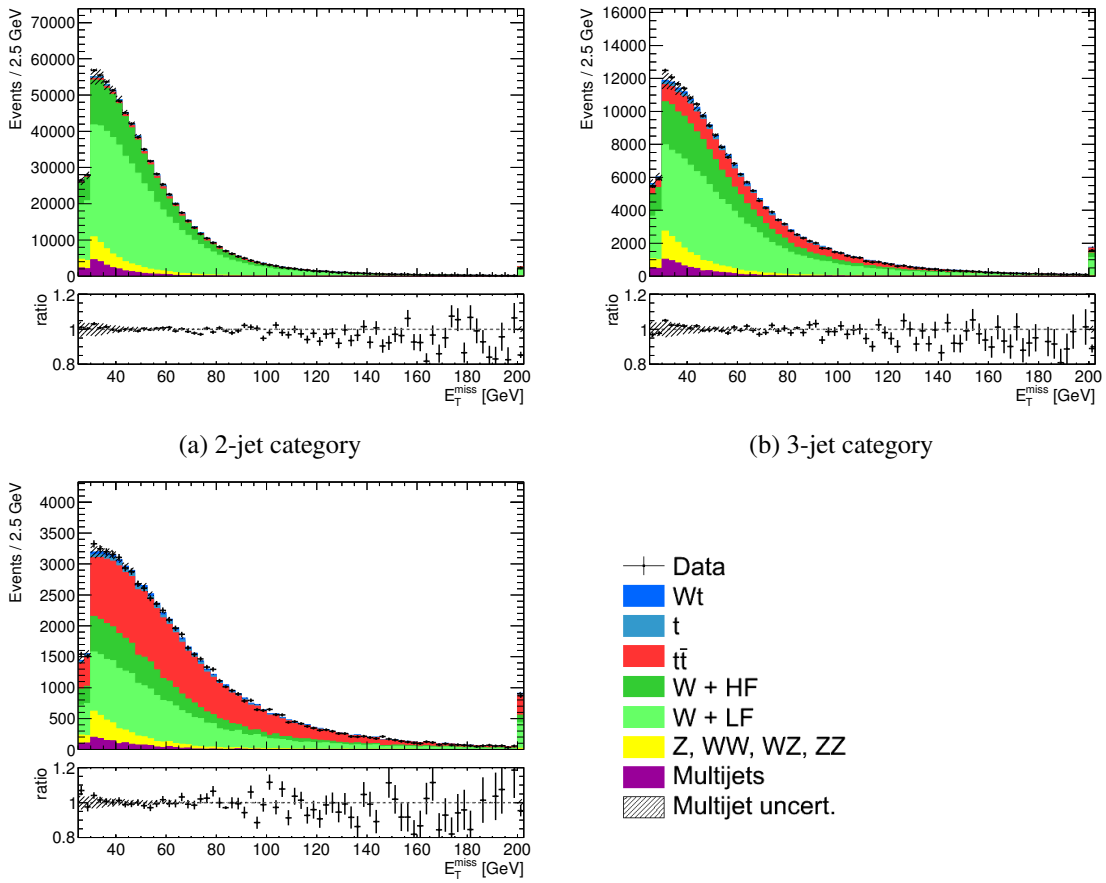
where $E_T = \sqrt{m^2 + p_T^2}$ and ϕ is the angle between the two particles in the transverse plane. In the electron channel, a single cut on $m_T(l\nu) > 30\,\mathrm{GeV}$ is favoured. The final state of Wt-channel production with one lepton contains three quarks, one of which is a b quark. The analysis therefore only considers events with two jets or more. It is then split into categories with two, three and four jets exclusively as each category has a unique background composition as detailed below.

	2 jets	3 jets	4 jets
Single top Wt	3060 ± 280	2910 ± 260	1280 ± 110
Single top t	7200 ± 300	3400 ± 140	1180 ± 50
Single top s	606 ± 25	219 ± 9	56.7 ± 2.4
$t\bar{t}$	18 100 ± 1700	32 500 ± 3100	28 600 ± 2700
$Wc\bar{c}, Wb\bar{b}, Wc$	170 000 ± 60 000	44 000 ± 19 000	11 000 ± 5000
W + jets	480 000 ± 160 000	100 000 ± 42 000	21 000 ± 10 000
Z + jets	36 000 ± 12 000	12 000 ± 5000	3300 ± 1600
WW, WZ, ZZ	8300 ± 800	2280 ± 230	460 ± 50
QCD multijets	27 000 ± 14 000	7500 ± 3800	1700 ± 900
Expected	740 000 ± 200 000	210 000 ± 50 000	69 000 ± 12 000
Data	765 063	201 264	66 428

Table 7.1: Expected and observed number of events after applying the preselection. The numbers are shown for the three categories of events with two, three or four jets. The number of events for QCD multijets is the result of the data-based estimation explained in Section 6.2. The uncertainties on the simulation data are given according to Section 5.4.

The resulting event yields for this selection are shown in Table 7.1 for the different simulated production channels compared to collision data. The overall agreement between expected and measured event numbers is on the level of a few percent and well within the uncertainties. It is very accurate given that the expectation partially has large uncertainties. The same number of signal events are expected in the two-jet and three-jet bins, while only about half the number of events are expected in the four-jet bin.

The by far largest background, especially in events with fewer jets, is the production of single W bosons accompanied by jets. In events with a larger number of jets, the production of top-quark pairs becomes more and more prominent. As a general trend it can be seen that the total number of expected and measured events diminishes with the number of required jets. An interesting fact seen in the top-quark pair production is that the number of expected events is lower in the four-jet bin than in the three-jet bin, while the former is the signal region of this channel in which all expected final state particles can be reconstructed. The reason for this effect is the cut on the number of jets with $p_{\mathrm{T}}(jet) > 25\,\mathrm{GeV}$ and $|\eta| < 2.5$. Apparently the requirements lead to an expected final state quark jet not to enter the selection for $t\bar{t}$ production. The same feature can also be observed in other production channels, like the Wt signal channel, and in the other jet bins.



(a) 2-jet category

(b) 3-jet category

(c) 4-jet category

Figure 7.1: Distributions of the missing transverse momentum for the three different jet categories applying the pretag selection. Electron and muon channels are merged. The simulation distributions are normalized according to the theory predictions. The number of simulation events is finally renormalized to match the number of collision events. The hashed area corresponds to the uncertainty resulting from the mulitjet estimation. The ratio of data and simulation is shown in the bottom plot.

Control distributions of the missing transverse momentum for all three jet categories are shown in Figure 7.1. The last bin on the horizontal axis contains the integrated overflow of the distribution. The same scheme is applied to all subsequent distributions. Overall the three distributions show a good agreement between simulation and collision data. The good agreement supports the models used for

the determination of the multijet background contributions. Deviations of the order of a few percent are visible in the lower E_T^{miss} region but they stay within the multijet uncertainty which is also maximal in this region. Another source of deviations is the large uncertainty on W+jets normalization which is redetermined in-situ in the later analysis. The distributions visualize again the major background contributions detailed above. With a signal-to-background ratio of 0.4 %, 1.4 % and 1.9 %, for the two-jet, three-jet and four-jet bins, it is impossible to extract a signal at this analysis stage. Therefore an improved selection is applied.

7.2 Tag selection

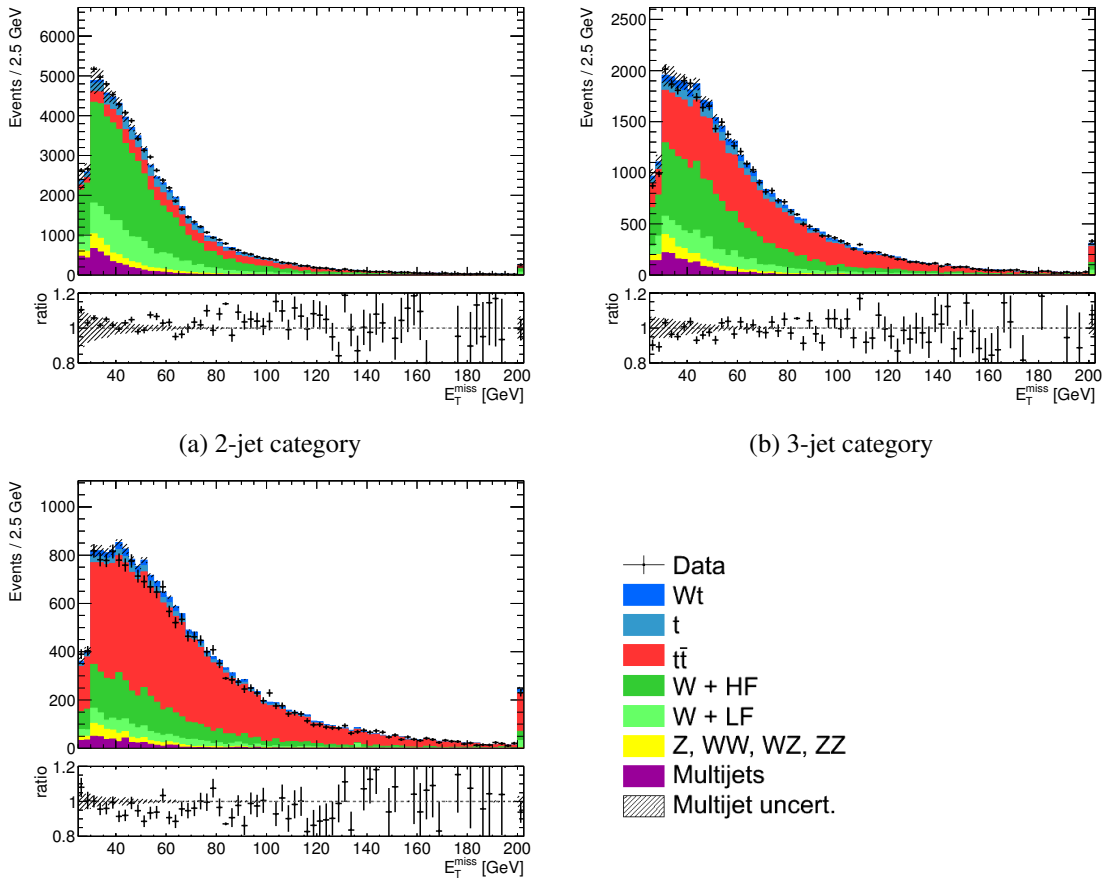
As the only difference between the $t\bar{t}$ and the Wt final states is the absence of a second b-quark jet in the Wt system, a cut on b-tagged jets is one of the few separation possibilities. In this refined selection only events with exactly one b-tagged jet are used. After applying this cut, the event yields shown in Table 7.2 are obtained.

	2 jets	3 jets	4 jets
Single top Wt	1680 ± 150	1670 ± 150	674 ± 61
Single top t	4340 ± 180	1800 ± 80	581 ± 24
Single top s	294 ± 12	100 ± 4	25.6 ± 1.1
$t\bar{t}$	9700 ± 900	16 000 ± 1500	12 700 ± 1200
$Wc\bar{c}, Wb\bar{b}, Wc$	35 400 ± 12 000	11 000 ± 5000	3300 ± 1600
W + jets	12 000 ± 4000	3700 ± 1600	1100 ± 500
Z + jets	2200 ± 800	1100 ± 500	420 ± 200
WW, WZ, ZZ	930 ± 90	315 ± 31	74 ± 7
QCD multijets	4400 ± 2200	1700 ± 900	480 ± 240
Expected	71 000 ± 15 000	38 000 ± 6000	19 300 ± 2100
Data	72 761	37 503	18 518

Table 7.2: Expected number of simulation and data events after applying the full event selection including the b-tagging requirement. The numbers are shown for the three categories of events with two, three or four jets. The number of events for QCD multijets are the result of the data-based estimation explained in Section 6.2. The uncertainties on the simulation data are given according to Section 5.4.

Again an agreement on the level of below five percent is reached, which allows the theoretical and data-driven predictions to be used to draw conclusions on the data composition. The number of signal events is reduced nearly by a factor of two compared to the preselection while in general the background is well suppressed. The highest signal-over-background ratio is now reached in the three-jet bin with 4.6 %. A ratio of 3.6 % is expected for events with four jets and two jets events still have the lowest ratio with 2.4 %. These numbers underline the successful background suppression of the cut. The largest reduction is achieved in the W+jets production with light-flavour quark jets (95 to 98 %). Events with Wc, $Wc\bar{c}$ and $Wb\bar{b}$ production are now the most prominent background in two-jet events with a reduction of 70 % and more over all jet-bins. The b-tag requirement also reduces the less prominent backgrounds, Z+jets, diboson and multijet production; the reduction ranges from 72 to 94 % in all three jet categories.

In general a clear trend on the most important backgrounds can be seen. The W+jets production dominates in the lowest jet bins and is reduced with the number of jets while the top-quark pair production



(a) 2-jet category

(b) 3-jet category

(c) 4-jet category

Figure 7.2: Distributions of the missing transverse momentum for the three different jet categories after applying the full selection. Electron and muon channels are merged. The simulation distributions are normalized to the measured luminosity. The hashed area corresponds to the uncertainty resulting from the mulitjet estimation. The ratio of data and simulation is shown in the bottom plot.

shows the opposite trend. In contrast to the preselection, nearly the same number of events are expected for both backgrounds in the three-jet category. In the $t\bar{t}$ signal region of events containing four jets, the W+jets background only plays a minor role.

In Figure 7.2 the same distributions of the missing transverse momentum as in the previous section are shown for the tightened selection. Only the reweighting of the number of simulated events to the number of collision events is given up. In all three categories a clear enhancement of top-quark production can be observed compared to the preselection, especially in the categories with three and four jets. The description of the collision data by the simulation is again fair and most seen deviations are statistical fluctuations as observable in the ratio plot.

From above considerations, the three-jet category offers the best foundations to extract a signal: the maximal signal share of all categories and a fully reconstructible Wt final state which can be used to build background-discriminating variables.

7.3 Selection performance

In the previous sections, the general agreement of the simulated data with the measured collision data has been shown. In this section, the key kinematic distributions of the reconstructed final state particles are discussed. Such control plots are important to spot possible problems with the theoretical models as well as the detector simulation. As the category with three jets shows the highest signal-over-background ratio, the following considerations will focus on events with exactly three jets in the final state, one of which is a b-quark jet.

In Figure 7.3, distributions of the electron transverse momentum, pseudorapidity and the polar angle are shown. In all three distributions a good agreement of data and simulation is seen. Some deviations can be seen in the lower p_T range. The peaking structure at p_T(e) around 35 GeV is visible in the W+jets production. It coincides with a peak in the multijet distribution, which is purely statistical. Similar statistical fluctuations can be seen in the other two plots. The reason is the low statistics input of the jet-electron model. It uses a very special combination of triggers, as explained in Section 6.2.1 and therefore only yields a small number of events. Large factors from the normalization rescaling enhance these fluctuations. The same fluctuations can also occur in the W+jets samples, where luminosity scaling enhances the problem. The merging of electron and muon channels reduces these problems significantly so that they do not impose any constraints on the later analysis and its results. The pseudorapidity distribution nicely shows the cut-out of the inefficient region of the calorimeter, while the events are isotropically distributed in ϕ which proves a regular efficiency of the calorimeter in this variable.

The same distributions for selected muons are shown in Figure 7.4. They show an even better agreement than the electron distributions. The multijet distribution shows much smaller fluctuations, which is a result of the used model. The applied matrix method, which is described in Section 6.2.2, yields a higher number of events as it generates events from the full data set of collision events. Large fluctuations in the muon channel only arise from the modelling of the W+jets background distributions. The η distribution nicely verifies the inefficient region at $\eta = 0$, where structural elements of the muon system prevent detection. Two clear dips are also visible in the η distribution and in the ϕ distribution at negative values. The reason for the two dips are the feet of ATLAS – support structures that stand on the ground and hold the detector. Fewer muon chambers are installed inside of these structures which lowers the efficiency.

After verifying the lepton distributions, the jet kinematics also need to be checked. Figure 7.5 shows the transverse momentum spectra of the three jets sorted from hardest to lightest jet in the event for the combination of electron and muon events. All three distributions show a very good agreement between collision data and simulation with deviations only in the order of a few percent. Additionally, the distributions of the b-tagged jet are shown. As correlated with the first three distributions, the transverse momentum is also well described. The same holds for the pseudorapidity and polar angle of the b-tagged jet. The statistical fluctuations for the two merged lepton channels are indeed much smaller and do not disturb the good modelling of the variables.

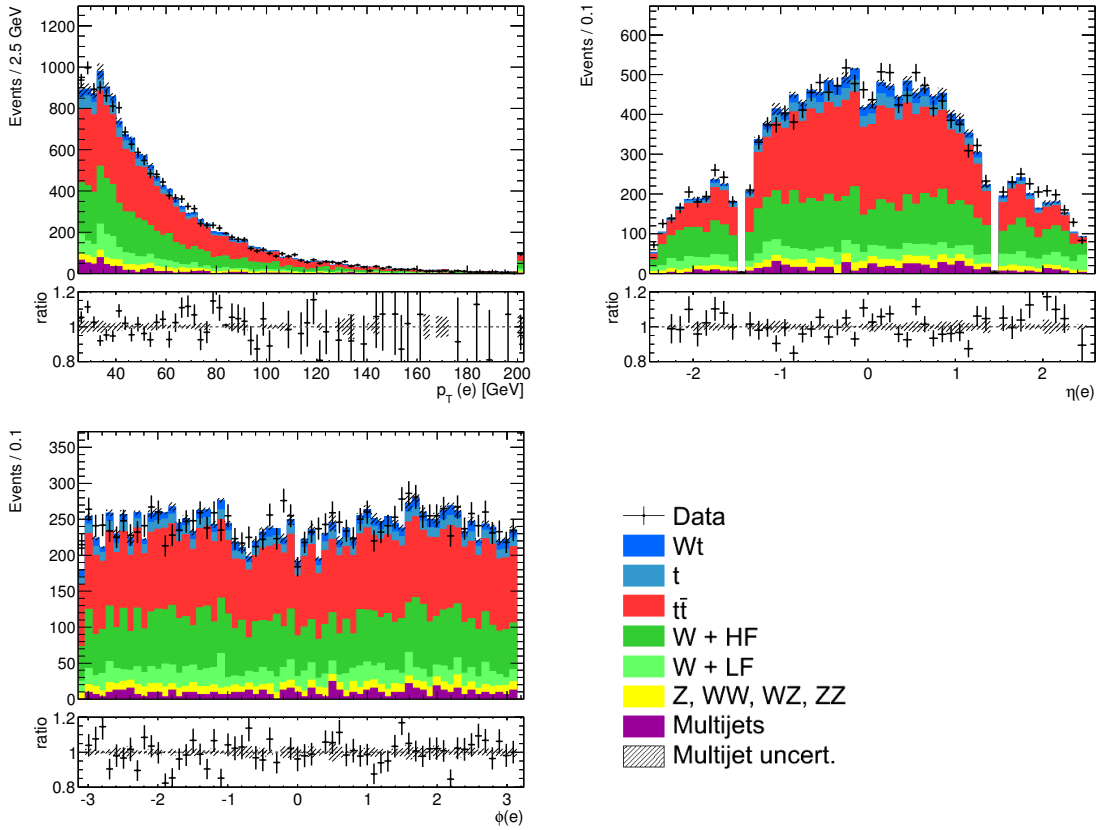


Figure 7.3: Control distributions of selected electron variables. The simulation distributions are normalized to the measured luminosity. The hashed area corresponds to the uncertainty resulting from the mulitjet estimation. The ratio of data and simulation is shown in the bottom plot.

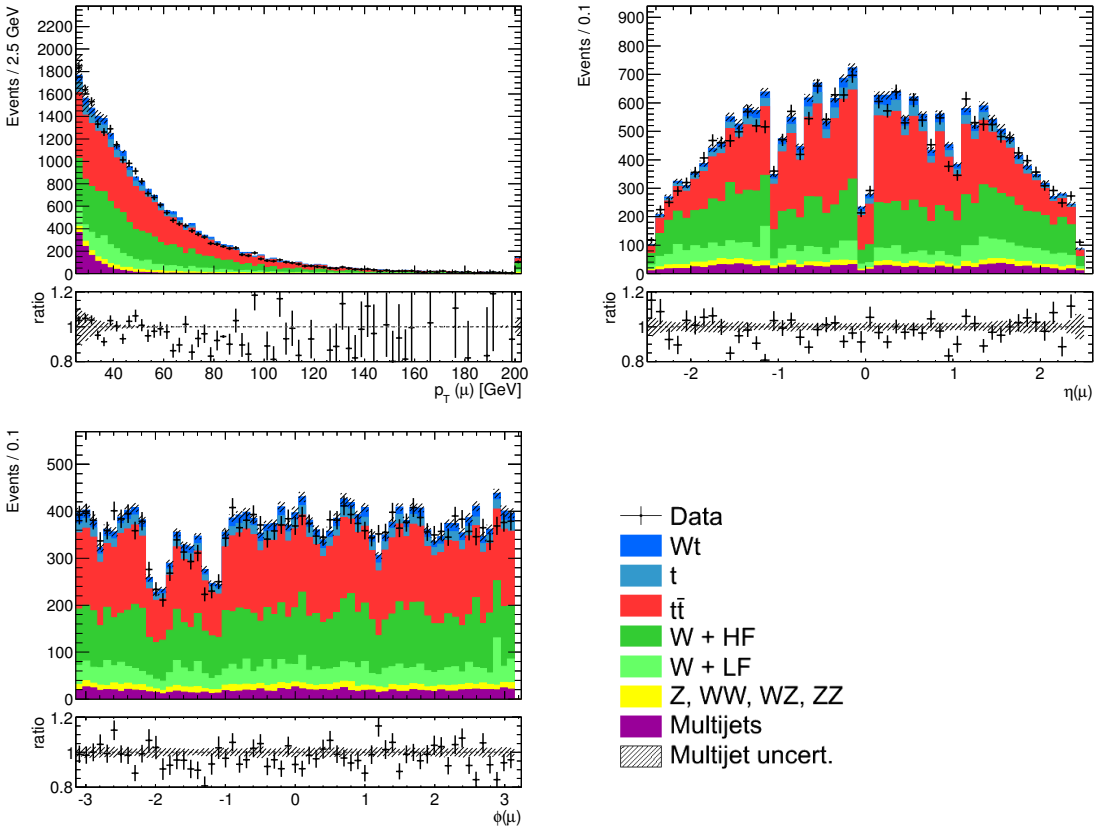


Figure 7.4: Control distributions of selected muon variables. The simulation distributions are normalized to the measured luminosity. The hashed area corresponds to the uncertainty resulting from the mulitjet estimation. The ratio of data and simulation is shown in the bottom plot.

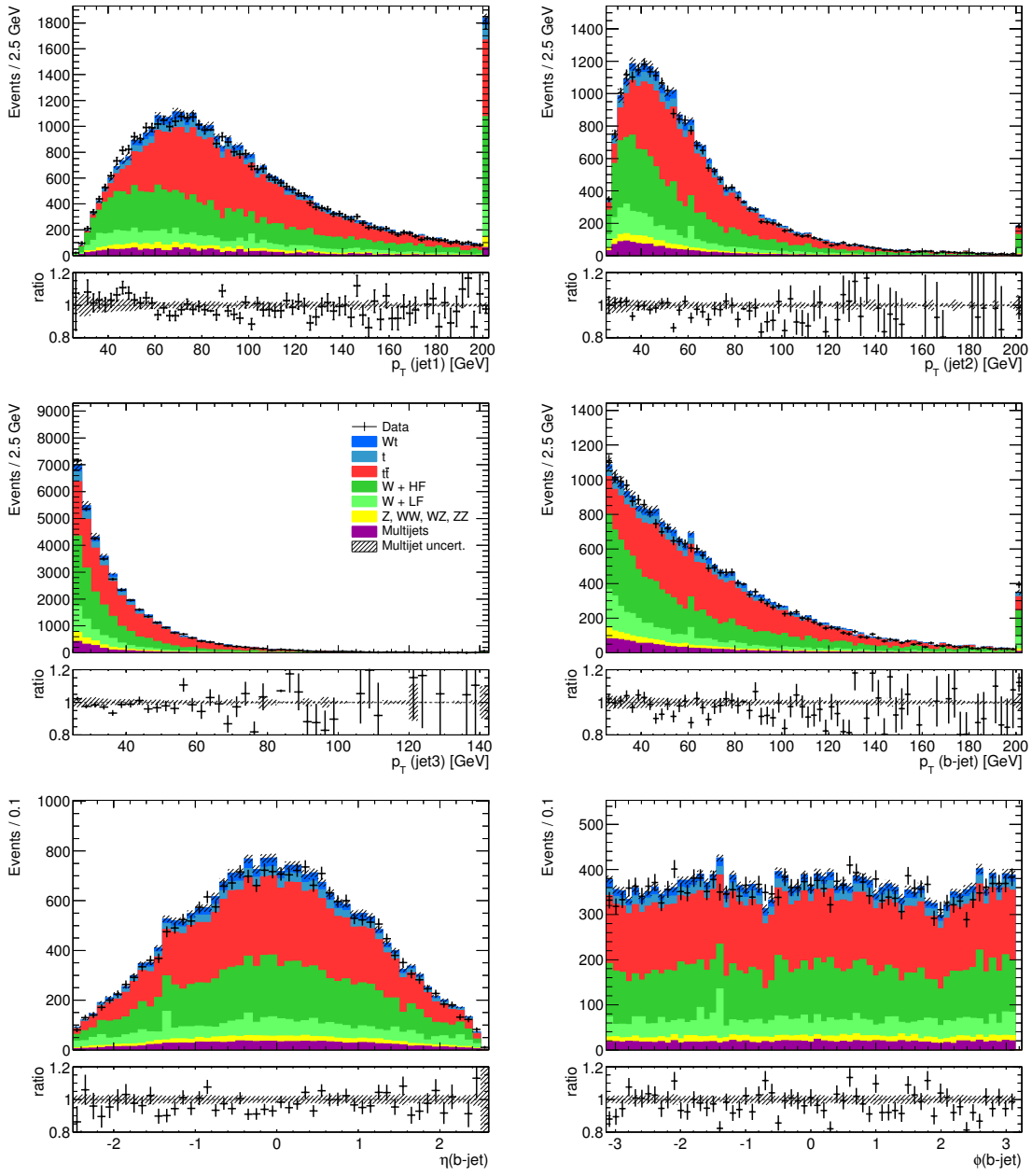


Figure 7.5: Control distributions of selected jet variables. Electron and muon channels are merged. The jet numbering follows the p_T ordering of the jets. Jet1 is the hardest jet in the event. The simulation distributions are normalized to the measured luminosity. The hashed area corresponds to the uncertainty resulting from the mulitjet estimation. The ratio of data and simulation is shown in the bottom plot.

CHAPTER 8

Construction of Discriminating Variables

To extract the single top-quark production in the Wt-channel it is important to find variables whose distributions show different shapes for the signal compared to the backgrounds. This is called the shape of a variable in the following. As the production of W+jets and top-quark pairs are the most important backgrounds, the focus on the search for good variables lies on these two. In the comparison of the variables' shapes, the other backgrounds will be merged into one distribution to allow a fast judgement using the plots. Two general paths are taken to construct variables: kinematic variables of the Wt decay particles and event property variables.

8.1 Wt decay kinematics

A first set of variables has been shown in Chapter 7 already: the transverse momenta, pseudorapidities and polar angles of the final-state leptons and jets. These variables were also used in the preselection and tag selection to suppress backgrounds significantly. Combining several particles one can also construct angular differences in η, ϕ and $\eta-\phi$ space as well as ratios of transverse momenta. An example for such a distribution is shown in the left plot in Figure 8.1 with the angular difference in the $\eta-\phi$ plane between the two light-quark jets. As in the following, the short-notation of the variable is the mathematical symbol followed by the objects which it applies to in parentheses. The shape of the distributions of this variable suggests high separation power against W+jets production as well as some separation from the $t\bar{t}$ background. The ATLAS reconstruction software provides the full four-vectors of final-state objects, which can be used to calculate additional kinematic variables. An example is the invariant mass of an object, defined as

$$M = \sqrt{E^2 - p_x^2 - p_y^2 - p_z^2}.$$

The invariant masses in turn can be used to derive variables for mass differences and ratios of different objects. For some tools it is easier to use the absolute values of the variables instead of the pure ones due to certain symmetries. The absolute values of many variables are therefore also taken into account in the search for the best variables. The right plot of Figure 8.1 shows the absolute value of the distance in η between the harder light-quark jet and the b-quark jet, $|\Delta\eta(\text{l-jet1}, \text{b-jet})|$. A good separation can be observed between signal and the W+jets production as well as the other backgrounds, while the shapes of the Wt and $t\bar{t}$ productions coincide. As the signal distribution is closer to zero in $\Delta\eta$, the signed distributions would show two crossings of the signal and the background shapes. A cut-based selection

could exploit the signed variable by applying two separate cuts, but other techniques can possibly benefit from the simpler shape of the distribution.

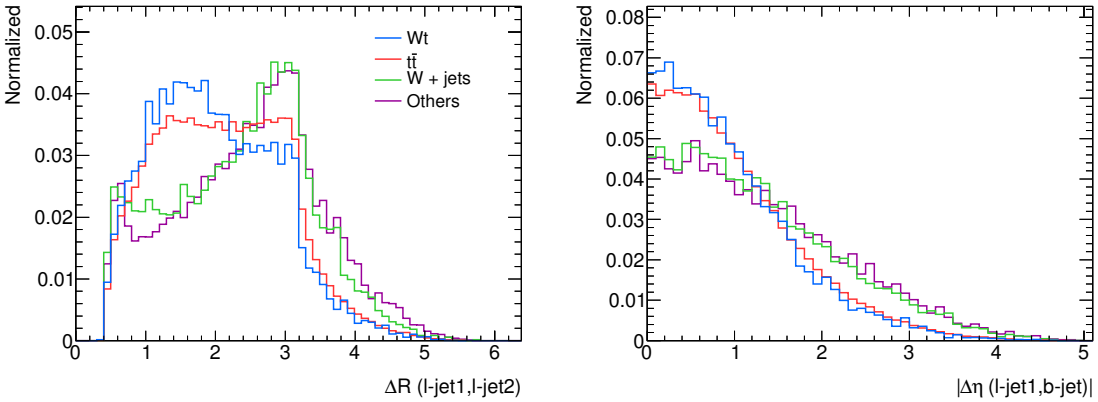


Figure 8.1: Distributions of ΔR between the two light-quark jets in the event (left) and the absolute $\Delta\eta$ between the harder light-quark jet and the b-quark jet (right) in the three-jet category normalized to unit area. Distributions are shown for the Wt signal and the two major backgrounds – the other backgrounds are merged.

Using the expected decay topology of the Wt-channel production as hypothesis, the above variables can all be extended to mother particles of the final-state ones. If there are at least three jets in the final state, one of which should be tagged as a b-quark jet, the full topology can be reconstructed. The only missing object in the expected final state is the neutrino which results from the leptonic decay of one of the W bosons. Assuming that the expected neutrino is the major source of missing transverse momentum, the components transverse to the beam axis can be reconstructed, while the component in z-direction remains unknown. To compute this missing parameter, a W-boson mass constraint is applied using the current PDG W-boson mass value. A detailed description and discussion of the calculation of $p_z(\nu)$ using this technique can be found in [138]. Figure 8.2 shows control distributions of key kinematics of the neutrino. The measured data is described very well by the simulation with only small deviations which are within the uncertainties.

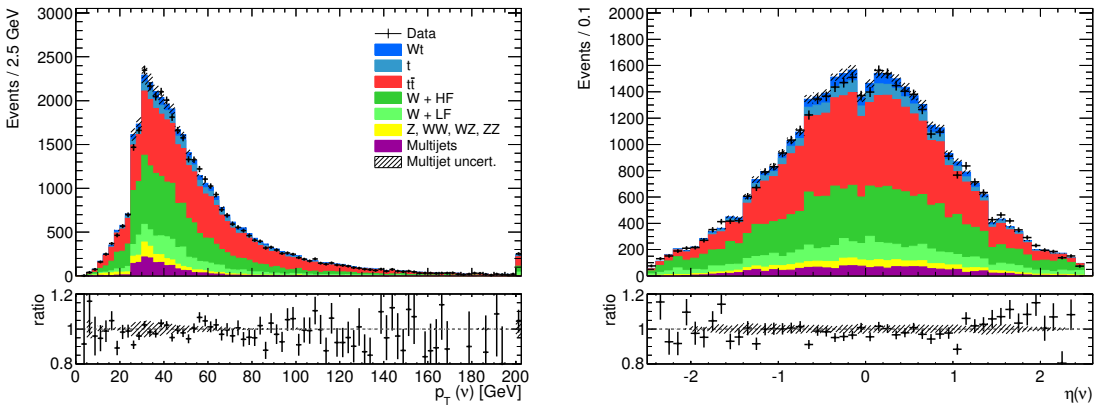


Figure 8.2: Control distributions of the transverse momentum (left) and pseudorapidity (right) of the neutrino. The neutrino p_z is reconstructed as explained in the text. The simulation distributions are normalized to the measured luminosity. The hashed area corresponds to the uncertainty resulting from the mulitjet estimation. The ratio of data and simulation is shown in the bottom plot.

With the knowledge of all four-vectors, the reconstruction of the full Wt decay topology can be carried out as follows. The hadronically decaying W boson, W^{had}, corresponds to the combination of the two light-quark jets, l-jet1 and l-jet2, while the leptonically decaying W boson, W^{lep}, corresponds to the combination of the lepton and the reconstructed neutrino. Together with the b-tagged jet, they can in turn be combined to reconstruct the initially produced top quark. It is per se not possible to distinguish if the top quark decays into a hadronic final state or a final state including a lepton. It may be possible to resolve this ambiguity using kinematic properties of the event. As this would only be possible with a certain efficiency, which is not expected to be very high, this analysis uses the inclusive approach where both hypotheses are always used in the reconstruction. In this way, all variables are always calculated for each event. Of course this results in combinatorial background, which can reduce the efficiency of a variable in terms of the separation of signal and background.

Using the four-vectors of all eight expected objects (t, W^{lep}, W^{had}, b-jet, l, ν, l-jet1, l-jet2) and additionally of the full combined Wt system, Wt, a large number of kinematic variables is constructed in the same way as explained for the final reconstructed objects. Two examples for the variables are shown in Figure 8.3: the pseudorapidity of the reconstructed leptonically decaying W boson, $\eta(W^{\text{lep}})$, and the ΔR between the b-quark jet and the reconstructed hadronically decaying W boson, $\Delta R(\text{b-jet}, W^{\text{had}})$. Both variables differ for signal and $t\bar{t}$ production compared to W+jets production and the other backgrounds. The combined distribution of other backgrounds has a shape between the other distributions as they are composed of the other single top-quark events, which behave like the Wt-channel and $t\bar{t}$, and diboson as well as multijet events. The shape of the distribution of the latter ones are similar to the W+jets production. The differences between invariant W-boson masses and their current PDG values are also evaluated as well as the mass difference of the top quark to the one set in the Monte Carlo generator of $M(t) = 172.5\,\text{GeV}$.

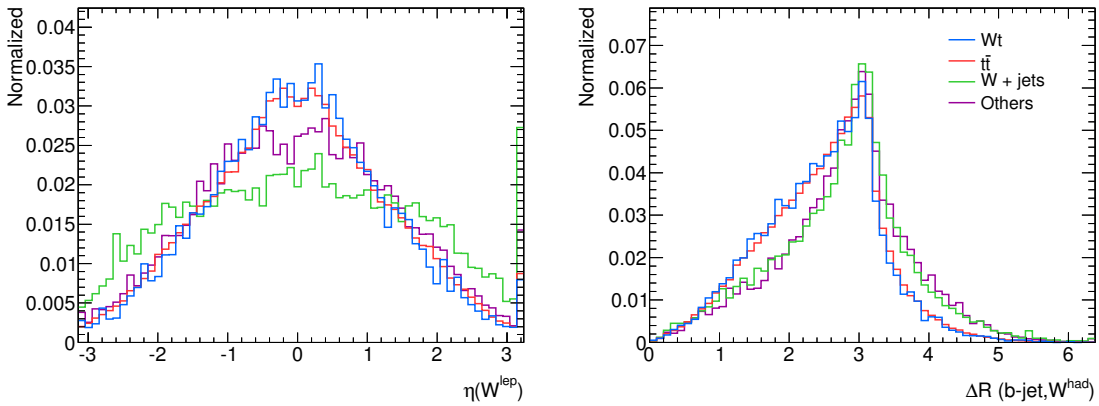


Figure 8.3: Distributions of the pseudorapidity of the reconstructed leptonically decaying W boson (left) and ΔR between the b-quark jet and the reconstructed hadronically decaying W boson (right) in the three jet category normalized to unit area. Distributions are shown for the Wt signal and the two major backgrounds – the other backgrounds are merged.

As a unique assignment of reconstructed objects to expected decay particles is only possible in the three-jet category, events with an additional or a missing jet have to be treated separately. In the two-jet category, a b-tagged jet and a light-quark jet are required. To anyway reconstruct the Wt system as far as possible, the single light-quark jet is treated as a hadronically decaying W-boson in the reconstruction of the decay. Of course this is just a hypothesis and variables like the mass of the said W boson are no longer sensible to use. It is possible to create a number of kinematic variables which can be used to

separate signal and background events. A detailed study of this category has been presented in [139]. In events with four jets in the final state, one of which is tagged as a b-quark jet, the reconstruction of the full decay system remains possible. The light-quark jets are assigned to the two expected ones in the order of their transverse momentum. The two hardest non-tagged jets are used to reconstruct the hadronically decaying W boson, while the softest jet remains as single object. This simplified ansatz is acceptable as the hardest objects in the event are expected to result from the decay of the top quarks. Resolving the combinatorics would again require kinematic assumptions, similar to the disentanglement of the leptonic and hadronic top-quark decay mentioned above, with possibly limited efficiency. As the four-jet category is mostly $t\bar{t}$ background-dominated, its separation potential is also very limited as the kinematics are very similar to the Wt-channel.

8.2 Event variables

While the first approach allows a large number of possible separating variables to be constructed using only few objects, the second class of variables is built from many or all particles in the final state of the event. One such variable, the missing transverse momentum, has already been introduced in detail and also been used in the selection. Another two variables are built from the reconstructed momenta of final-state particles: the sum of the transverse momenta of the charged lepton and all jets in the event (H_T) and the sum of transverse momenta divided by the sum of the momenta of the charged lepton and all jets, which is called centrality. Both distributions are shown in Figure 8.4 for the three-jet category normalized to unit area. The centrality discriminates well between signal and W+jets events as well as the less prominent backgrounds while the H_T distributions have slight shifts for each sample. The distribution of the Wt signal lies in-between the W+jets and the $t\bar{t}$ backgrounds as is the case for many other variables, too.

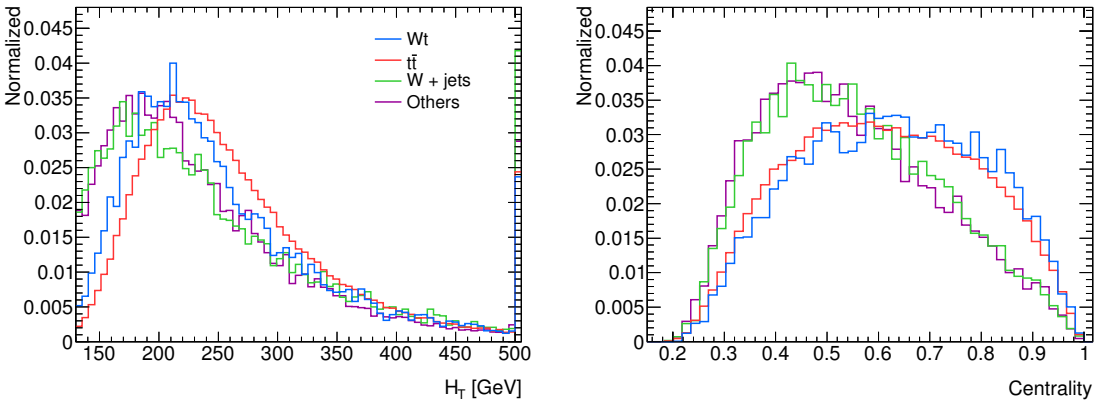


Figure 8.4: Distributions of H_T (left) and centrality (right) in the three jet category normalized to unit area. Distributions are shown for the Wt signal and the two major backgrounds – the other backgrounds are merged.

Finally, two other variables which can parametrize the kinematic shape of events are the sphericity and the aplanarity of the event. Both variables are calculated from the sphericity tensor ($S^{\alpha\beta}$), which is defined as

$$S^{\alpha\beta} = \frac{\sum_i p_i^\alpha p_i^\beta}{\sum_i |\vec{p}_i|^2},$$

where α and β are all permutations of the three momentum components and the momenta of the lepton

and all jets are again taken into account. This tensor has eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$ which follow $\lambda_1 + \lambda_2 + \lambda_3 = 1$. Using these eigenvalues, the sphericity is defined as

$$S = \frac{3}{2}(\lambda_2 + \lambda_3)$$

and ranges between zero, which corresponds to an event with two back-to-back jets, up to one, which corresponds to an isotropic event. The normalized distribution of the sphericity of three-jet events is shown in Figure 8.5. The events from Wt and $t\bar{t}$ production are slightly more isotropic than other events.

The aplanarity is defined as

$$A = \frac{3}{2}\lambda_3,$$

ranging from planar events at values close to zero to isotropic events with $A \approxeq 0.5$. In general the events are planar, as can be seen from the normalized distribution of the aplanarity of three-jet events shown in Figure 8.5. Again events with Wt and $t\bar{t}$ production show a more isotropic shape than others.

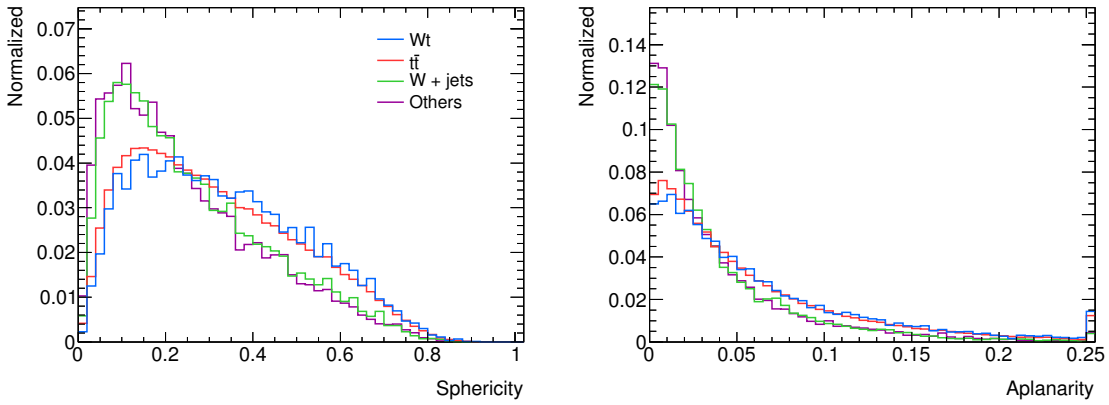


Figure 8.5: Distributions of sphericity (left) and aplanarity (right) of events in the three jet category normalized to unit area. Distributions are shown for the Wt signal and the two major backgrounds – the other backgrounds are merged.

8.3 Best discriminating variables

As motivated in the beginning of this chapter, the goal is to find variables with high discrimination power between signal and background events. It is also desirable to limit the amount of used variables to evade large correlations and reduce the impact of systematic variation effects on the measurement. To judge the discrimination power, all backgrounds are merged into a single distribution according to their cross-sections. The full background distribution is then normalized to the already luminosity-normalized signal distribution. The comparison of the two distributions allows the variables with the best separation between the single top-quark production in the Wt-channel and all its backgrounds to be found. The best separation is achieved for variables based on the mass of reconstructed primary particles, like W bosons and top quarks.

Two examples for such variables are shown in Figure 8.6 with the reconstructed hadronically decaying W boson, $M(W^{\text{had}})$, obtained from the combination of the two light-quark jets and the hadronically decaying top quark, $M(t^{\text{had}})$, which is reconstructed using all three jets present in the event. The distributions of the signal both show narrower peaks at the expected particle masses compared to the back-

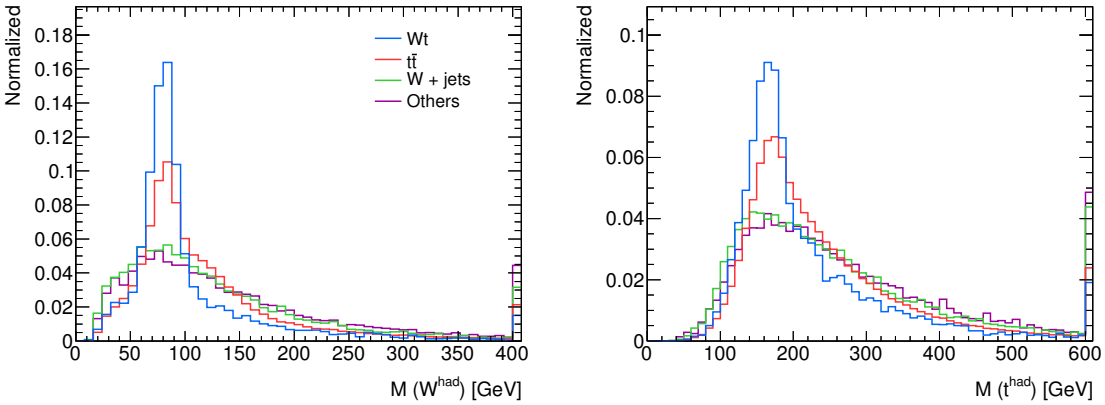


Figure 8.6: Distributions of the invariant masses of the reconstructed hadronically decaying W boson (left) and reconstructed hadronically decaying top quark (right) in the three jet category normalized to unit area. Distributions are shown for the Wt signal and the two major backgrounds – the other backgrounds are merged.

ground distributions. While the $t\bar{t}$ production still shows a peak, it vanishes in the distributions of the other backgrounds although they still show a maximum at the peak position. As the W+jets production only includes leptonically decaying W bosons, no peak is expected in the two distributions. In the other backgrounds, hadronically decaying W bosons are only expected in the diboson samples which include W bosons (WW, WZ). As shown in the previous chapter, these backgrounds are heavily suppressed by the applied kinematic selection.

Another two examples, which are shown in Figure 8.7, are mass ratios of reconstructed particles. The invariant mass of the hadronically decaying W boson is divided by the invariant mass of the leptonically decaying top quark, $M(t^{\text{lep}})$, which is reconstructed from the charged lepton, the neutrino and the b-tagged jet. This ratio again shows the most pronounced peak for the signal distribution while the backgrounds show a similar behaviour as before. The full Wt system is reconstructed from all final state particles found in the event and its invariant mass is divided by the invariant mass of the two light-quark jets. This variable shows a peaked region for all distributions, but the peak moves from smaller ratios for the less important backgrounds to a maximum for the signal distribution. All these variables have the advantage of an isolated behaviour of the signal distributions in the sense that their shapes do not lie in-between the two major backgrounds. The variables shown in the previous sections did not show this isolated behaviour which suggests that a separation of the $t\bar{t}$ background is possible.

In general it must be said that even the variables with the largest shape-differences between the distributions of events from Wt production and distributions of background events are not sufficient to isolate the signal by applying direct cuts. The kinematic differences, especially between the signal and the top-quark pair production, are not large enough to isolate the signal, and the small cross-section compared to the background channels prevent the application of a multitude of cuts as the statistics would drop too much. Possible solutions for this problem are the construction of super-variables using more elaborate combinations of the introduced variables, the application of multi-dimensional cuts and finally the application of other multivariate analysis techniques. The last approach is used in this analysis.

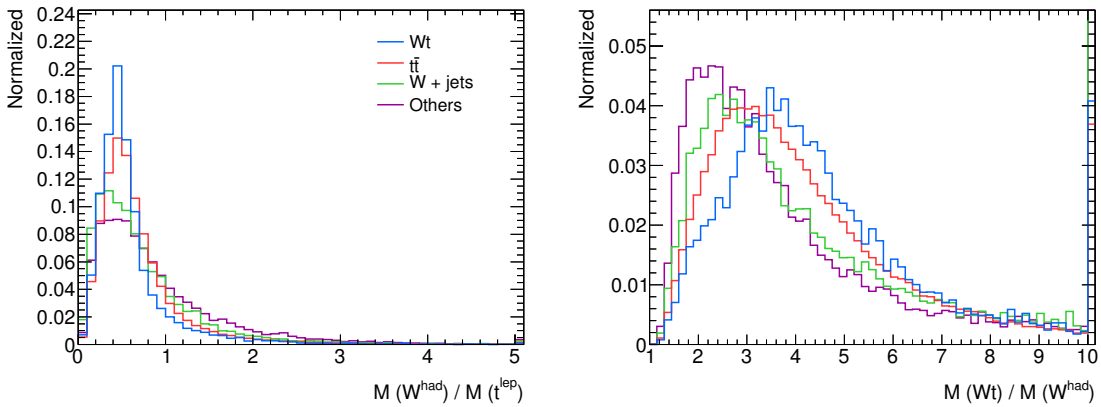


Figure 8.7: Distributions of the ratios of invariant masses of the reconstructed hadronically decaying W boson and the leptonically decaying top quark (left) and the full reconstructed Wt system and the hadronically decaying W boson (right) in the three jet category normalized to unit area. Distributions are shown for the Wt signal and the two major backgrounds – the other backgrounds are merged.

CHAPTER 9

Improved Signal Isolation Using Neural Networks

While the analysis approach of cutting on single variables to enhance the number of signal events over the number of background events is very simple and allows one to effectively execute measurements in many high-energy physics channels, it is also limited to very special conditions. If it is possible to find variables with large shape differences between signal and background events, this method can produce results using a well-controlled chain of cuts. Many searches for new physics processed at the LHC show very distinct signatures, which can be well isolated using cut-based analyses. As the investigations in Chapter 8 have shown, the search for single top-quark production in the Wt-channel is very different in two regards: the kinematics of the particles and also the event variables are very similar to those found in the background channels and the cross-section is small compared to the main backgrounds. To overcome these problems, multivariate techniques based on machine learning, including artificial neural networks, have recently been used successfully in high-energy physics analyses. Examples for other analyses using the NeuroBayes neural network software, which is used in this analysis, can be found in the field of single top-quark production in ATLAS [140, 141].

In the first part of this chapter (Section 9.1), NeuroBayes neural networks are introduced and their advantages are discussed. The following Section 9.2 shows the creation of a single discriminating variable using NeuroBayes. The inputs and outputs of the software package are exemplified and the performance is evaluated. The final Section 9.3 of this chapter shows the combination of two neural network trainings to improve the discrimination power against the very different major backgrounds.

9.1 Artificial neural network and preprocessing - NeuroBayes

The following explanations are based on the overview of multivariate techniques by Helge Voss [142]. More detailed information about the NeuroBayes package are taken from [143, 144] and the NeuroBayes user's guide [145].

Multivariate techniques exploit the whole variable space, which is spanned by all employed input variables, to find a single event classifier which separates signal and background events as effectively as possible. This approach, in contrast to isolated cuts, can use correlations between the input variables for a better classification. A further advantage of the application of a neural network over many other techniques is the usage of non-linear functions to build the classifier. Such a classifier can model boundaries

between signal and background distributions more accurately and lead to large improvements. In many analyses it has been observed that artificial neural networks achieve a higher classification efficiency at the same rate of false positives compared to simpler methods.

9.1.1 The neural network classifier

In the following, the construction of the event classifier is explained in a bit more detail focussing on the concrete implementation in NeuroBayes. There are of course many other implementations of artificial neural networks where different approaches are used in some steps.

The neural network classifier, y, can mathematically be described as a superposition of non-linear functions, h_m,

$$y(\vec{x}) = w_0 + \sum_{m=1}^{M} w_m h_m(\vec{x}),$$

where $\vec{x}$ is a vector of all used input variables. The dimension M of this superposition can be chosen freely, which allows complex classifiers to be modelled. The coefficients w_i, which are called weight factors, have to be determined. Instead of using different functions h_m it is also possible to use only one non-linear function h, transforming the output to

$$y(\vec{x}) = w_0 + \sum_{m=1}^{M} \left[w_m^{II} h\left(w_{0m}^{I} + \sum\nolimits_{i=1}^{D} w_{im}^{I} x_i \right) \right],$$

where D is the dimension of the variable space, i.e. the number of input variables. Weight factors with superscripts I are the ones used in the previous equation while the factors with superscript II is a scale of the non-linear function. Activation functions are used for the non-linear function h. To match the correct working points of the activation function, the additional bias w_{0m}^{I} has been introduced, which scales the linear combination of all input variables. The NeuroBayes software uses symmetric sigmoid functions which are defined by:

$$h(x) = S(x) = \frac{2}{1 + e^{-x}} - 1.$$

An example for such a function is shown in Figure 9.1. It maps the whole range of x to an interval of $[-1, 1]$. Additionally, in NeuroBayes, another sigmoid function is applied to the output and no overall weight w_0 is used. The classifier, or neural network output, is then defined by

$$y(\vec{x}) = S\left(\sum_{m=1}^{M} w_m^{II} S\left(w_{0m}^{I} + \sum\nolimits_{i=1}^{D} w_{im}^{I} x_i \right) \right).$$

In the binary classification used in this analysis, the classifier of the network is expected to have a value of -1 for background events and +1 for signal events. This expectation is called the neural network target. While the target is discrete, the calculated neural network output is a continuous distribution between -1 and +1 if NeuroBayes is used in binary classification mode. A cut on the neural network output can classify events into signal or background events with a certain efficiency. Such a cut represents a non-linear hyperplane in the variable space.

It is possible to create a graphical representation of the neural network output calculation, an example of which is shown in Figure 9.2. Each sigmoid corresponds to an internal node of the network and for each weight factor, a connection between the corresponding nodes is drawn. The superscripts of the weight factors in the above equations indicate the interconnection layers in Figure 9.2. The shown

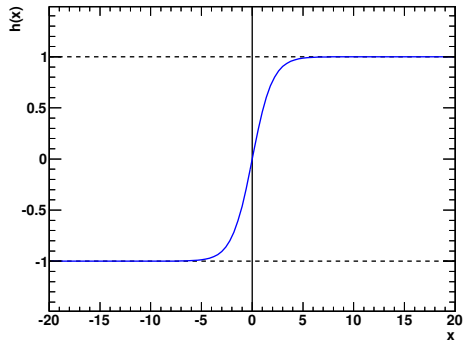


Figure 9.1: Example of a symmetric sigmoid function. Compared to the usual function it is shifted to project to the range of -1 to 1.

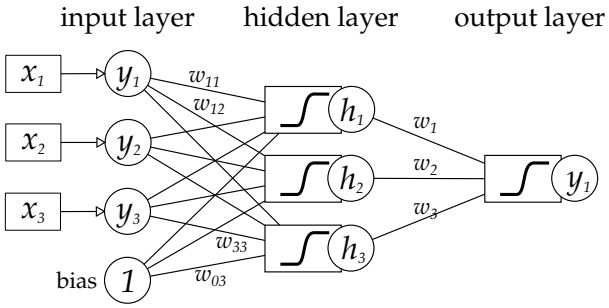


Figure 9.2: Graphical representation of a neural network with three input variables, three internal nodes in the hidden layer and a single output node. Each internal node corresponds to a symmetric sigmoid function. and each connection to a weight factor.

network consists of three input variables, three internal nodes in the hidden intermediate layer and one output node in the output layer and has the typical structure of a neural network used in the NeuroBayes software package. The graphical representation shows why such mathematical structures are called neural networks as they resemble the functionality of neurons in central nervous systems in nature. A node works like a neuron which receives a signal from another neuron and transmits the signal further based on a decision. The binding strength between neurons is given by the weight factors. The special network shown here is called feed-forward network or multilayer perceptron (MLP), because it contains no loops which bind nodes backward to previous layers.

9.1.2 Training of the neural network

As mentioned above, the weight factors in the network have to be determined to build an effective classifier. This step is accomplished in the training procedure of the neural network. The figure of merit for this procedure is a loss function, which is minimised. NeuroBayes uses the entropy loss function

$$E = \sum_{n=1}^{\text{events}} \log\left(\frac{1}{2}(1 + y_n^{\text{true}} \cdot y_n + \epsilon)\right),$$

where y_n^{true} is the target output for event n. As this function is too complicated for the application of standard minimization algorithms, simulation events are used to teach the signal and background distributions to the neural network, which in turn uses a step-wise minimisation of the loss function called back-propagation. Starting with random weights Gaussian-distributed around zero, the gradient of the loss function is calculated as a function of the weight factors for the events. The weight factors are then adjusted to reduce the loss function's size. NeuroBayes uses a procedure called mini-batch or quasi-online learning, which evaluates the loss function after 200 events only. This can drastically speed up the training procedure and is often precise enough. The simple structure of the network also becomes beneficial at this point, as it allows a fast calculation of the loss function gradient.

A risk in the learning procedure is the so-called overtraining. If the number of internal network nodes is chosen to be too large, the network training can create a classifier which follows statistical fluctuations in the signal and background distributions. This finally leads to a smaller efficiency of

the network than the best possible one, as the underlying true distribution is not modelled well. To prevent overtraining, NeuroBayes uses a Bayesian regularisation procedure. Essentially, regularisation parameters are included into the loss function to reduce the impact of weight factors that become very large. More details about this procedure can be found in [143].

9.1.3 Preprocessing of the input variables

A large advantage of the NeuroBayes software over other neural network implementations is the preprocessing module for the input variables. It executes three steps before handing the input variables over to the neural network described above: a variable transformation, the choice of the best separating variable set and the decorrelation of the input variables. In a first step, all continuous input variables are transformed into a flat distribution using a non-linear transformation. The result is a histogram with a variable bin width, where the sum of the number of signal and background events is the same over all bins. The transformation removes a possible saturation of single nodes in the later training by strong outliers. Such saturations would lead to a single node dominating the final network. The flattened distributions are used to derive a purity distribution using the definition

$$P_j = N_j^{\text{sig}} / \left(N_j^{\text{sig}} + N_j^{\text{bkg}} \right)$$

for the purity in each bin j calculated from the number of signal events (N_j^{sig}) and number of background events (N_j^{bkg}) in the respective bin. To remove a possible dependence on statistical fluctuations, a spline is fitted to the purity distribution. Finally the spline is transformed to a Gaussian distribution with mean zero, mapping the respective input variable to the same range as all other variables. This step facilitates the later network training. After this equalization process, the variables can be well compared.

To judge the quality of the variables, the correlation matrix of the input variables is calculated and the total correlation to the target (the true assignment to -1 for background and +1 for signal events in case of the binary classification) is calculated for all input variables. Afterwards an iterative ranking procedure is started. One variable is removed from the full set of variables and the total correlation to the target is calculated for the remaining variable set. This defines a loss in correlation. The variable leading to the smallest loss in total correlation to the target is then removed and the next iteration starts. This process creates a well-defined ranking procedure of variables. The analyser can then define which maximal loss in correlation to target is affordable and the software feeds only the variables into the neural network training whose removal leads to a larger loss. The significance of a variable is defined as the loss of global correlation caused by its removal at the corresponding step in ranking multiplied by $\sqrt{n}$, where n is the number of input events and expressed in units of standard deviations, σ, of the loss compared to the target. Other variables which are calculated in the process of the preprocessing are:

- *Single significance* – is the correlation to the target multiplied by $\sqrt{n}$ for only this variable without taking other variables into account. Using this variable, the NeuroBayes preprocessing can be exploited to find well-discriminating variables in a large variable set.
- *Significance loss* – is the loss in significance if only this variable is removed from the full input set. It corresponds to the loss in significance as defined above in the first iteration step of the ranking procedure.
- *Global correlation* – is the correlation of a variable to all other input variables.

This ranking procedure ensures that correlations between variables are taken into account. It often occurs that variables only have a high significance in combination with others.

After the variable transformation and ranking procedure, the finally chosen input variables for the neural network training are decorrelated as described in [143] which is again beneficial for the training procedure. The transformed, ranked and decorrelated set of input variables are then fed to the neural network training described above and the output of the trained neural network can be used as discriminating variable for the statistical analysis. More information about the training quality and the output interpretation are explained at the point of occurrence in the first network training, which is described in the next section.

9.2 Neural network training of signal against backgrounds

In a first attempt to create a good classifier for the isolation of the single top-quark production in the Wt-channel, a neural network is trained using simulated events for the signal and background production channels described in Chapter 5. The multijet background events determined by the data-driven method are not included in the training as they only play a minor role and the shapes of the multijet distributions resembles the ones of other backgrounds. The simulated events of all backgrounds are first reweighted according to their respective cross-sections as provided by the theory calculations. In a second step, the background events are reweighted to match the number of signal events. This avoids prior assumptions on the signal cross-section. The NeuroBayes preprocessing identifies the combination of variables which achieves the best separation of signal and background as detailed above. In this training, it is configured to only use variables with a significance of more than $3\,\sigma$. This configuration leads to a network with 14 input variables, which are listed in Table 9.1, sorted by their significances. The distributions of all used input variables can be found in Appendix B.1.

Variable	Significance [σ]	Single significance [σ]	Significance loss [σ]	Global correlation [%]
$M(W^{\text{had}})$	42.8	42.8	30.1	50
$\eta(l)$	17.0	18.6	11.4	36
p_{T}(jet1)	12.0	16.8	8.3	35
$\Delta\eta$(b-jet, W^{lep})	10.4	17.7	5.8	38
ΔR(l-jet1, t^{lep})	9.8	17.7	8.3	43
ΔM(b-jet, t^{had})	8.2	30.9	8.9	73
$\Delta M(Wt, t^{\text{lep}})$	7.6	17.6	8.6	73
ΔM(l-jet1, b-jet)	8.7	13.1	6.1	54
ΔM(l-jet1, t^{lep})	7.5	12.6	6.7	26
$p_{\text{T}}(l)$	7.5	8.8	6.7	31
$\Delta\phi(\nu, W^{\text{had}})$	6.3	8.6	5.7	29
ΔM(l-jet2, b-jet)	6.0	14.7	6.1	44
ΔR(l-jet2, b-jet)	5.5	13.5	5.5	22
ΔR(l-jet1, b-jet)	5.4	20.5	5.4	53

Table 9.1: Variables selected by the NeuroBayes preprocessing in the training of the signal against all backgrounds combined. The table is sorted by the significance of the variables. Significance, single significance, significance loss and the global correlation of each variable are given as defined in the text in Section 9.1.

The variable with the highest significance is the mass of the reconstructed hadronically decaying W boson, $M(W^{\text{had}})$, followed by the pseudorapidity of the lepton, $\eta(l)$, and the transverse momentum of

the hardest jet in the event, p_{T}(jet1). The list is dominated by mass differences. As explained before, the single significance is the same as the significance for the best variable, while it differs for the following ones. The table clearly shows that the variables do not necessarily need to have a high single significance to contribute to the overall network. This is a consequence of variable correlations: combinations of variables can reach a larger significance than the single ones. This is also shown by the least significant variable in the list, ΔR(l-jet1, b-jet). While it has a large single significance of 20.5 σ, the combination of other variables with much lower single significances outweigh its importance. To get an impression of the correlations, the covariance matrix is visualized in Figure 9.3.

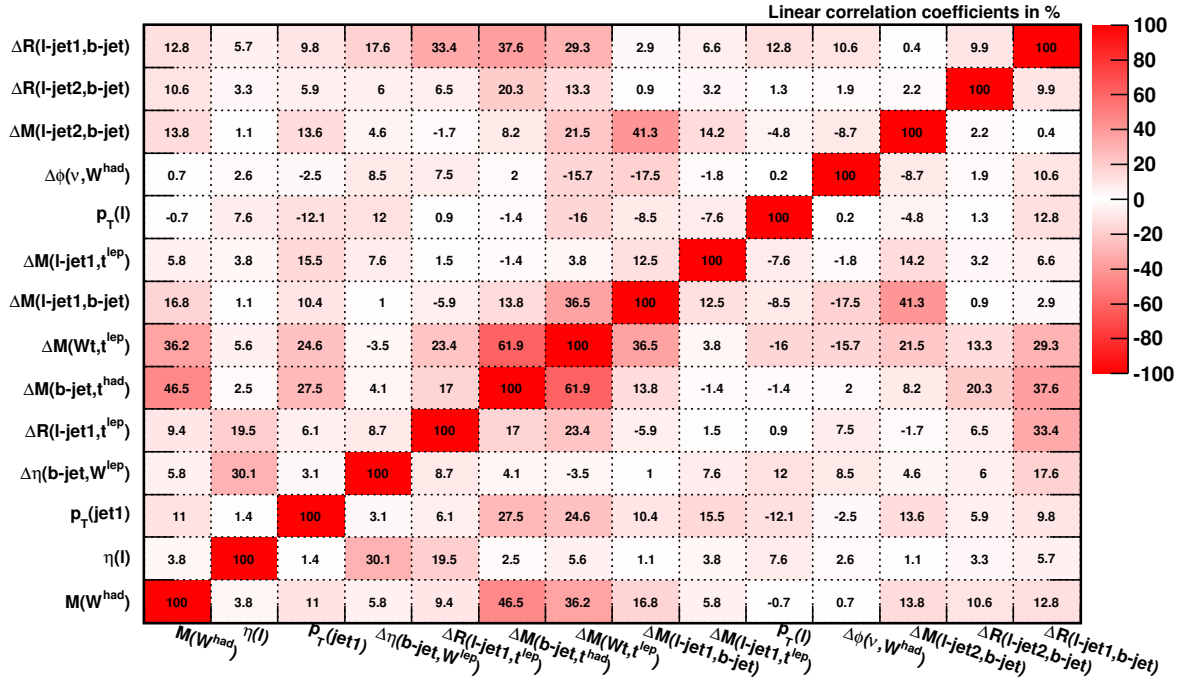


Figure 9.3: Covariance matrix of the variables selected by NeuroBayes preprocessing in the training of the signal against all backgrounds. The variables are sorted by their significance on the x-axis.

The largest correlation of 62 % is found between the mass difference of the reconstructed Wt-system and the leptonically decaying top quark, $\Delta M(Wt, t^{\mathrm{lep}})$, on one hand and the mass difference of the b-quark jet and the hadronically decaying top quark, ΔM(b-jet, t^{had}), on the other hand. The two variables with the second largest correlation (47 %) are again the mass difference of the b-quark jet and the hadronically decaying top quark and the mass of the hadronically decaying W boson, $M(W^{\mathrm{had}})$. These correlations are understandable from the particle combinations used in the variable calculations. The variable $\Delta M(Wt, t^{\mathrm{lep}})$ includes the full final state composed of the two light-quark jets, the b-quark jet and the two leptons and subtracts the mass of the leptonically decaying top quark which is reconstructed from the b-quark jet, the charged lepton and the neutrino. The remaining objects are the two light-quark jets which are used in the construction of the highly correlated variable $M(t^{\mathrm{had}})$, which is subtracted from the mass of the b-quark jet. The same argument holds for the other two variables. The correlation likely results from the two light-quark jets. Despite the large correlations, all three discussed variables still provide additional information and hence have large significances and are important for the final network. This argument is of course also valid in the opposite direction: the least important variable ΔR(l-jet1, b-jet) in the first row shows large correlations to some other variables, which is the reason for

its small global significance compared to the much larger single significance.

After the preprocessing, the neural network training is executed with a network topology of 15 input nodes, eleven internal nodes and a single output node. The training procedure calculates the weight factors of the connections which results in the network shown in Figure 9.4, where the line darkness is proportional to the weight factor and hence darker lines represent a stronger binding between the nodes. The numbering of the input nodes starts with the bias node at one and then subsequently adds all 14 input variables from node two to 15. In this training, the variables were ordered by their significance, so that the order of the input nodes in the graphical representation matches the one given in Table 9.1. It is visible that some nodes only show small weight factors which suggests that a network with fewer internal nodes would perform equally well. This has been verified in the course of the analysis by alternating the number of internal nodes over a wide range for several different network trainings. For a very small number of internal nodes, the separation of the neural network output variable breaks down, but for larger numbers, the training results have been very stable. The same has been observed in the studies in the two-jet bin performed in the framework of the presented analysis in [139]. A larger number of internal nodes slows the training procedure down significantly which is the reason for the simple choice taken here.

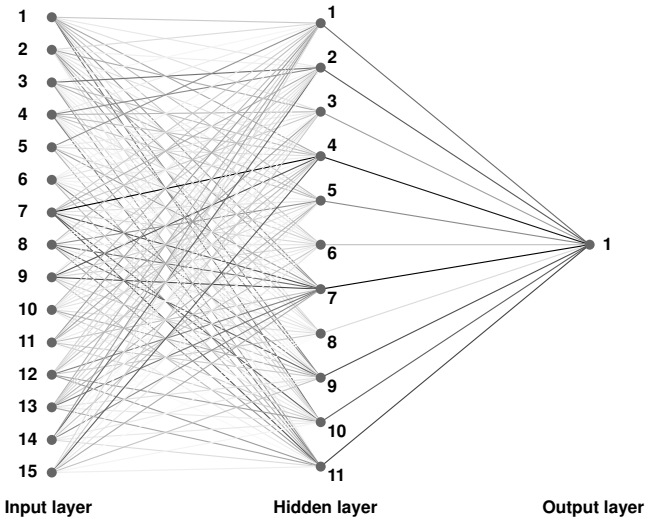


Figure 9.4: Graphical representation of the trained neural network for the training of signal against all backgrounds. The line darkness depends on the weight factors of the connections. The darker a line, the larger is the corresponding weight factor. The meaning of the numbers is discussed in the text.

Several distributions are provided by NeuroBayes to judge the quality of a training. The ones relevant for the analysis are be explained here and used subsequently in the following trainings. Figure 9.5 shows the neural network output for the training for signal and background events separately. Using the two distributions one can judge the quality of the separation of the training. Both span over the same area due to the normalization of the background events to the number of signal events before the preprocessing. Signal-like events are favoured in the region towards a network output value of 1 while background events tend to -1. A separation of signal from background events can be observed which is much stronger than the separation of any single variable evaluated in Chapter 8. The superiority of the neural network compared to single variables in the analysed Wt-channel is evident already at this point.

Despite the reasonable separation, considerable background contamination can still be observed in all filled bins, which is problematic knowing that the cross-sections of the background processes are expected to be much larger than the cross-section of the signal process predicted by the Standard Model. A final judgement is only possible after plotting the distribution using the correct scaling according to the cross-sections of all processes.

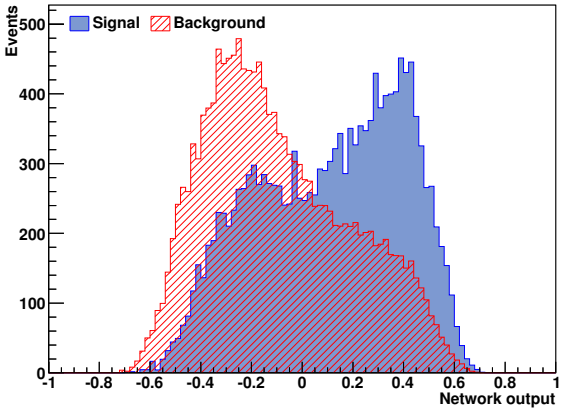


Figure 9.5: Distributions of the neural network output for signal and background events used for training. The number of background events is normalized to the number of signal events.

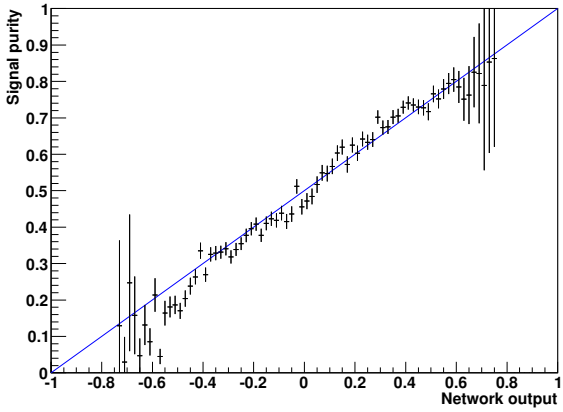


Figure 9.6: Distribution of the signal contribution (purity) as a function of the neural network output for the training of the signal against all backgrounds. The diagonal visualizes the expected linear behaviour given that the number of background events is normalized to the number of signal events.

In Figure 9.6, the purity P is plotted as a function of the neural network output, where purity is defined as the number of signal events over the number of all events. A linear distribution along the diagonal of the plot can be observed, which indicates a good training of the neural network. Deviations from the diagonal would suggest problems in the minimization process of the error function. This can be understood by writing the mean χ^2 of events with the output y according to [143] as

$$\chi^2 = P(1-y)^2 + (1-P)(-1-y)^2,$$

where the first term of the sum corresponds to the signal and the second term to the background events. In the case of an optimal training, the error function gets minimal which corresponds to a vanishing first derivative and hence leads to the relation

$$P = \frac{y+1}{2},$$

which is the linear behaviour observed in Figure 9.6.

A useful measure for the separation power of a network is the Gini index[142]. Its construction is visualized in Figure 9.7, which shows the signal efficiency as a function of the overall efficiency of the neural network output. The efficiencies are calculated applying consecutive cuts on the neural network distributions shown in Figure 9.5 from low values of the neural network output to large values. The signal efficiency is defined as the sum of all signal events to the right of this cut divided by the overall number of signal events in the training. The overall efficiency is defined accordingly with the sum of signal and background events. The corresponding blue curve in the plot is created by interpolation between the resulting points. Finally the Gini index can be calculated by dividing the area between the efficiency curve and the diagonal by the area below the diagonal. Larger indices indicate a better

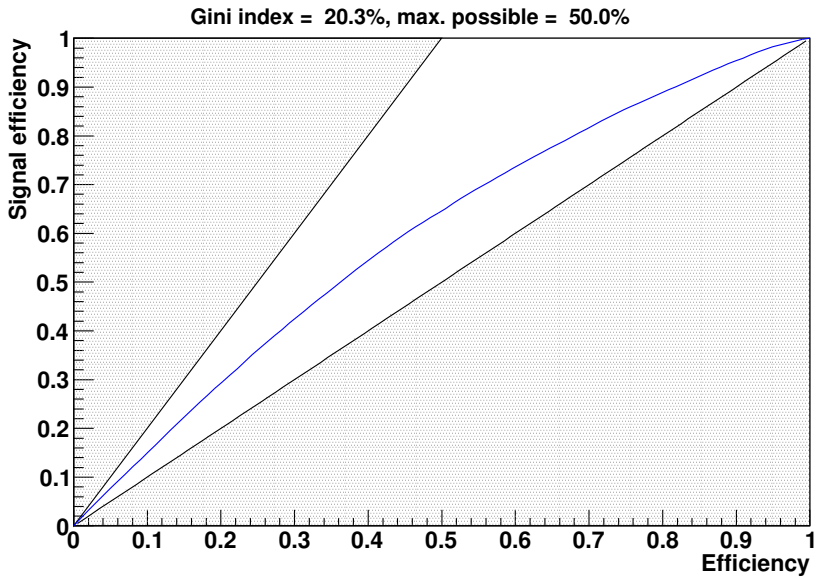


Figure 9.7: Dependence of the signal efficiency on the full efficiency for the training of the signal against all backgrounds. The non-shaded area denotes the allowed efficiency range. The Gini index is given as defined in the text, the maximum allowed value results from the same number of signal and background events in the training.

separation. The theoretically achievable maximum in a training where signal and background events are normalized to the same number is 50 %, which is indicated by the lower edge of the shaded area above the efficiency curve. The presented training of the signal events against all backgrounds leads to a Gini index of 20.3 % as also shown in the plot.

After finalizing the training, the neural network is fixed and the output of the network can be calculated for any given event. Figure 9.8 shows the distributions of the neural network output for collision and simulation data. A fair description of the collision data by the simulation is given with deviations mostly in the background-enhanced region below -0.5. The simulated data shows a steeper rise than the collision data, which results from the distribution of W-boson and light-flavour production which is the dominant contribution in this region. To better judge the signal extraction potential, Figure 9.9 shows the neural network distribution only in the most signal-like region from 0.5 to 0.75. The first three bins in this plot show a downward-fluctuation, while the simulation data in the other bins describe the collision data very well. At this point a possible signal extraction using a fitting method already seems problematic given the size of the statistical error of the data and assuming that systematic variations will also play a role. It is also obvious from this distribution that the top-quark pair production presents the largest contamination of the signal region with background events.

To further understand the distribution of background events from the different production channels, Figure 9.10 shows the neural network output distributions for the signal and the two main backgrounds normalized to unit area. The shapes of the distributions clearly show that the separation of the signal against the W+jets background and the smaller backgrounds is good. The top-quark pair production remains the most problematic background channel. The shape suggests a peak towards the background region but also shows a distinct shoulder with a small peak in the signal region. Given that the top-quark pair production has a much larger cross-section compared to the expected signal cross-section, it is very important to further reduce the contamination of the signal region with $t\bar{t}$ events.

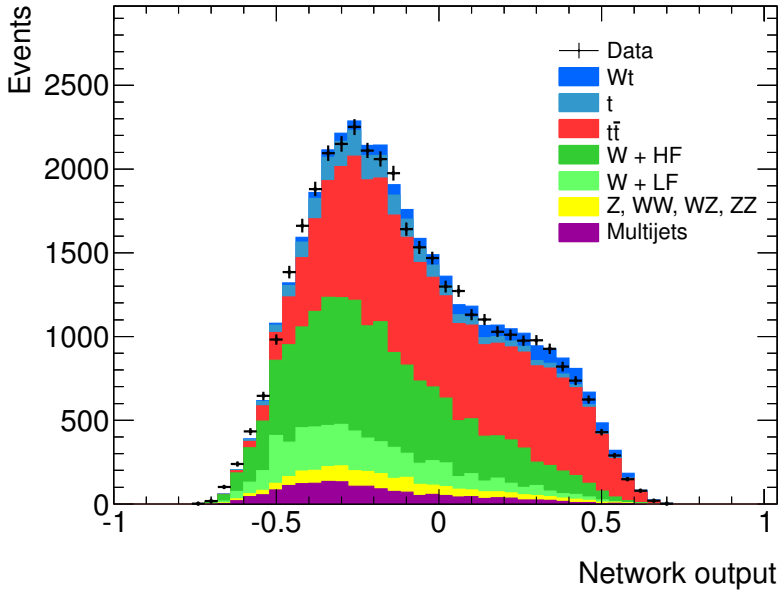


Figure 9.8: Neural network output distribution for the training of the signal against all backgrounds. All simulation contributions are weighted to the data luminosity according to their theoretically predicted cross-sections.

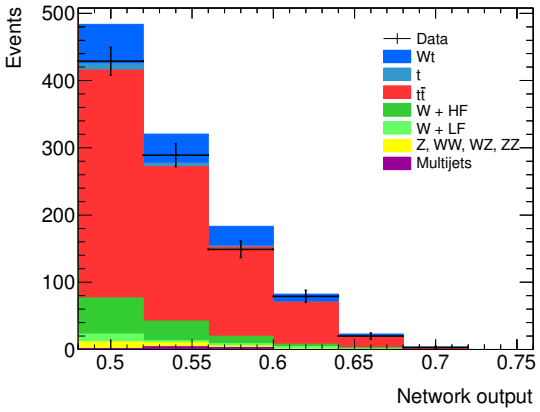


Figure 9.9: Neural network output distribution for the training of the signal against all backgrounds for the signal-like region only. All simulation contributions are weighted to the data luminosity according to their theoretically predicted cross-sections.

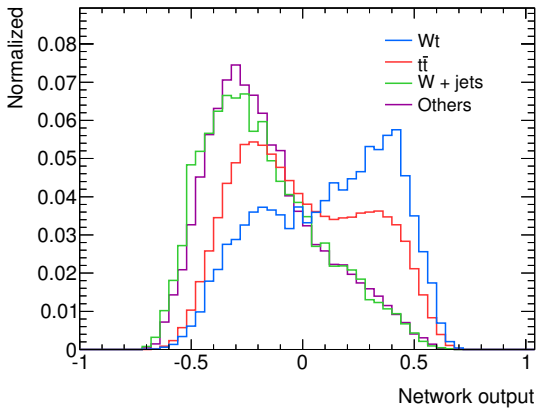


Figure 9.10: Simulated distributions of the neural network output for the training of the signal against all backgrounds normalized to unit area. Only the signal and major background distributions are shown. Other backgrounds are merged into “Others”.

9.3 Combination of two dedicated neural network trainings

A better separation of the events resulting from the single top-quark production in the Wt-channel and the top-quark pair production is the goal of the enhanced analysis method presented here. One variant to accomplish this goal could be to change the event weights for the inputs of the training and overweight $t\bar{t}$ events. While this would motivate the training procedure to construct a neural network with better separation against this background, which would then be considered as the main one, it would reduce the separation against all other backgrounds. To avoid this downside, the presented method creates two discriminant variables using two separate trainings: one variable to separate signal from top-quark pair production and one to discriminate against all other background channels. The two variables are then combined into a two-dimensional distribution which is finally used for the signal extraction in the next chapter.

It was observed in the evaluation of variables with large separation power in Chapter 8 that the shapes of the variable distributions often show either large separation of the signal to $t\bar{t}$ or to the W+jets background. In all times, the respective other background's shape was found to be more similar to the signal in these cases. Separating the training of signal events against $t\bar{t}$ events allows the resulting neural network to be specialised on this purpose and the optimal set of variables for this task to be found. The second network is also expected to produce a better discriminator for all other backgrounds exploiting corresponding variables. This procedure relies even more on the good modelling of the differences between $t\bar{t}$ and single top-quark production. Due to the focus of the training, modelling problems have a larger impact on the final result.

9.3.1 Discrimination of signal against top-quark pair production

The first training uses only simulated events from top-quark pair production as background and Wt associated production as signal. The 21 best variables identified by the preprocessing are used in the network training cutting on a significance of $4.0\,\sigma$ as shown in Table 9.2. The cut value is adjusted to limit the number of input variables, as the significance values generally differ between the trainings. Comparing the variables with the training shown in Section 9.2, only three variables are used in both trainings, namely the mass of the reconstructed hadronically decaying W boson, the pseudorapidity of the charged lepton and the difference between the invariant masses of the harder light-quark jet and the b-quark jet. This gives a hint on the variables that separate well the signal from top-quark pair production in the previous training. The largest number of used variables are based on differences in the azimuthal angle, ϕ, although most of these variables enter the training with low significance. In contrast to the previous training, the invariant masses are used much less, while azimuthal angles are added. It is also notable that the significances are generally lower compared to the ones in the previous training. This is another hint on the problematic separation between signal and events with top-quark pairs.

The neural network training using the described events achieves a Gini index of 17.6 % and its output distribution is shown in Figure 9.11. The modelling of the collision data by simulation is better on the low values of the network in this case, while the region between 0 and 0.3 is not perfectly well described. Figure 9.12 shows the area normalized distributions for different simulated production channels. The distributions of Wt associated production and $t\bar{t}$ are much better separated than in the previous training, which was the goal of the split trainings. The $t\bar{t}$ distribution clearly peaks towards the background-enhanced region and then steeply falls for larger values of the neural network output. The distribution of W+jets shows a worse separation and peaks in the middle of the two former ones. This bad separation doesn't present a large problem as the second network will be trained to suppress mostly this background. It is also visible from this plot that the joint distribution of other backgrounds follows the

Variable	Significance [σ]	Single significance [σ]	Significance loss [σ]	Global correlation [%]
$M(W^{\mathrm{had}})$	33.9	33.9	22.9	55
$p_{\mathrm{T}}(Wt)$	26.7	30.3	22.3	40
$\Delta R(l, W^{\mathrm{lep}})$	11.1	7.4	6.1	48
H_{T}	11.9	17.7	7.5	51
$\Delta\phi(\nu$, b-jet)	5.9	6.1	7.6	47
$M(t^{\mathrm{had}})$	5.7	24.2	5.0	65
$p_{\mathrm{T}}(W^{\mathrm{lep}})$	5.0	6.4	6.9	35
$E_{\mathrm{T}}^{\mathrm{miss}}$	5.5	13.3	6.9	49
ΔR(l-jet2, W^{had})	4.8	9.6	7.3	85
$\Delta\phi$(l-jet1, l-jet2)	4.9	7.8	5.0	84
$\Delta\phi$(b-jet, t^{lep})	4.9	6.0	5.0	61
ΔM(l-jet1, b-jet)	5.0	10.9	5.0	35
$\eta(l)$	4.8	5.4	5.7	34
$\Delta\eta$(b-jet, t^{lep})	4.5	4.1	4.6	38
$\eta(W^{\mathrm{had}})$	4.5	6.6	4.4	45
$\Delta\phi(Wt$, l-jet2)	2.8	3.4	5.2	43
$\Delta\phi(Wt$, l-jet1)	2.2	4.9	5.6	61
$\Delta\phi(Wt, W^{\mathrm{lep}})$	2.3	1.0	4.6	51
$\Delta\phi(Wt, \nu)$	3.1	3.7	4.8	55
$\Delta\phi(Wt, l)$	3.0	5.1	4.6	58
$\Delta\phi(Wt$, b-jet)	4.3	6.5	4.3	57

Table 9.2: Variables selected by the NeuroBayes preprocessing in the training of the signal against top-quark pair production events only. The table is sorted by the variables' significances. Significance, single significance, significance loss and the global correlation of each variable are given as defined in the text in Section 9.1.

one of the W+jets events with a shift to smaller neural network output values avoiding the need for a dedicated handling of these minor background contributions.

9.3.2 Discrimination of signal against W+jets and other backgrounds

In the second training, the signal is trained against the other main background, W+jets production, as well as against all the minor background contributions. The expectation is to achieve a very good separation, as the inclusive training already showed a large shape difference between W+jets production and the signal in the neural network output distribution and the second main background does not need to be discriminated against. In the training, the 16 best variables from the preprocessing are used which corresponds to a significance cut below $5.0\,\sigma$. The list of used variables and their key properties as calculated by the preprocessing procedure are shown in Table 9.3.

This training shares seven variables with the inclusive one from Section 9.2, one of which is transformed: the mass of the W boson as given by the PDG is subtracted from the reconstructed mass of the hadronically decaying W boson. The other shared variables are pseudorapidity of the lepton, ΔR between the hardest light jet and the reconstructed leptonically decaying top quark, difference of the invariant masses of the b-quark jet and the reconstructed hadronically decaying top quark, difference of the invariant masses of the reconstructed Wt system and the reconstructed leptonically decaying top quark,

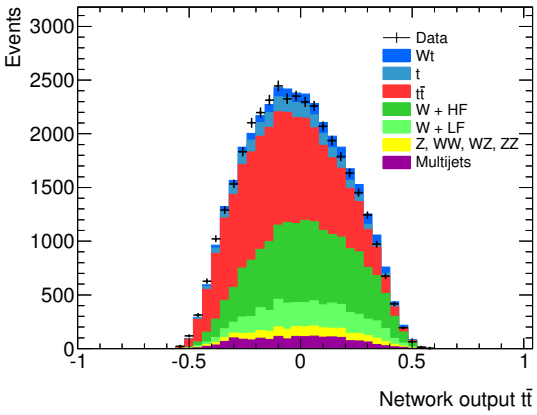


Figure 9.11: Neural network output distribution for the training of the signal against $t\bar{t}$ events. All simulation contributions are weighted to the data luminosity according to their theoretically predicted cross-sections.

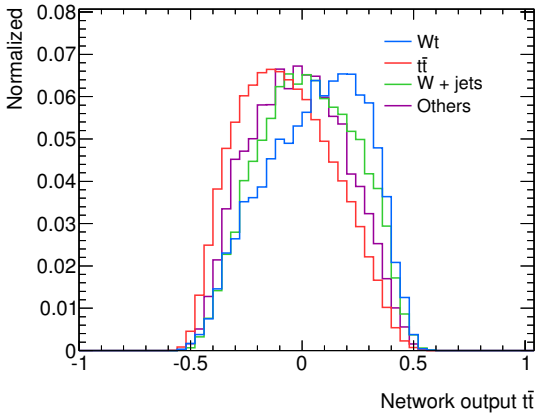


Figure 9.12: Simulated distributions of the neural network output for the training of the signal against $t\bar{t}$ events normalized to unit area. Only the signal and major background distributions are shown. Other backgrounds are merged into "Others".

ΔR between the second hardest light-quark jet and the b-quark jet and ΔR between the hardest light-quark jet and the b-quark jet. Five variables are also common to both trainings in the two-dimensional approach, namely the invariant mass of the reconstructed hadronically decaying W boson, considering the different transformation using the PDG mass as stated above, the transverse momentum of the reconstructed Wt system, the missing transverse momentum, the difference of the invariant masses of the harder light-quark jet and the b-quark jet and the pseudorapidity of the lepton. This shows that there are some kinematic variables which generally show good separation between the different production channels. This was already shown in Chapter 8. The employment of separate trainings solves the problem of the signal peaking in-between the two main backgrounds discussed in that chapter.

The evaluation of the variables shows that this training mostly uses kinematic properties like the invariant mass (six times), the transverse momentum (three times) and ΔR (three times) of reconstructed objects or variables based on these, while the only used event-wide variable is the missing transverse momentum. As in the other two trainings, a variable based on the invariant mass of the reconstructed hadronically decaying W boson is the most significant one.

The neural network training achieves the best separation of all networks shown so far with a Gini index of 31 %. This good result is expected as the backgrounds used in the training show larger kinematic differences from the signal in general as already pointed out in Chapter 8 and the inclusive training in Section 9.2. The distribution of the neural network output of this training is shown in Figure 9.13 for all simulation events, weighted with the appropriate cross-section, and collision events. It shows a good description of the data by the simulation and a clear enrichment of W+jets production towards lower values of the network output. Figure 9.14 shows the normalized distributions of the major simulation channels and again demonstrates the good separation of W+jets events. A comparable shape difference is also achieved for the minor backgrounds which entered the training while the shape of the distribution of $t\bar{t}$ events closely follows the signal one. This has already been motivated above considering the difficult discrimination of these events by the previous training.

Variable	Significance [σ]	Single significance [σ]	Significance loss [σ]	Global correlation [%]
$M(W^{\text{had}}) - M(W^{\text{PDG}})$	45.2	45.2	31.3	47
$\eta(l)$	22.4	24.9	12.4	41
p_{T}(b-jet)	21.5	24.9	13.3	39
ΔR(l-jet1, b-jet)	20.1	28.4	8.3	51
$E_{\text{T}}^{\text{miss}}$	16.8	21.2	13.6	32
$p_{\text{T}}(Wt)$	14.0	14.7	11.7	30
$M(t^{\text{lep}}) - M(t^{\text{PDG}})$	12.7	21.6	11.6	30
p_{T}(l-jet2)	12.4	13.3	8.2	39
ΔM(b-jet, t^{had})	10.4	33.1	10.7	69
ΔR(l-jet2, b-jet)	8.1	18.9	7.9	24
ΔR(l-jet1, t^{lep})	6.9	24.1	9.4	48
$\Delta M(Wt, t^{\text{lep}})$	6.7	18.5	9.0	70
p_{T}(l-jet1)	7.3	16.6	8.2	47
ΔM(l-jet1, b-jet)	7.7	15.1	7.3	44
$M_{\text{T}}(W^{\text{lep}})$	7.4	12.8	7.5	29
$\Delta\eta(l$, b-jet)	7.2	24.4	7.2	42

Table 9.3: Variables selected by the NeuroBayes preprocessing in the training of the signal against W+jets production and other backgrounds. The table is sorted by the variables' significances. Significance, single significance, significance loss and the global correlation of each variable are given as defined in the text in Section 9.1.

9.3.3 Neural network combination

After creating a single discriminant for each of the two main backgrounds of Wt production, this section describes their combination into a single variable which in turn can be used for signal extraction. In a first step, the neural network outputs are plotted against each other in a two-dimensional histogram. The distributions of the simulated events separated by channels are shown in Figure 9.15.

In the two-dimensional representation, the signal- and background-enhanced regions can be seen easily. The simulated distributions of W+jets production and the minor backgrounds peak towards the left part of the respective plots and can be separated from the signal in both directions as described in the trainings before. The signal events are strongly concentrated in the top right corner while $t\bar{t}$ events are shifted down a bit, which demonstrates the separation due to the respective network. It is also visible that $t\bar{t}$ events spread over a large region of the second neural network training, which further improves the separation from signal events. The distribution shows an interesting second peak structure around network output of -0.3 for the $t\bar{t}$ training and zero for the W+jets training, which is most pronounced in this correlation representation. A closer look at the shapes for the single trainings reveals a small enhancement in the training against W+jets, which gets enhanced as the training against $t\bar{t}$ peaks in the questionable region. The dominating peak in the two-dimensional representation is still fixed by the maximum of the neural network output distribution for the W+jets network.

The two remaining steps before the signal extraction are of a more technical nature: a rebinning of the two-dimensional histograms and the transformation into one-dimensional distributions. In the distributions of Figure 9.15 many bins remain empty (white regions) as the two trainings only have a limited separation. Empty bins can be problematic for the template fit presented in the next chapter, if they contain collision data events but no simulation events or only events of one production channel.

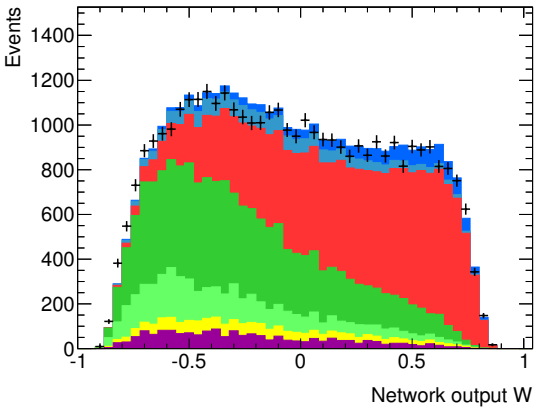


Figure 9.13: Neural network output distribution for the training of the signal against W+jets events and other backgrounds. All simulation contributions are weighted to the data luminosity according to their theoretically predicted cross-sections. The channels are coloured as in Figure 9.11.

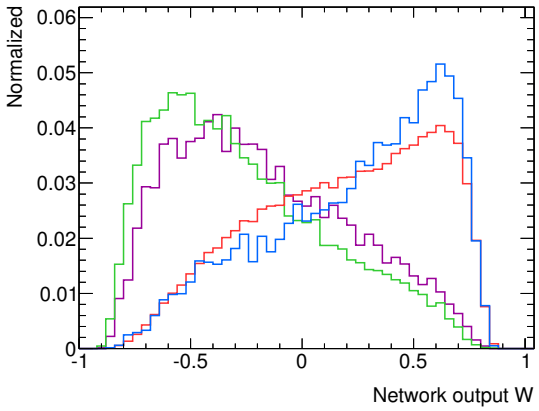
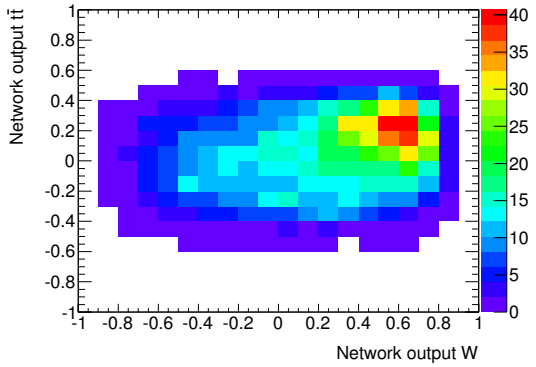
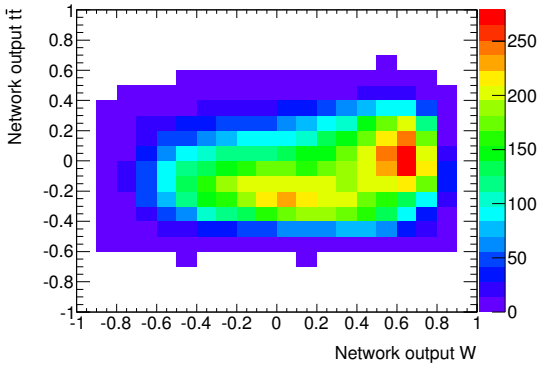


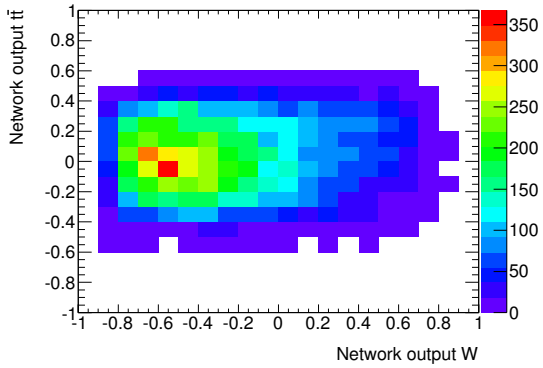
Figure 9.14: Simulated distributions of the neural network output for the training of the signal against W+jets events and other backgrounds normalized to unit area. Only the signal and major background distributions are shown. Other backgrounds are merged into "Others".



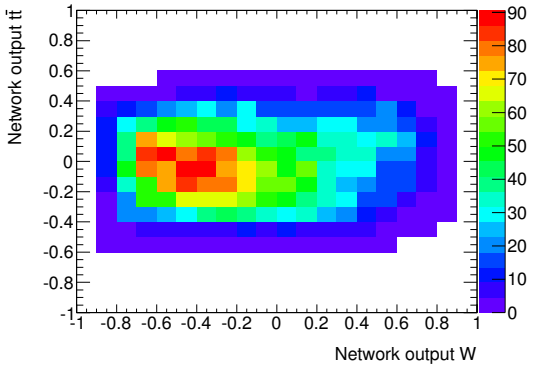
(a) Wt-channel production



(b) $t\bar{t}$ production



(c) W+jets production

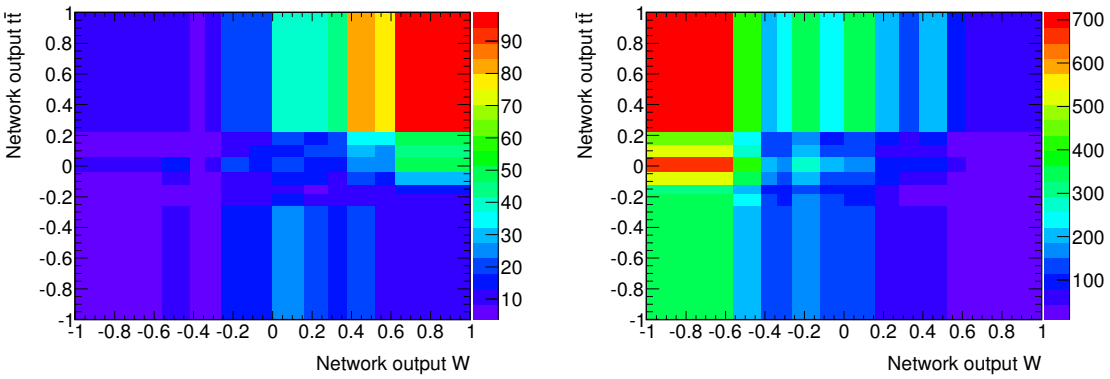


(d) Sum of minor background production

Figure 9.15: Simulated distributions of the neural network output for the training of signal against $t\bar{t}$ events plotted against the neural network output for the training of signal against W+jets and minor background events.

Therefore an iterative rebinning procedure has been applied which is described in the following. The base requirements for the development of the procedure were a minimal number of entries per bin for simulation and collision data distributions, which can be set before the optimisation and not to create too many bins with a large number of entries, which could reduce the separation power of the later fit. The optimization procedure uses a finely-binned reference histogram. In this analysis the sum of all simulated production channels is used as reference where both axes are split into 100 bins each with a bin-width of 0.02. This ensures that all bins containing data also contain a sufficient number of simulated events used by the fit.

In the beginning of the algorithm, empty rows are merged with their respective neighbours starting from the low end of the histogram. If the last row is empty, it is merged with the previous one. The same algorithm is subsequently also applied to all columns. In the next step, the bin with the lowest number of entries is identified, starting from the bin with the lowest value for both networks. The bin with the smallest entry number is merged with the neighbouring row or column that holds the smallest integral. This procedure is repeated until all bins have a minimum of 150 entries. The choice of 150 bins ensures that even templates with small number of bin entries in the extreme regions do not show too large statistical fluctuations. Other minimum values have been tested and no large changes in the results were seen. The resulting binning consists of twelve bins for the network output of the training against W+jets and minor backgrounds and eight bins for the network output of the training against $t\bar{t}$ events. Two example distributions with the new binning are shown in Figure 9.16 for the signal channel and the W+jets production.



(a) Wt-channel production (b) W+jets production

Figure 9.16: Examples of the rebinned simulated distributions of the neural network output for the training of the signal against W+jets and minor background events plotted against the neural network output for the training of signal against $t\bar{t}$ events.

The tool which performs the binned likelihood fit presented in the next chapter uses one-dimensional histograms as inputs. The two-dimensional distributions are thus transformed into a one-dimensional distribution using a procedure that we call unrolling. A new histogram with the same number of bins as the two-dimensional one is created and the bins contents and their errors are then written sequentially into the new histogram. The procedure is started in the most background-like bin and the values are copied column-wise to the new histogram always starting with the lowest bin in the network trained against $t\bar{t}$ events. The resulting combined neural network output is shown in Figure 9.17 for collision data and the simulation.

In this plot is now hard to judge how well signal and backgrounds are separated from each other as

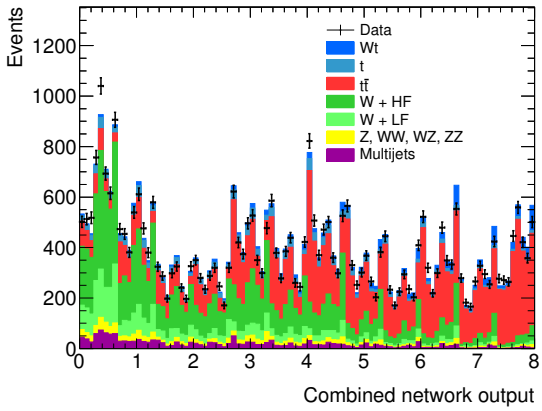


Figure 9.17: Combined network output distribution obtained by applying the unrolling procedure described in the text. All simulation contributions are weighted to the data luminosity according to their theoretically predicted cross-sections.

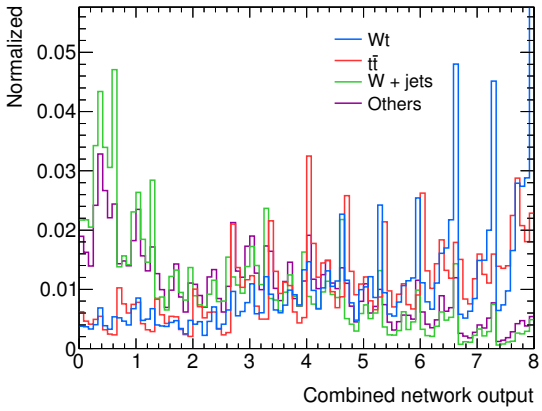


Figure 9.18: Simulated distributions of the combined network output obtained by applying the unrolling procedure described in the text normalized to unit area. Only the signal and major background distributions are shown. Other backgrounds are merged into "Others".

the right-most bins are not necessarily the most significant bins any longer. It is still possible to see that the contamination with events from W+jets diminishes from the left to the right while $t\bar{t}$ production is enhanced. The agreement between simulation prediction and collision data shows disagreements in the lower region of the distribution, which results from the mismodelling of the lower region of the network trained against W+jets seen in Figure 9.13. Another observation is that the sum of predicted events in some bins with large signal content is larger than the number of collision events. Figure 9.18 shows the normalized distributions of the signal and the main backgrounds which allows a better judgement of the shapes. The superposition of the two-networks can be clearly seen here as groups of eight bins are visible which correspond to one column in the two-dimensional representation which shows the distribution for the network trained against $t\bar{t}$ events. In each of these groups the number of W+jets events increases from the first to the last bin seen in Figure 9.16(b). The same behaviour can also be observed for signal events. In the overall distribution, the number of W+jets events is reduced from left to right while the number of signal events increases. This is the effect of the network trained against the W+jets and other backgrounds. A separation in the shapes of Wt and $t\bar{t}$ events can be seen in each of these groups. An unrolling procedure by rows of the two-dimensional histogram would show the exact opposite behaviour. As a binned likelihood fit is applied in the end, both approaches are equivalent and just the bin order is different.

CHAPTER 10

Signal extraction

The previous chapter explained the construction of a combined variable with high separation between the associated production of a W boson and a top quark on the one hand and all the background channel production on the other. This chapter describes how the variable is used for the extraction of the signal cross-section. The procedure includes systematic uncertainties which influence the measurement of the signal cross-section. A description of each source of systematic uncertainty which has been studied and how it is evaluated is given in the second part of this chapter.

10.1 Binned log likelihood fit

A binned maximum likelihood fit of all simulated events to collision data is used to extract the signal and background contributions. It uses the same fitting procedure as the analysis in [141] and the description of the technique follows the one in [146]. The method also has been applied by the CDF collaboration before [147]. The fit is performed using the entire distribution of the combined neural network output. This has several advantages over just cutting the distribution in one or the other way to enrich the signal. First of all, the usage of the full spectrum can exploit the full number of signal events in all bins of the distribution, while using sub-spectra would require an optimization procedure for the cut placement. The result of the signal extraction is also independent of small migrations between the bins in the signal-enriched region of the spectrum. A discrete cut value could have larger effects, especially as the combined neural network output distribution shows larger jumps at various points of the spectrum as seen in Figure 9.17. Placing a cut close to these jumps could introduce a larger bias.

Another important advantage is the constraint of backgrounds in the fit. The full combined neural network output contains different signal-depleted regions which have enhanced contributions of the major backgrounds. As explained in the previous chapter, the two main backgrounds dominate in different regions of the distribution and can therefore be constrained independently. Using separate templates for the production of W bosons and light-flavour jets as well as the production of W bosons and heavy-flavour jets, the fit also fixes the contribution from the different flavours which is not well modelled by the Monte Carlo simulation.

The employed likelihood function is defined as

$$L(\beta_j^b;\beta^s) = \prod_{k=1}^{M} P(n_k|\mu_k) \times \prod_{j=1}^{B} G(\beta_j^b|\Delta_j),$$

where the fit parameters β^s and β^b_j are the scale factors of signal and each background sample j and M is the number of bins of the distribution. The function is composed of two larger building blocks. The first part is the product of Poisson likelihoods of each histogram bin. Each Poisson factor of the product describes the probability to observe n_k events in the bin k given the expected number of events μ_k for that same bin:

$$P(n_k;\mu_k) = \frac{\mu_k^{n_k} \mathrm{e}^{-\mu_k}}{n_k!}.$$

The expected number of events for a bin is calculated by

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

are the expected number of signal (s) and background (b) events. $\tilde{\nu}$ depicts the predicted number of events given by the simulation. α_k is the fraction of events of the full template in the bin k. It follows the normalization $\sum_k \alpha_k = 1$ for each signal or background template. All the factors α_k together thus correspond to the shape of a template distribution.

The second block of the likelihood function is a product of Gaussian functions which include the prior knowledge about the production rate of each background. It is defined as

$$G(\beta_j^b;\Delta_j) = \frac{1}{\sqrt{2\pi\Delta_j^2}} \exp\left\{-\frac{(\beta_j^b - 1)^2}{2\Delta_j^2}\right\}.$$

The Gaussian for each background features a mean of one and a width Δ_j representing the relative rate uncertainty of the respective background. It limits the variation of the scale factor for the corresponding background, β_j^b. This limitation supports a stable fit in the background-enhanced region where several background contributions overlap. The uncertainties utilized in this analysis are provided in Table 10.1.

Process	Δ_j(3 jets)
Single top s channel	0.04
Single top t channel	0.04
$t\bar{t}$	0.10
W+light flavour,	0.42
W+heavy flavour	0.42
Z+jets, WW, WZ, ZZ	0.42
Multijets	0.50

Table 10.1: Relative uncertainties used to constrain the fit of the scale factors of each background contribution. Uncertainties as introduced in Section 5.4. The larger uncertainty of Z+jets production is applied to the combination with diboson production.

In the maximization of the likelihood, very large numbers can occur, which would render the method numerically unstable. Instead, the fit minimizes the negative logarithm of the likelihood function using the MINUIT[148] minimization tool. The evaluation of the likelihood fit corresponds to a measurement of the cross-section as the acquired signal template scale factor, β_s connects the provided theory cross-section with the estimated one

$$\hat{\sigma} = \beta_s \sigma_{\text{theo}}.$$

The statistical uncertainty on the cross-section is estimated using pseudo-experiments created by a Monte Carlo method. As introduced above, each background process j has an expected number of events $\tilde{\nu}_j$ which is given by the simulated events and the normalization to the integrated luminosity of the data sample using the theoretically calculated cross-sections. In each pseudo-experiment a random number n_j is thrown according to a Poisson distribution around a mean of $\tilde{\nu}_j$. Afterwards, n_j values are randomly chosen from background template j, which corresponds to the combined neural network output in this analysis. The resulting distributions are fitted using the original templates obtaining scale factors β for the signal and each background process. The root mean square, RMS, of the resulting distribution for β_s is then an estimate for the expected uncertainty on the signal cross-section measurement. The ensemble of all pseudo-experiments represents our statistically limited a-priori knowledge. This method has the advantage that systematic uncertainties can be included into the pseudo-experiments as described in Section 10.2.

10.2 Systematic uncertainties

The outcome of a measurement is always influenced by systematic effects which lead to an increased uncertainty, the systematic uncertainty. It is a combination of uncertainties from very different sources, e.g. limited precision of input factors such as the energy of measured particles, the efficiency in the reconstruction, and the modelling of processes. In this section the treatment of systematic uncertainties in the analysis is described. Systematic uncertainties are evaluated according to the prescriptions centrally provided by the ATLAS top group for all analyses involving top quarks. Some prescriptions are common to all ATLAS analyses using the same dataset and are given by the same working groups which also studied the reconstruction of particles and missing transverse energy as described in Chapter 4. The first part of this section explains how the systematic uncertainties are included into the binned maximum likelihood fit followed by a brief description of the different systematic uncertainties which are studied for this analysis.

10.2.1 Systematic uncertainty estimation

Systematic uncertainties can belong to two classes: rate and shape uncertainties. Rate uncertainties change the number of accepted signal and selected background events. Shape uncertainties describe migrations between the bins inside the combined neural network output distributions. Most sources of systematic uncertainties contribute to both classes. In some cases, they only affect single production channels. In the following it is described how the uncertainty resulting from the theoretical predictions of the background cross-sections is calculated and how the two classes of systematic uncertainties are included into the maximum likelihood fit. An additional uncertainty stemming from the limited statistics of the simulation is also introduced.

Cross-section uncertainty estimation

To include the influence of the uncertainty on the cross-sections of the corresponding templates, random numbers, β^{gen}, are created for each template according to a log-normal density function

$$\ln \mathcal{N}(x;\mu,\sigma) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\ln x - \mu)^2}{2\sigma^2}\right\},$$

where μ corresponds to the mean of the distribution of $\ln x$ and σ to its standard deviation. Choosing this type of function avoids the appearance of negative expectation values, which would be unphysical. In

the determination of the statistical uncertainty, a Gaussian distribution with a mean of one and a width Δ_j has been used. Setting the same expectation value, E and variance, V, here, we obtain the following equations for the two factors:

$$E = e^{\mu+\frac{\sigma^2}{2}} \stackrel{!}{=} 1 \quad \text{and} \quad V = (\mathrm{e}^{\sigma^2} - 1)e^{2\mu+\sigma^2} = (\mathrm{e}^{\sigma^2} - 1)E^2 \stackrel{!}{=} \Delta_j^2.$$

This leads to a mean of the distribution of $\mu = -\sigma^2/2$ and $\sigma^2 = \ln(\Delta_j^2+1)$ as fixed parameters. Compared to the Gaussian distribution used for the randomization of the number of events of a template before, the log-normal distribution has the advantage that it is not defined for negative values preventing unphysical scale factors β^{gen}. The scale factor of the signal process does not need to be varied in the estimation as it is used to measure the signal cross-section and the uncertainty is estimated at the cross-sections predicted. It is only included during hypothesis testing.

Rate uncertainty estimation

The rate uncertainties allow a larger fluctuation of the β values for a template. This means, that the template can have more or fewer entries, but synchronously for all bins. The relative sizes of the up-fluctuation, ϵ_{ij+}, and down-fluctuation, ϵ_{ij-}, for each systematic uncertainty i are obtained from running the analysis with up and down variations of the corresponding systematic effect, respectively. They correspond to the one-sigma deviation compared to the nominal analysis. The fluctuations have to be included into the expectation values, $\tilde{\nu}_j$, for each process which is in turn used in the generation of the pseudo-experiments as explained in the previous section. To do so, a nuisance parameter δ_i is introduced which specifies sign and magnitude of a systematic deviation. This nuisance parameter is randomly chosen from a Gaussian distribution of width one and mean zero for each systematic during the creation of each pseudo-experiment. The new expected number of selected events (including the cross-section uncertainty explained above) is calculated by

$$\nu_j = \tilde{\nu}_j \, \beta_j^{\text{gen}} \left\{ 1 + \sum_i |\delta_i| \left(\mathcal{H}(\delta_i) \, \epsilon_{ij+} + \mathcal{H}(-\delta_i) \, \epsilon_{ij-} \right) \right\},$$

where $\mathcal{H}$ denotes the Heaviside step function which decides if an up-fluctuation or a down-fluctuation is applied in the pseudo-experiment. The factor ν_j is again used to create pseudo-experiments as explained previously and the RMS of the resulting distribution of β_s is a measure for the uncertainty due to the systematic effects.

Shape uncertainty estimation

Systematic effects can also lead to bin migrations in the template of the combined neural network output. By using the modified events of the systematically changed model and evaluating the corresponding combined neural network output, new factors α^+_{jki} and α^-_{jki} are extracted for upward and downward variations of the respective effect for each systematic i. During the creation of pseudo-experiments the new template is created by a linear interpolation between the shape of the default template and the systematically altered one using the strength of the variation defined by the nuisance parameter δ_i. The factor is the same one as is used in the incorporation of the rate variation above. The new shape is obtained exploiting the Heaviside step function again:

$$\alpha_{jk} = \tilde{\alpha}_{jk} + \sum_i |\delta_i| \left\{ (\alpha^+_{jki} - \tilde{\alpha}_{jk}) \mathcal{H}(\delta_i) + (\alpha^-_{jki} - \tilde{\alpha}_{jk}) \mathcal{H}(-\delta_i) \right\}.$$

Simulation statistics uncertainty estimation

A limited number of simulation events available for the creation of the fit templates imposes another source of systematic uncertainty. Assuming Gaussian errors, the statistical uncertainty of the number of events, n_k, in a bin k is given by $\sqrt{n}$. To evaluate the systematic uncertainty, the number of entries in a bin is set to a random number distributed according to a Gaussian distribution with a mean n_k and a width of $\sqrt{n_k}$ for each pseudo-experiment. Technically, the bin error of the histogram is used as the width of the Gaussian as the template histograms are scaled to the collision data luminosity.

10.2.2 Considered systematic uncertainty sources

Many possible sources of systematic uncertainties have been studied in the analysis apart from the limited statistics of simulation data samples and cross-section uncertainty which were already mentioned above. In the following, a short explanation of each studied source is given with an emphasis on how the templates or rate uncertainties are calculated. All systematic uncertainties, except for the one due to the luminosity determination, are evaluated applying the event selection to one or more event samples which have the variation applied. Afterwards, the combined neural network output distribution is created for the variations and templates are produced as described below to account for the systematic variation of the shapes. A large number of the effects are related to the reconstruction of particles in the ATLAS detector, which was detailed in Chapter 4.

Lepton energy scale The electron energies in collision data are corrected with a scale factor by default as explained in Section 4.2. Varied scales as given in [94] are applied to simulation events to evaluate systematic deviations. The variation of the scale factors depends on the pseudorapidity, η, of the electron's calorimeter shower and the transverse energy, E_{T}. The correction is binned in six η-regions for the central electrons used in the analysis. Templates for both upward and downward variations are created and compared to the default. Muon momenta in collision data are also scaled by default according to the p_T and η of the muon. To calculate the systematic uncertainty, the selection is applied to events without the momentum scale offset. The combined neural network templates are created for the variation and are then symmetrized around the default distribution. This also means that the rate differences are symmetrized around the default event numbers.

Lepton energy resolution Lepton energies in Monte Carlo events are always smeared to take the finite resolution into account. The systematic uncertainty is evaluated by increasing and decreasing the width of the Gaussian distribution used as a reference for the randomization of the factor used for the energy variation. Templates for both variations are created and compared to the default.

Lepton identification and trigger efficiency Inefficiencies observed in data are reproduced by applying scale factors to the simulation events which are extracted from measurements of $Z \to \ell\ell$. [94, 97]. Decays of the type $W \to e\nu_e$ were also included in the analysis of the electron scales. The scale factors are varied according to their systematic uncertainties depending on their momentum and pseudorapidity. The uncertainties were found to be around 1 % for electrons and muons. Templates for both variations are created and compared to the default.

Jet energy scale The uncertainties on the measured jet energies are given based on the transverse momentum of a jet, as shown in Figure 4.3, and its pseudorapidity. They vary between 2.5 % and 8 % in the central region and between 3.5 % and 14 % in the forward region of the detector [99]. Terms

accounting for the flavour composition of the sample of simulated events and the presence of close-by jets are included. The latter reach their maximum at 5 % in the central region and 7 % in the detector's forward region. An additional uncertainty of 2.5 % for true b-jets in Monte Carlo is added to account for different results compared to light and gluon jets obtained in the studies. The jet energies are varied up and down by $1\,\sigma$ of the combined uncertainty and the changes in jet energy from the variation are propagated into the missing transverse energy calculation. Templates for all simulation samples and both variations are created and compared to the default.

Jet energy resolution The finite resolution on the jet energy is taken into account by smearing the jet energies in Monte Carlo events to match the resolution in data based on the transverse momentum and the rapidity of the jets. To evaluate the systematic uncertainty, an additional smearing of the Monte Carlo energies is applied. The new jet energies are propagated into the missing transverse energy calculation. Using the varied events, new templates are generated and symmetrized around the default distribution to obtain upward and downward variations.

Jet reconstruction efficiency The reconstruction is only inefficient for jets with $p_T < 30\,\text{GeV}$ and found to be 2.7 % at most. To account for uncertainties in the reconstruction efficiency, jets are randomly dropped and removed from the event according to a probability to be badly reconstructed. The sample is then used to create new templates which are finally symmetrized around the distributions with no dropped jets.

Jet vertex fraction As explained in Section 4.4.2, the events are reweighted with a combined jet vertex fraction scale factor of all jets which are calculated based on four efficiency scale factors. For the evaluation of the systematic effect of the scale factor application, the four individual scale factors are varied based on their $1\,\sigma$ uncertainties. A combined systematic variation factor is applied on the default scale, resulting in upward and downward variation templates.

b-quark, c-quark and light flavour scale factors All jets in Monte Carlo events are recalibrated based on their simulated flavour. Currently b-quark, c-quark and light flavour jets are distinguished. Calibration scale factors are supplied by the ATLAS flavour tagging group together with uncertainties on the scale factors for upward and downward variations. Varied scale factors are then applied separately for b-quark, c-quark and heavy flavour efficiencies as well as mistag efficiencies to Monte Carlo samples and the varied results are evaluated to create fitting templates.

Luminosity measurement A fixed uncertainty of 1.8 % is applied to the acceptances of all Monte Carlo channels and evaluated symmetrically.

Monte Carlo generator, parton shower and hadronisation model There are systematic effects in the generation of events which are based on the choice of Monte Carlo and parton shower generator combination. To study the effect on the analysis, the production of $t\bar{t}$ events and signal events has been studied using different combinations. The default combination for both channels is the MC@NLO generator interfaced to the HERWIG generator, which is used for the generation of parton showers and the hadronisation. For the production of top-quark events, samples for the combination POWHEG and PYTHIA are used to create templates which are then symmetrized around the default. Only the single top t-channel production is not varied, as only the AcerMC generator is available to provide a correct modelling.

Initial and final state radiation Higher order corrections in perturbation theory lead to additional gluon radiation in the initial and final states of the production reaction. Three parameters in the Pythia generator are used to vary the strength of the radiation, two for the initial state radiation (ISR) and one for the final state radiation (FSR). In space-like showers `PARP(67)` defines the maximum parton virtuality altering the Q^2-scale in the hard process and thus sets an upper limit to the momentum transfers in initial state radiation. The parameter defaults to 1.0 and is varied to 1.4 to increase ISR and 0.6 to decrease it. `PARP64` changes the squared transverse momentum evolution scale, which allows to scale the strong coupling constant α_s. The default is again 1.0 and the reduction to 0.9 increases the probability for ISR while the increase to 3.5 lowers the probability. In time-like showers, `PARP72` sets the QCD scale Λ_{QCD} value used in α_s. The default is set to 0.26 GeV. A value of 0.37 GeV is used to enhance FSR while a value of 0.11 GeV is used to decrease the probability for FSR.

Two Monte Carlo samples are available for each production channel, one with increased initial and final state radiation and one with decreased radiation. The samples were created using the AcerMC generator for all top-quark production channels. To study the impact on the analysis, the combined neural network output distribution is created for both samples. The two distributions are then subtracted and the differential distribution is rescaled by a factor of 0.5. The final differential distribution is added and subtracted from the distribution obtained from the default generator to create the two final templates which enter the systematics evaluation in the fit.

Diagram subtraction and diagram removal As introduced in Chapter 2, the signal can be simulated using two different strategies: diagram removal (DR) and diagram subtraction (DS), where DR is the used default. A Monte Carlo sample simulated using DS is used and the combined neural network output distribution is symmetrized around the according DR distribution to create templates for the fitting procedure.

W+jets shape variation When simulating W+jets events with the Alpgen generator, several parameters have to be set which were not yet fully confirmed with data. These parameters can alter the shapes of kinematic distributions of the generated event samples. To account for such changes, several parameters have been changed and the effect on the distribution shapes has been studied. Using a reweighting procedure based on the transverse momentum of the hardest jet these shape variations can be propagated to the analysis based on the default samples without generating new events. Two generator parameters are changed separately using this reweighting procedure in this analysis to evaluate the systematic uncertainty: the functional form of the factorisation scale, IQOPT, and the minimal parton p_T generated by ALPGEN, PTJMIN. The default function for IQOPT is $Q^2 = m_W^2 + \sum p_{\mathrm{T}}^2(\mathrm{jet})$ and is changed to $Q^2 = m_W^2 + p_{\mathrm{T}}^2(W)$ for a first sample. The templates for the new shapes are symmetrized around the central value and the systematic effect is evaluated in the fit. A second sample is created reducing IQOPT from 15 GeV to 10 GeV and the templates are created in the same way.

10.2.3 Inputs to the systematic uncertainty estimation

Using the methods explained in Section 10.2.1 the impact of the systematic rate and shape deviation on the cross-section measurement is estimated. As most sources of systematics affect both classes of uncertainties, the rate deviations and the shape templates of the combined neural network output distribution have to be evaluated for each production channel as inputs for the creation of pseudo-experiments. An overview of the rate deviations for each studied systematic effect is given in Table 10.2.

The largest deviations of the event rate are visible in the systematics due to the jet vertex fraction, initial and final state radiation, the b-tagging calibration scale factors of real b quarks and the jet energy

Source of uncertainty		Wt	$t\bar{t}$	t	W+HF	W+LF	Z, WW, WZ, ZZ
e energy resolution	up	0.0	0.0	0.1	−0.0	0.0	−0.7
	down	0.0	0.0	0.1	0.1	0.1	−0.5
e energy scale	up	0.3	0.2	0.1	0.0	0.7	0.1
	down	0.1	−0.1	−0.2	−0.3	−0.4	−0.4
e reconstruction	up	1.1	1.1	1.0	1.1	1.0	0.9
	down	−0.9	−1.0	−1.0	−0.9	−0.8	−1.5
μ momentum resolution ID	up	0.1	0.0	0.1	−0.1	−0.2	−0.2
	down	0.0	0.0	−0.0	−0.1	0.2	−0.5
μ momentum scale	up	0.1	0.1	0.1	0.2	0.2	−0.1
	down	−0.1	−0.1	−0.1	−0.2	−0.2	0.1
μ reconstruction	up	1.0	1.0	0.9	1.0	1.1	0.5
	down	−0.8	−0.8	−0.9	−0.8	−0.8	−1.1
Jet reconstruction	up	0.3	0.0	0.1	0.1	−0.2	−0.3
	down	−0.3	−0.0	−0.1	−0.1	0.2	0.3
Jet energy resolution	up	−1.3	−0.6	1.4	1.5	3.1	7.5
	down	1.3	0.6	−1.4	−1.5	−3.1	−7.5
Jet energy scale	up	2.0	−2.7	5.8	10.8	12.7	11.8
	down	−2.4	2.1	−5.7	−9.7	−11.1	−10.9
Jet vertex fraction	up	9.6	11.1	8.4	8.6	9.3	9.9
	down	−9.0	−10.0	−8.6	−8.5	−8.6	−9.9
b-tagging (b quark)	up	3.0	−0.9	0.8	0.9	0.4	0.7
	down	−2.8	0.5	−1.1	−0.8	−0.1	−1.3
b-tagging (c quark)	up	−0.5	−0.6	−0.2	10.1	0.8	6.0
	down	0.8	0.7	0.2	−10.0	−0.6	−6.7
b-tagging (light quark)	up	0.1	−0.1	−0.1	1.4	21.2	5.4
	down	0.2	0.2	0.2	−1.2	−21.1	−6.0
Wt DS / DR	up	1.8					
	down	−1.8					
POWHEG + PYTHIA	up	−0.6	−2.1	0.0			
	down	0.6	2.1	−0.0			
Initial/final state radiation	up	−9.4	−3.4	5.4			
	down	9.4	3.4	−5.4			
W+jets shape (Iqopt3)	up	0.1	0.1	0.0	0.9	0.9	−0.3
	down	−0.1	−0.1	−0.0	−0.9	−0.9	0.3
W+jets shape (Ptjmin10)	up	0.1	0.1	0.0	0.0	−0.0	−0.3
	down	−0.1	−0.1	−0.0	−0.0	0.0	0.3
Luminosity	up	1.8	1.8	1.8	1.8	1.8	1.8
	down	−1.8	−1.8	−1.8	−1.8	−1.8	−1.8

Table 10.2: Relative rate uncertainties for analysed sources of systematic effects. All numbers are given in percent.

scale. This behaviour is expected as the event selection already shows a large sensitivity of the analysis on the number of jets in the event and their properties.

Examples for the influence of a systematic effect on the shape of the combined neural network distributions are given in Figures 10.1, 10.2 and 10.3, which show the default distribution together with the upward and downward variations due to the jet energy scale systematic for each of the six templates which are prone to this systematic. The distribution for multijet production is not shown, as it is not varied and only estimated once using the default events. As one can see, the largest effect is on the production of top quarks: top-quark pair production, single top-quark production and signal channel. The other channels show larger statistical errors due to the limited number of simulated events. This shows the requirement for larger statistics of the simulation samples as most of the systematic variations are covered by the statistical errors and are more fluctuations, as seen in the deviation plot of Figure 10.1 for the signal channel.

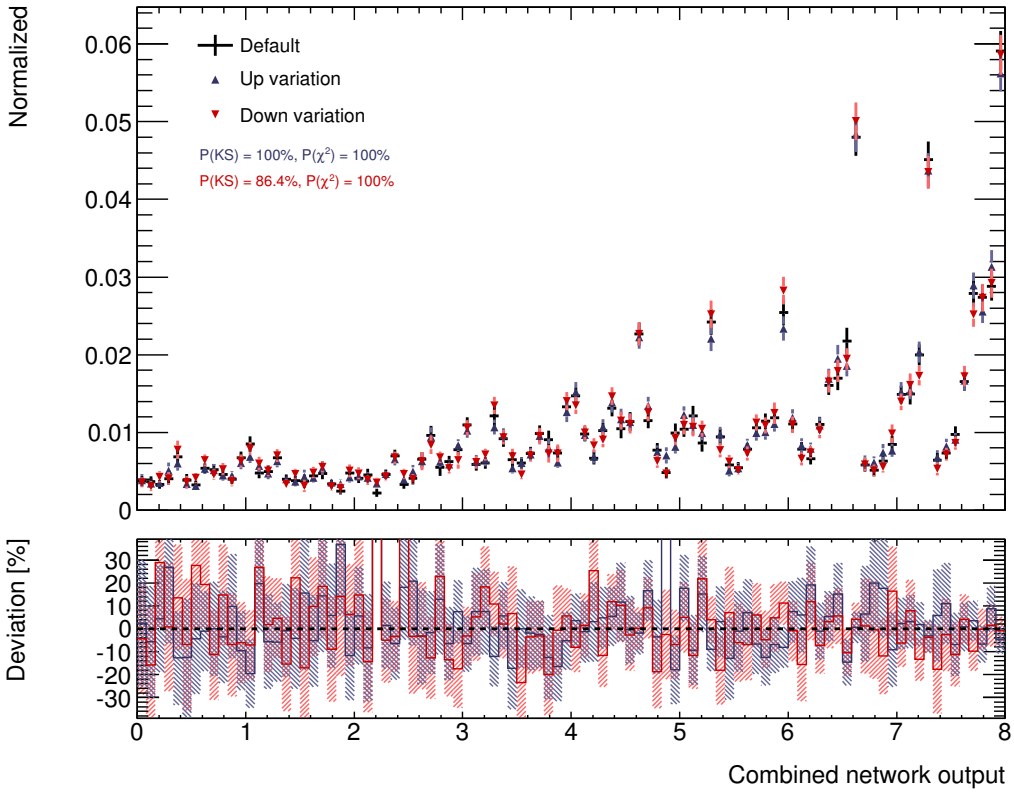


Figure 10.1: Distributions of the combined neural network output for Wt production normalized to unit area used as template in the fit. The plot shows the default distribution compared to the upward and downward variations of the jet energy scale systematic. The shown error bars correspond to the statistical uncertainty of the available Monte Carlo sample only. Probabilities resulting from a Kolmogorov-Smirnov test and a χ^2 test for the agreement between variation and default are given for both variations. The lower plot shows the deviation of the systematics including the combined statistical errors as hashed areas.

As mentioned at the end of the previous chapter, a group of eight bins in the distribution combines to the neural network training of the signal against $t\bar{t}$ events. In Figure 10.2 this shape can be identified again: eight bins show a tendency to less entries, while the ninth bin shows a pronounced jump upwards. This can be best spotted in the middle of the distributions. The jet energy scale variation for $t\bar{t}$ events is particularly interesting because of a shift in the correlation of the network response to the direction

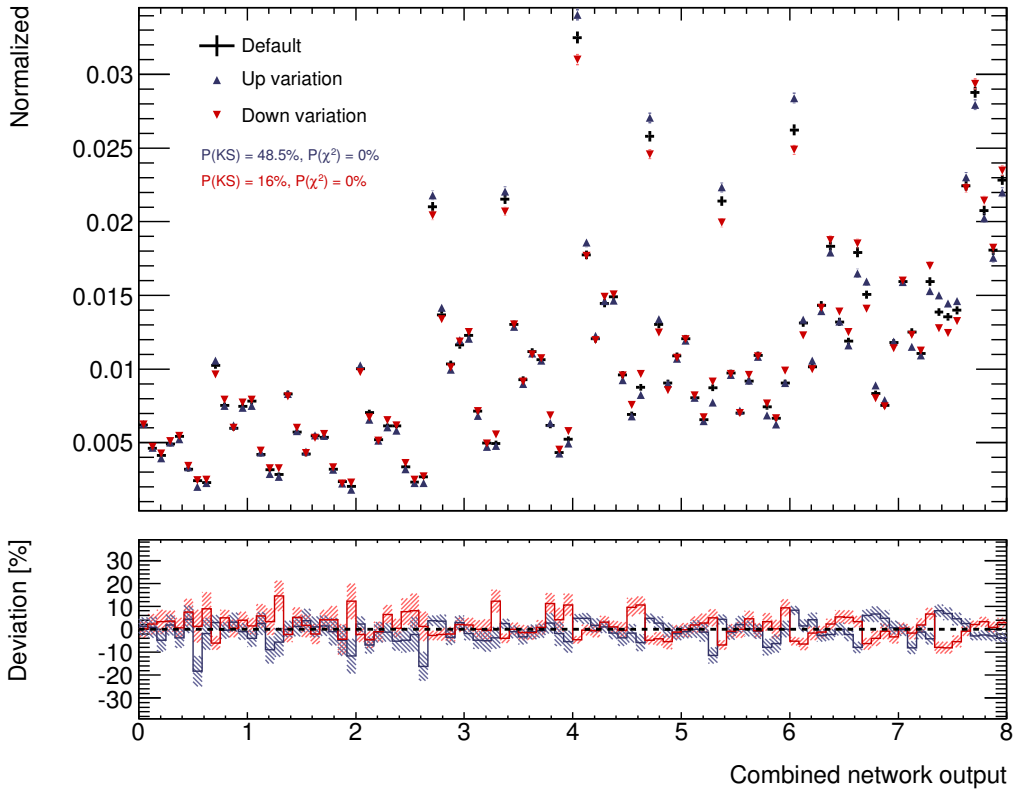
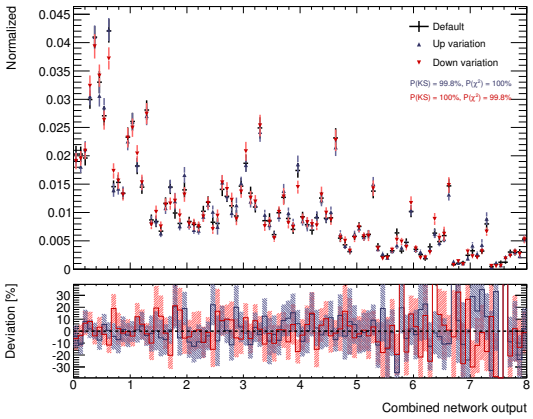
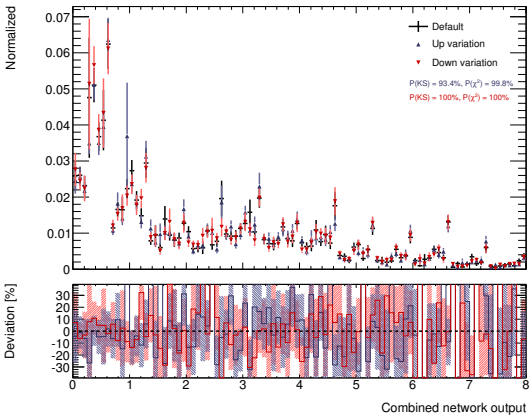


Figure 10.2: Distributions of the combined neural network output for $t\bar{t}$ production normalized to unit area used as template in the fit. The plot shows the default distribution compared to the upward and downward variations of the jet energy scale systematic. The shown error bars correspond to the statistical uncertainty of the available Monte Carlo sample only. Probabilities resulting from a Kolmogorov-Smirnov test and a χ^2 test for the agreement between variation and default are given for both variations. The lower plot shows the deviation of the systematics including the combined statistical errors as hashed areas.

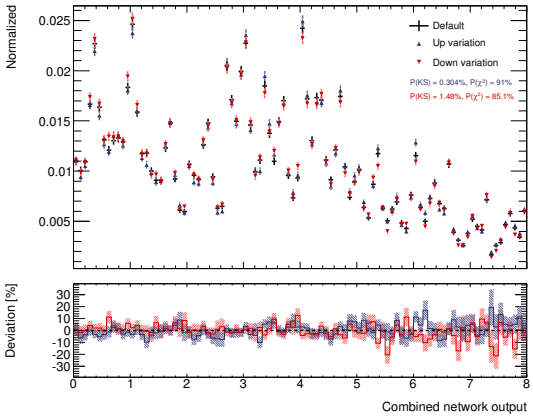
of the variation of the jet energy scale. In the first group of nine bins, the shape uncertainty vanishes in the first bins while it increases to the end of the bin group. In all eight bins of the group, the downward variation lies above the default distribution which is an anti-correlation and is also visible in the up-scaling of the jet energies. Interestingly, looking at the least W+jets-like events at the high end of the combined network output distribution, a crossing between correlation and anti-correlation can be observed. Wt-like events show the anti-correlation, while background-like events show a correlation between the direction of the jet energy scale variation and the change of the number of events: less jet energy also means fewer events in that region. This is a very good example how the neural network training can differentiate between events based on their properties. Exploiting the correlations on top of the pure combination of kinematic variables, signal-like and background-like events also show different responses to systematic variations. As the top-quark pair production is such a prominent and difficult to separate background, this observed effect plays a major role on the uncertainty of the cross-section measurement.



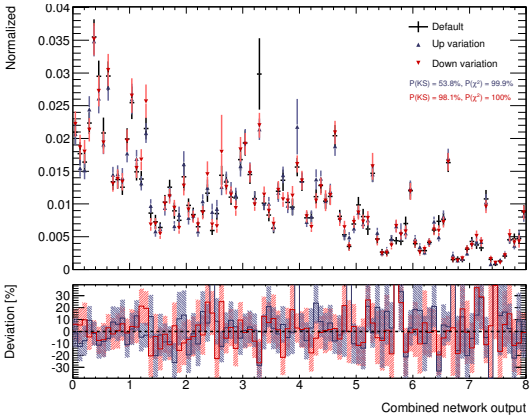
(a) W+heavy flavour jets production



(b) W+light flavour jets production



(c) t-channel and s-channel production



(d) Z+jets, WW, WZ, ZZ

Figure 10.3: Distributions of the combined neural network output for different backgrounds normalized to unit area used as template in the fit. Each plot shows the default distribution compared to the upward and downward variations of the jet energy scale systematic. The shown error bars correspond to the statistical uncertainty of the available Monte Carlo sample only. Probabilities resulting from a Kolmogorov-Smirnov test and a χ^2 test for the agreement between variation and default are given for both variations. The lower plot shows the deviation of the systematics including the combined statistical errors as hashed areas.

10.3 Significance estimation

To estimate the significance of the measurement two hypotheses are tested. The null hypothesis, H_0, is the background-only hypothesis. It assumes that there is no signal contribution and that the data can be described by the background contributions alone. The second hypothesis, H_1, assumes the presence of a signal contribution with the rate predicted by the theoretical calculations based on the Standard Model. For both hypotheses, pseudo-experiments are created accordingly: the pseudo-data for H_0 does not include signal events, while the data for H_1 does. To test the hypotheses, each pseudo-experiment of a hypothesis is fitted, including all systematic uncertainties and also the theoretical uncertainty on the cross-section of the signal, and a log-likelihood ratio is computed:

$$Q = -2 \ln \frac{L(\beta^s = 1; \hat{\beta}_j^b)}{L(\beta^s = 0; \hat{\hat{\beta}}_j^b)},$$

where $\hat{\beta}_j^b$ and $\hat{\hat{\beta}}_j^b$ are the fitted scale factors for each background j. The result of the Q-value computations are two distributions q_0 and q_1 for the null hypothesis and the signal hypothesis respectively. The overlap between these two distributions is a measure for the separation of the two hypotheses: the smaller it is, the better it is possible to distinguish between the two. The significance of the measurement is given by the p-value, that is the probability of a measurement which is at least as signal-like as the one observed given the null hypothesis is true. According to this definition the significance is larger the smaller the p-value gets. The p-value is related to the log-likelihood ratio Q by

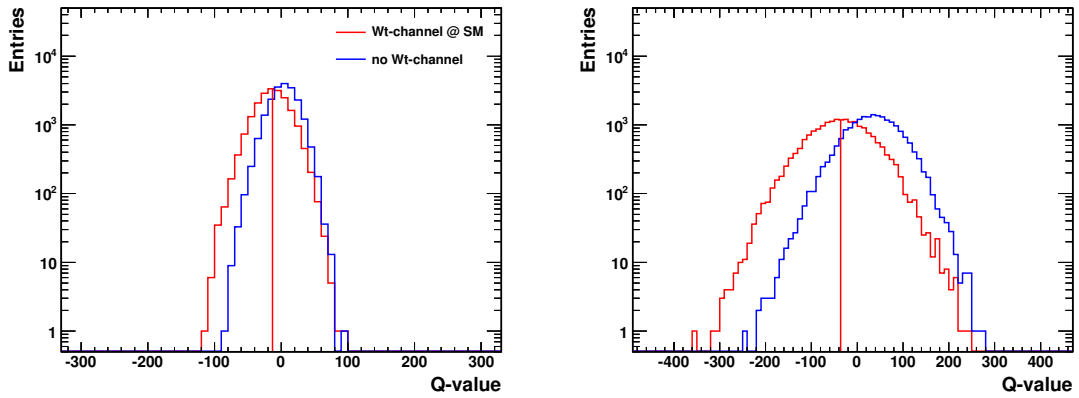
$$p(Q_0) = \frac{1}{I_q} \int_{-\infty}^{Q_0} q_0(Q')\,\mathrm{d}Q', \text{ where } \quad I_q = \int_{-\infty}^{\infty} q_0(Q')\,\mathrm{d}Q'.$$

Geometrically, the p-value is the fraction of measurements of q_0 which lie in the range $-\infty$ to Q_0. The range limit Q_0 is defined separately for the expected and the measured significance. In the first case, the median of the q_1 distribution is used, in the latter case, Q_0 corresponds to the Q-value determined by fitting the data distribution. The usual convention in high-energy particle physics is to express the significance in terms of the Gaussian standard deviation σ. The p-value can be converted into these, using one-sided integrals on the tails of a normal distribution. The $3\,\sigma$ evidence of for a signal then corresponds to a p-value smaller than 1.35×10^{-3} and the $5\,\sigma$ observation to a p-value less than 2.87×10^{-7}.

10.4 Expected significances and uncertainties

In a first step, the expected significances of the two signal isolation methods are evaluated. This allows to compare the performance of the single training and the combined training methods on a quantitative basis including all systematic effects explained in the previous section. Evaluating only the expected significance avoids the introduction of a bias, as the observed significance can fluctuate within the expected uncertainties. The Q-value distributions used in the significance calculation are shown in Figure 10.4 for both methods. They yield to expected significances of $0.7\,\sigma$ for the single neural network training and $1.1\,\sigma$ for the combination of the two separate neural network classifiers.

To investigate the source of the low significance, it is helpful to look at the breakdown of the expected uncertainties on the cross-section measurement. As all sources of uncertainties are included into the likelihood definition as nuisance parameters, a simple investigation of the effects of single sources is not possible. Instead, fits based on pseudo-data including only single sources of uncertainties are performed



(a) Single neural network

(b) Combined neural networks

Figure 10.4: Distributions of Q-values determined using 20 000 pseudo-experiments evaluating the non-signal hypothesis in blue and the signal hypothesis in red for the single neural network training (a) and the combination of two neural network trainings (b). The vertical red line indicates the median of the distribution for the signal hypothesis.

in this analysis and the relative uncertainty on the cross-section measurement is evaluated. While this treatment can hint at the dominant sources of systematic uncertainties, the total systematic uncertainty is not calculated simply from the sum of the squares but extracted from the full likelihood. The sum of squares is only used to combined statistic and systematic uncertainties, as they are uncorrelated. The results of the breakdown for both approaches is shown in Table 10.3 for all included sources of uncertainties described above. The larger significance of the combination method is reflected by a smaller overall uncertainty on the cross-section measurement and the size of the uncertainties is responsible for the generally low expected significances. The breakdown also shows that the total uncertainty is dominated by systematic effects, which can be seen by comparing the combined total with the total from systematics only: it only differs by a few percent at most.

The dominant source of systematic uncertainty in both cases is the variation of initial and final state radiation, followed by the jet energy scale the jet energy resolution and the b-tagging calibration scale factors of real b quarks. Most of these sources were already identified to have a large rate uncertainty in Section 10.2.3. Single important sources of systematic uncertainties are the jet vertex fraction, the background normalization and the b-tagging calibration scale factors of real c quarks for the single neural network approach and the generator systematic comparing to POWHEG + PYTHIA for the combination of two classifiers. The fact that the uncertainties from these sources only grow large in one of the two approaches leads to the assumption that the choice of input variables of the neural networks are responsible. As explained in Section 9.1.1, the neural network classifier is a function of all utilised input variables, which are prone to systematic variations themselves. Hence, a variable showing large systematic variations can enter the neural network classifier with a considerable large weight, which can result in an unstable classifier in terms of the same source of systematic. This assumption is exploited to reduce the impact of systematic effects in the following.

10.5 Optimisation of the expected significance

The prescriptions for the variation of systematic effects for the analyses are centrally worked out within the ATLAS collaboration and shared by many analyses using a similar selection and the same collision

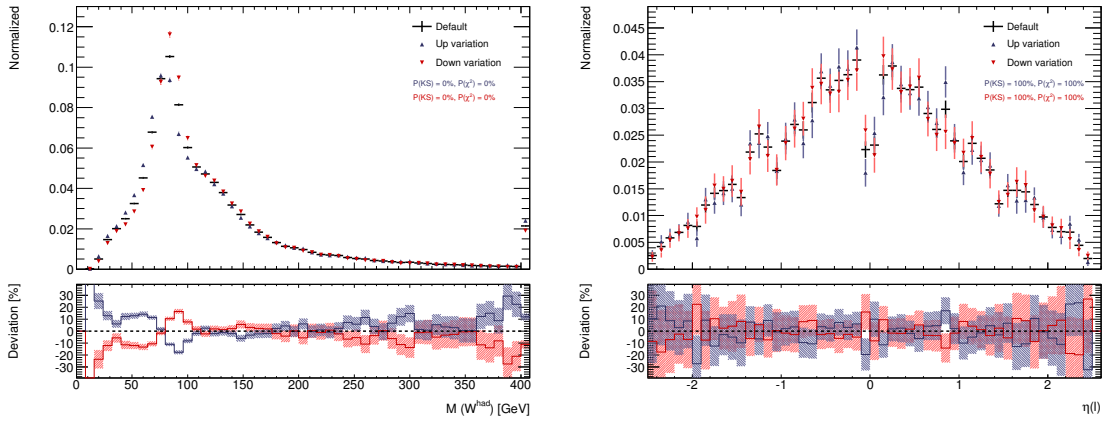
Source of uncertainty	$\Delta\sigma/\sigma$ (single) [%]	$\Delta\sigma/\sigma$ (combined) [%]
Data statistics	±28	±17
Monte Carlo statistics	±20	±10
e energy resolution	±2.3	+1.5/−1.3
e energy scale	+7.3/−4.5	+2.0/−2.6
e reconstruction	+3.2/−2.9	+0.7/−1.0
μ momentum resolution ID	+5.4/−2.9	+3.0/−3.2
μ momentum scale	±2.8	±1.6
μ reconstruction	+2.2/−1.7	+1.0/−1.3
Jet energy resolution	±35	±28
Jet energy scale	±43	±39
Jet reconstruction	±8.7	±2.8
Jet vertex fraction	±23	+3.6/−5.0
b-tagging (b quark)	±31	±26
b-tagging (c quark)	±16	+3.4/−3.5
b-tagging (light quark)	+3.6/−3.4	+2.8/−2.9
Wt DS / DR	±9.7	±3.1
POWHEG + PYTHIA	±1.2	±23
Initial/final state radiation	±100	±71
W+jets shape (Iqopt3)	±1.8	±7.5
W+jets shape (Ptjmin10)	±1.6	±4.2
Background normalization	±20	±3.2
Luminosity	±3.1	±2.2
Total (systematic)	±140	±92
Total (systematic + statistic)	±140	±93

Table 10.3: Expected relative uncertainties on the cross-section of the single top-quark production in the Wt-channel of the single and combined neural network approaches.

dataset. As such, it has been tried to reduce the impact of such variations on the analysis using various approaches.

In a first attempt, all variables were checked for large systematic variations. Distributions, which show the default shape of a variable and the two variations corresponding to each systematic effect, like the ones in Figure 10.5, were therefore created. A Kolmogorov-Smirnov test and a χ^2 test are included to quickly identify larger inconsistencies between the default and the varied distributions. Small probabilities are an indication for differences and hence systematic uncertainties. Variables with such deviations could then be removed from the trainings of the neural networks. While this approach lead to a small reduction of the systematic uncertainties, it also turned out that variables, which are important for the discrimination between signal and background events, can show large deviations in the systematic variations. This creates a dilemma in the choice of variables. A second observation was that the reduction of systematic uncertainties was not fundamental, as the judgement only from the distributions is very tedious and error-prone for the huge number of employed variables. It was then decided that only an automatised approach could improve the choice of variables.

The only sensible figure of merit in the analysis, which includes both, the separation between signal



(a) $t\bar{t}$ production (b) Wt-channel production

Figure 10.5: Example distributions compared to the upward and downward variations of the initial-state and final-state radiation systematic. All distributions are normalized to unit area. The shown error bars correspond to the statistical uncertainty of the available Monte Carlo sample only. Probabilities resulting from a Kolmogorov-Smirnov test and a χ^2 test are given for both variations.

and background processes and the effect of systematic variations, is the expected significance of the cross-section measurement. The disadvantage of this choice is the large computation demand. For each change in the choice of variables, the neural network has to be retrained, the neural network classifier has to be calculated for all collision and simulation events and pseudo-experiments have to be created to evaluate the expected significance of the measurement. In the course of the analysis, a program for the unsupervised execution of these steps for a given set of training input variables has been developed, allowing to flexibly test different iterative optimisation algorithms. All three steps depend on specific requirements, which have to be fulfilled and are partially special to the used computing element.

The NeuroBayes Expert library, employed in the training process, requires a host-bound licence, which is only available on a specific host. The program therefore executes remote training sessions on the training host. It also starts with a run of the preprocessor only, allowing the user to refine the choice of used variables by adjusting the cut on the variable significance as explained in Chapter 9. Alternative functions are included to predefine the significance cut-value and to keep a specific number of variables in the training. It is also possible to set the number of internal network nodes, which was not used in this analysis as the network training results prove to be very stable over a wide range of internal node configurations, as detailed before and in [139]. After the preprocessor step, the full training, including the preprocessing, is run on the chosen set of variables and the network configuration is saved for the next step. The program also creates the standard control distributions for the training and tables with the variable ranking in LaTeX format.

The neural network response has to be calculated for every available data event. As the evaluation is independent for each event, this step can be parallelised easily, using a computing cluster, which can save a considerable amount of time compared to the iterative processing. The developed program uses a predefined scheme, iteratively computing the neural network output for the default distributions and the systematic variations in parallel, resulting in 28 jobs. A further splitting would be desirable in the case of a growing dataset, but was not necessary for the 2011 ATLAS data. The program automatically sends the jobs to the computing farm and regularly checks for their completion. The completion of all computations are necessary for the next step, which is the creation of templates for pseudo-data generation and the binned log-likelihood fit. A separate plotting program is called for this step. It also

determines the event yields of the selection and rate deviations caused by the systematic variations, which are required by the fitting tool.

Using the created templates, pseudo-experiments are generated for the significance determination. This process is also executed on the computing farm, as each pseudo-experiment is random and therefore independent from the generation of all others. The program verifies the completion of all jobs again and finally calculates the expected significance, as explained above. The value can then be used for different optimisation schemes.

For this analysis, two different optimisation schemes have been applied. The first scheme was only applied to the combination of two networks and started with a base set of three variables, p_T, η and ϕ of the charged lepton, adding one additional variable in each iteration. After each iteration, the change of the expected significance was evaluated and the variable only kept, if the significance increased. The optimisation started with the training against the $t\bar{t}$ background and modified the second network only after the first network had reached its best variable choice. Overall, the result was not satisfactory. The maximum expected significance was $0.8\,\sigma$ and therefore worse than the default training. The main reason is supposedly the lost exploitation of variable correlations by the neural network when adding single variables only. This is supported by the fact that the best network configuration only uses seven variables in the training against the $t\bar{t}$ background and did not add any variable in the optimisation of the training against all other variables. The procedure also took very long, as 187 variables were checked for each network, summing to over a week of execution time.

The final applied scheme uses the opposite strategy starting from the default trainings as introduced in Chapter 9, where the preprocessing of NeuroBayes chooses the variable combination with the largest correlation to the target, i.e. the best separation of signal and background events. The significance of the this default training is used as a starting point and one variable is removed from the list of input variables at each iteration, testing the effect on the expected significance. This approach has the advantage that the optimisation starts with the highest significance reached so far and only a small number of variables need to be tested, limiting the required number of iterations. Table 10.4 documents the significance changes for each iteration. It lists the variable that is removed from the training and the according increment or decrement of the expected significance of the cross-section measurement. The table shows that the most significant variable in the training, $M(W^{\text{had}})$ is removed, adding $0.11\,\sigma$ to the measurement significance. This is at the same time the largest significance gain by removing a variable and confirms the large effect of the systematic variation, as seen in Figure 10.5(a). Apparently, the significance loss due to the increment of the systematic uncertainty is larger than the separation gain of this variable. An increment of the same order is reached by removing $\Delta R(\text{l-jet2}, \text{b-jet})$ and only minimal gain by removing $\Delta R(\text{l-jet1}, \text{b-jet})$ from the training. These variables are the least important ones in the training, which shows the potential of the presented method.

Adding a single variable can have dramatic effects on the expected significance. This statement is valid for both: incrementation and decrease of significance. The expected significance highly depends on anti-correlations of systematic variations as well. Removing the sixth best variable, $\Delta M(\text{b-jet}, t^{\text{had}})$, from the training, even returns a negative expected significance, which results from the fact that there is no lower limit of zero for the likelihood fit. In this case, the fits of the pseudo-experiments return negative values for the signal contribution. Large impacts on the significance are also observed for many of the other variables, even the ones with lower importance in the neural network training. Using the optimisation, the expected significance can be increased from 0.7 to $0.9\,\sigma$, corresponding to an improvement of 30 %.

The same optimisation procedure has also been applied to the combination of two neural networks. The results of the optimisation of the training against the $t\bar{t}$ background in Table 10.5(a) show several improvements compared to the default network. First of all it is worth mentioning that the most sig-

Variable	Significance change [σ]
$M(W^{\text{had}})$	0.11
$\eta(l)$	−0.52
p_{T}(jet1)	−0.09
$\Delta\eta$(b-jet, W^{lep})	−0.21
ΔR(l-jet1, t^{lep})	−0.23
ΔM(b-jet, t^{had})	−2.04
$\Delta M(Wt, t^{\text{lep}})$	−0.03
ΔM(l-jet1, b-jet)	−0.41
ΔM(l-jet1, t^{lep})	−0.13
$p_{\text{T}}(l)$	−0.42
$\Delta\phi(\nu, W^{\text{had}})$	−0.55
ΔM(l-jet2, b-jet)	−0.55
ΔR(l-jet2, b-jet)	0.09
ΔR(l-jet1, b-jet)	0.00

Table 10.4: List of variables and corresponding change of the expected significance of the cross-section measurement when removing it from the training for the single neural network approach.

nificant variable in the training, $M(W^{\text{had}})$, does not get removed in this case. Obviously, the loss of separation power is more dramatic for this separate network, than for the combined training and is not over-compensated by diminishing the effect of systematic variations. A total of six variables can be removed with a gain of expected significance for this training, which corresponds to more than a quarter of all used variables. The variable with the largest gain is $\Delta\phi(Wt, \text{l-jet2})$, which improves the significance by 0.11 σ. The optimisation of this one network in the combined approach is capable of increasing the significance from 1.1 to 1.5 σ, corresponding to an improvement of 31 %, which is in the same range as the former optimisation of the single network.

Additionally to the optimisation of the exclusive training against the $t\bar{t}$ background, the procedure is also applied to the training against the W+jets and smaller backgrounds. The changes in significance for each variable is summarised in Table 10.5(b). It shows that no improvement is reached at all by employing the introduced method. Instead, the removal of a variable only worsens the expected significance. A probable cause could be the separation of single variables. As shown in the comparison of the shapes of the distributions for the signal and the W+jets background, which has the largest impact in this training, in Chapter 8, the differences are more pronounced than the ones between signal and the $t\bar{t}$ distributions. This impression is supported by a trend for larger single significances of the variables used in the training against the non-$t\bar{t}$ backgrounds, given in Tables 9.2, compared to the ones for the training against $t\bar{t}$ given in Table 9.3. Removing variables with better separation power has a larger effect on the final network performance, which is obviously not compensated by the reduction of systematic uncertainties. This shows that the presented method is most effective in the case of large potential reduction of systematic uncertainties and for distributions that are hard to separate. It also supports the statement that the top-quark pair production is the single problematic background for the single top-quark production in the Wt-channel.

Variable	Significance change [σ]
$M(W^{\text{had}})$	−0.05
$p_{\text{T}}(Wt)$	−0.33
$\Delta R(l, W^{\text{lep}})$	0.02
H_{T}	−0.08
$\Delta\phi(\nu$, b-jet)	−0.01
$M(t^{\text{had}})$	−0.07
$p_{\text{T}}(W^{\text{lep}})$	−0.09
$E_{\text{T}}^{\text{miss}}$	0.01
ΔR(l-jet2, W^{had})	−0.05
$\Delta\phi$(l-jet1, l-jet2)	0.03
$\Delta\phi$(b-jet, t^{lep})	−0.02
ΔM(l-jet1, b-jet)	0.03
$\eta(l)$	−0.02
$\Delta\eta$(b-jet, t^{lep})	0.09
$\eta(W^{\text{had}})$	−0.16
$\Delta\phi(Wt$, l-jet2)	0.11
$\Delta\phi(Wt$, l-jet1)	−0.08
$\Delta\phi(Wt, W^{\text{lep}})$	−0.08
$\Delta\phi(Wt, \nu)$	−0.08
$\Delta\phi(Wt, l)$	−0.08
$\Delta\phi(Wt$, b-jet)	−0.08

(a) Training against $t\bar{t}$ background

Variable	Significance change [σ]
$M(W^{\text{had}}) - M(W^{\text{PDG}})$	−0.03
$\eta(l)$	−0.05
p_{T}(b-jet)	−0.07
ΔR(l-jet1, b-jet)	−0.09
$E_{\text{T}}^{\text{miss}}$	−0.05
$p_{\text{T}}(Wt)$	−0.19
$M(t^{\text{lep}}) - M(t^{\text{PDG}})$	−0.15
p_{T}(l-jet2)	−0.18
ΔM(b-jet, t^{had})	−0.10
ΔR(l-jet2, b-jet)	−0.17
ΔR(l-jet1, t^{lep})	−0.07
$\Delta M(Wt, t^{\text{lep}})$	−0.19
p_{T}(l-jet1)	−0.10
ΔM(l-jet1, b-jet)	−0.17
$M_{\text{T}}(W^{\text{lep}})$	−0.13
$\Delta\eta(l$, b-jet)	−0.08

(b) Training against other backgrounds

Table 10.5: Lists of variables and corresponding change of the expected significance of the cross-section measurement when removing it from the training for the combined neural network approach.

10.6 Cross-sections

The neural network classifiers obtained by the optimisation methods are used for a cross-section measurement. The templates of all simulated production channels are therefore fitted to the collision data distribution and the scale factors β are obtained. The results of the fits for the single network and combined network methods are summarized in Table 10.6. The table lists the event yields for the simulation prediction as explained in Chapter 7, the scale factors of the fit and the rescaled number of predicted events with the according scale factor. The fitted scale factors for the Wt-channel differ largely for the two methods: while the fit extracts a value of $\beta = 0.64 \pm 0.39$ for the single network, the combined method only shows a value $\beta = 0.07 \pm 0.16$. This result shows that the fit has problems fixing a signal cross-section given the bad separation between signal and background of the classifiers. The overall comparison shows a very good agreement between the expected number of events and the observed number of collision events. The number predicted by the combined method is a bit lower than the observed, which could be due to the small signal contribution in the fit. The difference is still well covered by the statistical error. The combined method shows a tendency for smaller errors on the backgrounds while the error on the signal is much larger. These facts are arguments for the assumption that the potential of the combined neural network classifier cannot be fully exploited by the binned log-likelihood fit.

Process	Prediction Events	Single β	Single Events	Combined β	Combined Events
Single top Wt	1670 ± 150	0.64 ± 0.39	1100 ± 700	0.07 ± 0.16	110 ± 270
Single top t	1830 ± 80	1.00 ± 0.04	1830 ± 80	1.07 ± 0.02	1960 ± 40
Single top s	100 ± 4	1.00 ± 0.04	100 ± 4	1.07 ± 0.02	106.7 ± 1.9
$t\bar{t}$	16 000 ± 1500	0.93 ± 0.06	15 000 ± 900	0.98 ± 0.04	15 700 ± 700
$Wc\bar{c}, Wb\bar{b}, Wc$	11 000 ± 5000	1.09 ± 0.08	12 600 ± 900	0.96 ± 0.06	11 100 ± 700
W + jets	3700 ± 1600	1.02 ± 0.17	3800 ± 600	1.10 ± 0.14	4100 ± 500
Z + jets	1100 ± 500	0.99 ± 0.10	1110 ± 110	0.98 ± 0.10	1100 ± 110
WW, WZ, ZZ	315 ± 31	0.99 ± 0.10	313 ± 31	0.98 ± 0.10	310 ± 31
QCD multijets	1700 ± 900	1.00 ± 0.00	1700 ± 900	1.00 ± 0.00	1700 ± 900
Expected	38 000 ± 6000	–	37 500 ± 1800	–	36 200 ± 1400
Observed	37 503	–	37 503	–	37 503

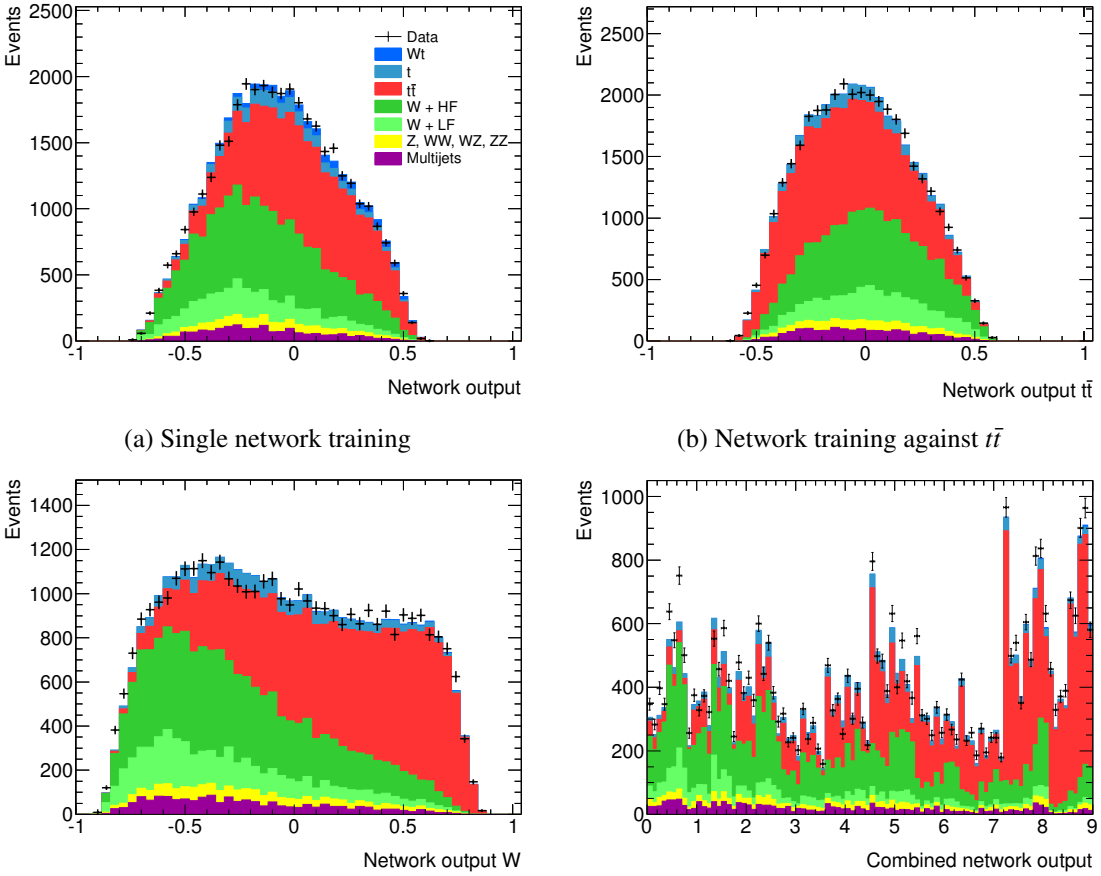
Table 10.6: Number of events with three jets predicted by the simulation and estimated by the fit of the single and combined neural network classifiers. The number of events for QCD multijets are the result of the data-based estimation explained in Section 6.2. The uncertainties on the simulation data are given according to Section 5.4. The errors on the fitted scale factor β and the estimated events are statistical only.

Figure 10.6 shows the optimised neural network distributions after rescaling the simulation templates using the β-factors estimated by the two fits. A good description of the collision data can be observed in the fitted distributions 10.6(a) and 10.6(d). The neural network output for the training against $t\bar{t}$ in Figure 10.6(b) also shows a good agreement while the training against the other backgrounds in Figure 10.6(c) shows the previously observed discrepancies on the edges of the distribution. In this respect, the obtained β-factors appear to be reasonable although the combined neural network classifier still shows potential for an increased signal strength in several bins. The reason for the low signal contribution using this approach remains unclear.

10.6.1 Statistic and systematic uncertainties

It has already been explained before, how the expected statistic and systematic uncertainties can be evaluated. The varied templates for the main systematics for both approaches compared to the default template are shown in Appendix C. Applying the scale factors determined by the fit of collision data to the input event yields of each process, it is possible to extract also the observed uncertainties. The breakdown of the expected and observed uncertainties for both analysis approaches are shown in Table 10.7. The absolute uncertainties are given here, to compare the expected and observed uncertainties, which would otherwise be relative to the obtained cross-sections the central values of which differ a lot. Comparing the expected relative uncertainties to Table 10.3 it is possible to derive the improvements reached by the optimisation method. The dominating uncertainty in the breakdown, which is due to the variation of initial-state and final-state radiation, is now reduced to around 50 % for the single training and 60 % for the combined training. Large reductions for the single network can be seen in the b-tagging systematics for real b quarks. The relative uncertainties drop from 31 to 22 %. At the same time, the uncertainties to do jet energy scale and resolution, the jet vertex fraction as well as due to variation of the W+jets shape rise substantially. For the single neural network methods, the variable selection obviously has a large impact on the systematic uncertainties which become dominant.

A different observation can be made for the combined network approach. The reduction of uncertain-



(a) Single network training

(b) Network training against $t\bar{t}$

(c) Network training against W+jets and others

(d) Combination of networks

Figure 10.6: Distributions of neural network classifiers. The simulation distributions are normalized according to the fitted scale factors.

ties is of general nature and affects all sources of uncertainties. While the total systematic uncertainty of the approach is still very large, the reduction to below 80 % for the combined network approach is promising. Regarding the observed uncertainties, this positive improvement contradicts with the low value of the scale factor determined in the fit of collision data and its large error. Nevertheless, the table shows that the absolute uncertainties agree fairly well between expectation and observation. A tendency to smaller observed uncertainties can be seen. The combined training approach yields generally smaller uncertainties, especially visible for jet energy resolution, jet energy scale and jet vertex fraction and the background normalization. It depends stronger on the choice of generator, though. The comparison between statistic and the other uncertainties listed in the table identifies the systematic uncertainties as the largest problem of the analysis again, despite the small separation between signal and $t\bar{t}$ background in the network classifiers.

10.6.2 Cross-section measurement

Using the theoretically predicted cross-section for the production of single top quarks in the Wt-channel, the fit results can be translated into cross-section measurements. In case of the single neural network

	Single		Combined	
Source of uncertainty	$\Delta\sigma_{\text{exp}}$ [pb]	$\Delta\sigma_{\text{obs}}$ [pb]	$\Delta\sigma_{\text{exp}}$ [pb]	$\Delta\sigma_{\text{obs}}$ [pb]
Data statistics	±5	±5	±2.6	±2.5
Monte Carlo statistics	±4	±4	±1.6	±1.6
e energy resolution	+0.6/−0.5	+0.8/−0.7	+0.11/−0.03	±0.22
e energy scale	+1.1/−0.8	+1.2/−0.9	+0.20/−0.26	+0.13/−0.24
e reconstruction	+0.8/−1.1	+0.9/−1.2	+0.22/−0.21	±0.21
μ momentum resolution ID	±0.5	±0.4	+0.13/−0.14	+0.15/−0.19
μ momentum scale	±1.2	+1.1/−1.0	±0.5	±0.4
μ reconstruction	+0.3/−0.9	+0.8/−1.2	±0.14	±0.27
Jet energy resolution	±8	±6	±5	±3
Jet energy scale	±8	±7	±4	±4
Jet reconstruction	+0.5/−0.4	+0.8/−0.7	±0.12	±0.12
Jet vertex fraction	±6	±6	+1.0/−1.1	+2.2/−2.3
b-tagging (b quark)	±4	+3.0/−3.1	±4	±3.1
b-tagging (c quark)	+2.4/−2.6	+2.3/−2.5	±0.7	±0.4
b-tagging (light quark)	+0.7/−1.1	+0.8/−1.2	+0.33/−0.32	±0.15
Wt DS / DR	±0.34	±0.19	±0.27	±0.16
POWHEG + PYTHIA	±0.23	±1.0	±3.1	±5
Initial/final state radiation	±8	±6	±9	±6
W+jets shape (Iqopt3)	±1.9	±2.1	±0.9	±0.9
W+jets shape (Ptjmin10)	±1.7	±1.8	±0.4	±0.5
Background normalization	±4	±4	±0.31	±0.27
Luminosity	±0.7	±0.8	±0.3	±0.23
Total (systematic)	+18/−19	+16/−17	±12	±10
Total (systematic + statistic)	+19/−20	+17/−18	±13	±11

Table 10.7: Expected and observed absolute systematic uncertainties on the cross-section of the single top-quark production in the Wt-channel for the single and combined neural network approaches.

approach a cross-section of

$$\sigma_{\text{exp}}^{Wt} = 16 \pm 5\,(\text{stat.})\,^{+18}_{-19}\,(\text{syst.})\,\text{pb} = 16^{+19}_{-20}\,\text{pb}$$

is expected and the observed cross-section is

$$\sigma_{\text{obs}}^{Wt} = 10 \pm 5\,(\text{stat.})\,^{+16}_{-17}\,(\text{syst.})\,\text{pb} = 10^{+17}_{-18}\,\text{pb}.$$

These values give no preference for the existence or absence of single top-quark production in the Wt-channel due to the large systematic uncertainties. The value for the observed cross-section has a tendency for its existence.

The approach of the combination of two neural network classifiers yields an expected cross-section of

$$\sigma_{\text{exp}}^{Wt} = 16 \pm 3\,(\text{stat.}) \pm 12\,(\text{syst.})\,\text{pb} = 16 \pm 13\,\text{pb},$$

where the combined error corresponds to an uncertainty of less than 100 %. The low result of the fit of

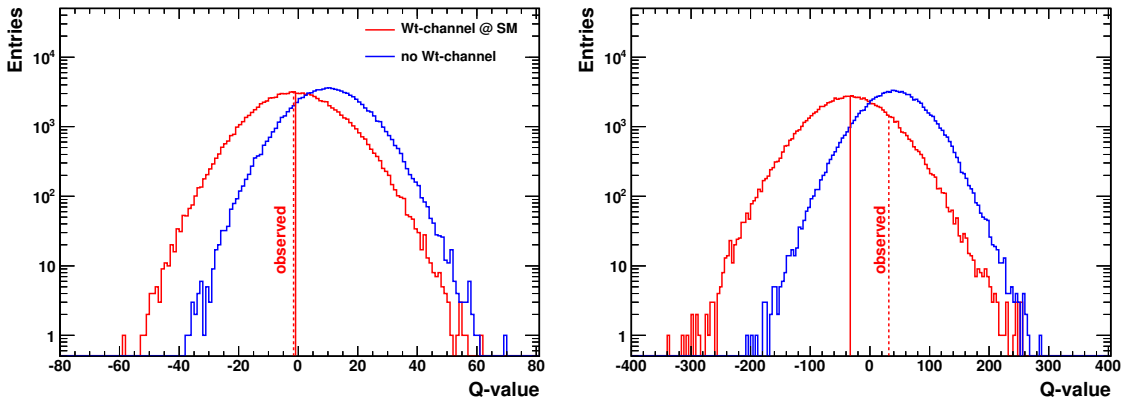
collision data leads to an observed cross-section of

$$\sigma_{\text{obs}}^{Wt} = 1.0 \pm 2.5\,(\text{stat.}) \pm 10\,(\text{syst.})\,\text{pb} = 1 \pm 11\,\text{pb},$$

which is just over one standard deviation away from the Standard Model prediction.

10.6.3 Measurement significance

As introduced in Section 10.3, a log-likelihood ratio is used to test the signal against the fluctuation of backgrounds only. A number of 10^6 pseudo-experiments are used to generate the Q-value distributions of the background-only hypothesis and the hypothesis including signal events for both approaches, which are shown in Figure 10.7. The single network approach, depicted in the left plot, has an expected



(a) Single neural network (b) Combined neural networks

Figure 10.7: Distributions of Q-values determined using 100 000 pseudo-experiments evaluating the non-signal hypothesis in blue and the signal hypothesis in red for the single neural network training (a) and the combination of two neural network trainings (b). The vertical red line indicates the median of the distribution for the signal hypothesis corresponding to the Q-value. The dashed vertical line indicates the position of the observed Q-value.

Q-value of -0.9, which corresponds to a p-value of 0.18 and a significance of $0.9\,\sigma$. The observed Q-value is very similar with -1.55, corresponding to a p-value of 0.15 and yielding a slightly larger significance of $1.0\,\sigma$. The proximity of both significances is a good indication for the stability of the method in the same way it was already used in other analysis of the single top-quark production sector.

The tests of the combined network approach, shown in Figure 10.7(b), show a better expected result with a Q-value of -32 corresponding to a p-value of 0.07 and the highest significance of $1.5\,\sigma$. The notably low fit value could not exploit the potential, showing a large Q-value of 33. The resulting p-value is 0.4 and the observed significance drops to 0.19.

CHAPTER 11

Conclusions

An analysis using $4.7\,\mathrm{fb}^{-1}$ proton collision data at a centre-of-mass energy of $\sqrt{s} = 7\,\mathrm{TeV}$ recorded by the ATLAS experiment at the LHC was presented, which shows a search for the single top-quark production in the Wt-channel. As the top quark exclusively decays into a W boson and a b quark, the channel is composed of two W bosons, each of which can decay to two quarks or to a charged lepton and a corresponding neutrino. Events with a single charged lepton were selected, where the lepton is an electron or a muon – leptonically decaying taus were treated inclusively. The expected full final state consists of the charged lepton, a neutrino, which manifests itself in the form missing transverse momentum in the detector, and three jets, one of which is a b-quark jet and two of which are light-quark jets. Using the constraint of the known W-boson mass to recover the third component of the neutrino, the four-vectors of the full decay chain of the process were reconstructed.

The problems with the separation of the Wt-channel signal and the two major backgrounds, top-quark pair production and W+jets production, were discussed in detail. To reach a maximal separation many variables, based on the properties of the reconstructed objects, were regarded. Additionally, mass and angular differences between the reconstructed particles as well as variables based on all objects in the final state were used. It was concluded that a cut-based approach would not reach the necessary purity to extract a signal. Using events with exactly three jets in the final state, an artificial neural network was trained using the NeuroBayes software. The included preprocessing mechanism was exploited to select the best combination of variables to maximise the separation out of the large number of variables regarded to separate the signal from the background. To further improve the separation, a combination method of two neural network outputs into a single classifier was employed and refined to the analysis. In this method, one neural network was trained against $t\bar{t}$ production events, which are hardest to separate, while the other was trained against all other background event production channels, dominated by the W+jets production. An automatic binning procedure was developed to improve the combined classifier and ease the signal extraction, removing all empty bins and asserting a minimal number of bin entries.

A binned maximum likelihood fit was finally used to fit the simulation templates of all processes to the collision data. Using the full range of the neural network classifiers, the fit can extract the signal cross-section and normalize the backgrounds in-situ. This is very important, as the normalization of W+jets production is not well predicted by theoretical calculations. The expected and observed uncertainties on the cross-section measurement, which based on systematic variations, were estimated using pseudo-experiments. Additionally, the expected and observed significances of the two methods were calculated

based on a logarithmic likelihood ratio using pseudo-experiments.

It was pointed out that the expected significance of the method represents a unique figure of merit for possible improvements, as it takes the separation provided by the neural network training as well as the systematic uncertainties into account. Provided that the neural network classifiers solely depend on the input variables, it was concluded that the systematic uncertainties are governed by these. An iterative optimisation method for the variable selection was thus developed, removing single variables from the neural network and re-evaluating the expected significance of the whole method. The optimisation procedure reached improvements of about 30 % in both methods, where the combination method of two networks reached the maximum expected significance of $1.5\,\sigma$, reducing the expected relative systematic uncertainties below 80 %. Evaluating the results of the combined network method, it could also be concluded that the optimisation is most effective for processes that are hard to separate, which is the case for the Wt-channel and $t\bar{t}$ production, while it shows minor or no improvements if the neural network can separate the signal from the background well. This is of course related to the fact that systematic effects also drastically decrease their impact in the latter case.

A final measurement of the cross-section was carried out for both methods. The single neural network training yields an expected cross-section for the single top-quark production in the Wt-channel of

$$\sigma_{\text{exp}}^{Wt} = 16 \pm 5\,(\text{stat.})\,^{+18}_{-19}\,(\text{syst.})\,\text{pb} = 16^{+19}_{-20}\,\text{pb}$$

with a significance of $0.9\,\sigma$, while the measured cross-section is

$$\sigma_{\text{obs}}^{Wt} = 10 \pm 5\,(\text{stat.})\,^{+16}_{-17}\,(\text{syst.})\,\text{pb} = 10^{+17}_{-18}\,\text{pb}$$

and has a slightly larger significance of $1.0\,\sigma$. Using approach of the combination of two neural network classifiers an expected cross-section of

$$\sigma_{\text{exp}}^{Wt} = 15.7 \pm 2.6\,(\text{stat.}) \pm 12\,(\text{syst.})\,\text{pb} = 16 \pm 13\,\text{pb},$$

can be deduced with a significance of $1.5\,\sigma$, while the observed cross-section is

$$\sigma_{\text{obs}}^{Wt} = 1.0 \pm 2.5\,(\text{stat.}) \pm 10\,(\text{syst.})\,\text{pb} = 1 \pm 11\,\text{pb}$$

and has a significance of $0.19\,\sigma$. The results show a stable but less significant measurement for the single network method and point at larger possible significances of the combined network method. The fits for the latter method yield a small value of the signal production cross-section compared to the Standard Model prediction but the increased significance in general shows a larger potential compared to the simpler one.

The presented analysis has shown the complications of the isolation of the single top-quark production in the Wt-channel in general, especially regarding the final state with a single charged lepton. While evidence for the Wt-channel production could be found by ATLAS using only $2.05\,\text{fb}^{-1}$ of the same dataset [61, 62] in the final state with two charged leptons, the analysed decay into a single lepton has proven to be one of the most challenging environments for a potential signal extraction in proton collisions predicted by the Standard Model. It has been shown in this thesis that the investigation and potential reduction of systematic uncertainties is the key for a cross-section measurement, especially in the field of simulation modelling, where increased Monte Carlo statistics for the background channels and a more precise description of the uncertainty of initial-state and final-state radiations are a prerequisite. The results of this analysis can certainly hint at methods to be used also in future measurement attempts and can be employed generically for cross-section measurements.

APPENDIX A

Used Simulation Samples

Process	$\sigma[pb]$	Generator	Events
Wt (DR, inclusive)	16	MC@NLO + HERWIG	900 000
Wt (DS, inclusive)	16	MC@NLO + HERWIG	1 000 000
Wt (DR, inclusive)	16	POWHEG + PYTHIA	940 000
t-channel ($W \to e\nu_e$)	7.0	AcerMC + PYTHIA	1 000 000
t-channel ($W \to \mu\nu_\mu$)	7.0	AcerMC + PYTHIA	1 000 000
t-channel ($W \to \tau\nu_\tau$)	7.0	AcerMC + PYTHIA	1 000 000
s-channel ($W \to e\nu_e$)	0.50	MC@NLO + HERWIG	300 000
s-channel ($W \to \mu\nu_\mu$)	0.50	MC@NLO + HERWIG	300 000
s-channel ($W \to \tau\nu_\tau$)	0.50	MC@NLO + HERWIG	300 000
s-channel ($W \to l\nu_l$)	1.5	POWHEG + PYTHIA	900 000
$t\bar{t}$ (no full-hadronic)	91	MC@NLO + HERWIG	15 000 000
$t\bar{t}$ (no full-hadronic)	91	POWHEG + PYTHIA	10 000 000

Table A.1: Monte Carlo simulation samples used in the analysis to model processes including top quarks. The cross-sections include branching ratios and correspond to the theoretical predictions with NLO precision. All cross-sections are rounded to two significant digits. The full ATLAS detector simulation is used in the generation process of all samples.

Process	$\sigma[pb]$	Generator	Events
Wt (DR, inclusive)	16	AcerMC + PYTHIA	1 000 000
t-channel ($W \to l\nu_l$)	7.0	AcerMC + PYTHIA	3 000 000
s-channel ($W \to l\nu_l$)	0.50	AcerMC + PYTHIA	600 000
$t\bar{t}$ (no full-hadronic)	91	AcerMC + PYTHIA	30 000 000

Table A.2: Monte Carlo simulation samples used in the analysis to model the effects of initial-state and final-state radiation variations for processes including top quarks. The cross-sections include branching ratios and correspond to the theoretical predictions with NLO precision. All cross-sections are rounded to two significant digits. The ATLFAST2 detector simulation is used in the generation process of all samples.

Process	$\sigma[pb]$	Generator	Events
$W \to e\nu_e$ + 0 parton	8300	ALPGEN + HERWIG	17 000 000
$W \to e\nu_e$ + 1 parton	1600	ALPGEN + HERWIG	10 000 000
$W \to e\nu_e$ + 2 partons	450	ALPGEN + HERWIG	3 800 000
$W \to e\nu_e$ + 3 partons	120	ALPGEN + HERWIG	2 000 000
$W \to e\nu_e$ + 4 partons	31	ALPGEN + HERWIG	1 000 000
$W \to e\nu_e$ + 5 partons	8.4	ALPGEN + HERWIG	620 000
$W \to \mu\nu_\mu$ + 0 parton	8300	ALPGEN + HERWIG	17 000 000
$W \to \mu\nu_\mu$ + 1 parton	1600	ALPGEN + HERWIG	5 000 000
$W \to \mu\nu_\mu$ + 2 partons	450	ALPGEN + HERWIG	3 800 000
$W \to \mu\nu_\mu$ + 3 partons	120	ALPGEN + HERWIG	2 000 000
$W \to \mu\nu_\mu$ + 4 partons	31	ALPGEN + HERWIG	950 000
$W \to \mu\nu_\mu$ + 5 partons	8.4	ALPGEN + HERWIG	620 000
$W \to \tau\nu_\tau$ + 0 parton	8300	ALPGEN + HERWIG	3 400 000
$W \to \tau\nu_\tau$ + 1 parton	1600	ALPGEN + HERWIG	2 500 000
$W \to \tau\nu_\tau$ + 2 partons	450	ALPGEN + HERWIG	3 800 000
$W \to \tau\nu_\tau$ + 3 partons	120	ALPGEN + HERWIG	1 000 000
$W \to \tau\nu_\tau$ + 4 partons	31	ALPGEN + HERWIG	500 000
$W \to \tau\nu_\tau$ + 5 partons	8.4	ALPGEN + HERWIG	65 000
$W \to l\nu_l + c$ + 0 parton	980	ALPGEN + HERWIG	6 500 000
$W \to l\nu_l + c$ + 1 parton	310	ALPGEN + HERWIG	4 100 000
$W \to l\nu_l + c$ + 2 partons	77	ALPGEN + HERWIG	1 000 000
$W \to l\nu_l + c$ + 3 partons	17	ALPGEN + HERWIG	310 000
$W \to l\nu_l + c$ + 4 partons	4.3	ALPGEN + HERWIG	30 000
$W \to l\nu_l + c\bar{c}$ + 0 parton	150	ALPGEN + HERWIG	1 300 000
$W \to l\nu_l + c\bar{c}$ + 1 parton	130	ALPGEN + HERWIG	1 000 000
$W \to l\nu_l + c\bar{c}$ + 2 partons	62	ALPGEN + HERWIG	1 000 000
$W \to l\nu_l + c\bar{c}$ + 3 partons	20	ALPGEN + HERWIG	340 000
$W \to l\nu_l + b\bar{b}$ + 0 parton	57	ALPGEN + HERWIG	1 300 000
$W \to l\nu_l + b\bar{b}$ + 1 parton	43	ALPGEN + HERWIG	1 000 000
$W \to l\nu_l + b\bar{b}$ + 2 partons	21	ALPGEN + HERWIG	1 000 000
$W \to l\nu_l + b\bar{b}$ + 3 partons	9.1	ALPGEN + HERWIG	340 000

Table A.3: Monte Carlo simulation samples used in the analysis to model processes of W+jets production. The cross-sections include branching ratios and k-factors for NLO corrections. All cross-sections are rounded to two significant digits. The full ATLAS detector simulation is used in the generation process of all samples.

Process	$\sigma[pb]$	Generator	Events
$Z \to ee$ + 0 parton	840	ALPGEN + HERWIG	6 600 000
$Z \to ee$ + 1 parton	170	ALPGEN + HERWIG	1 300 000
$Z \to ee$ + 2 partons	51	ALPGEN + HERWIG	2 000 000
$Z \to ee$ + 3 partons	14	ALPGEN + HERWIG	550 000
$Z \to ee$ + 4 partons	3.6	ALPGEN + HERWIG	150 000
$Z \to ee$ + 5 partons	1.0	ALPGEN + HERWIG	50 000
$Z \to \mu\mu$ + 0 parton	840	ALPGEN + HERWIG	6 600 000
$Z \to \mu\mu$ + 1 parton	170	ALPGEN + HERWIG	1 300 000
$Z \to \mu\mu$ + 2 partons	50	ALPGEN + HERWIG	2 000 000
$Z \to \mu\mu$ + 3 partons	14	ALPGEN + HERWIG	550 000
$Z \to \mu\mu$ + 4 partons	3.4	ALPGEN + HERWIG	150 000
$Z \to \mu\mu$ + 5 partons	0.96	ALPGEN + HERWIG	50 000
$Z \to \tau\tau$ + 0 parton	840	ALPGEN + HERWIG	11 000 000
$Z \to \tau\tau$ + 1 parton	170	ALPGEN + HERWIG	3 300 000
$Z \to \tau\tau$ + 2 partons	50	ALPGEN + HERWIG	1 000 000
$Z \to \tau\tau$ + 3 partons	14	ALPGEN + HERWIG	510 000
$Z \to \tau\tau$ + 4 partons	3.5	ALPGEN + HERWIG	140 000
$Z \to \tau\tau$ + 5 partons	0.96	ALPGEN + HERWIG	45 000
$Z \to ee + b\bar{b}$ + 0 parton	8.2	ALPGEN + HERWIG	410 000
$Z \to ee + b\bar{b}$ + 1 parton	3.1	ALPGEN + HERWIG	160 000
$Z \to ee + b\bar{b}$ + 2 partons	1.1	ALPGEN + HERWIG	60 000
$Z \to ee + b\bar{b}$ + 3 partons	0.49	ALPGEN + HERWIG	30 000
$Z \to \mu\mu + b\bar{b}$ + 0 parton	8.2	ALPGEN + HERWIG	410 000
$Z \to \mu\mu + b\bar{b}$ + 1 parton	3.1	ALPGEN + HERWIG	160 000
$Z \to \mu\mu + b\bar{b}$ + 2 partons	1.1	ALPGEN + HERWIG	60 000
$Z \to \mu\mu + b\bar{b}$ + 3 partons	0.49	ALPGEN + HERWIG	30 000
$Z \to \tau\tau + b\bar{b}$ + 0 parton	8.2	ALPGEN + HERWIG	150 000
$Z \to \tau\tau + b\bar{b}$ + 1 parton	3.1	ALPGEN + HERWIG	100 000
$Z \to \tau\tau + b\bar{b}$ + 2 partons	1.1	ALPGEN + HERWIG	40 000
$Z \to \tau\tau + b\bar{b}$ + 3 partons	0.49	ALPGEN + HERWIG	9000
WW	17	HERWIG	2 500 000
WZ	5.5	HERWIG	1 000 000
ZZ	1.3	HERWIG	250 000

Table A.4: Monte Carlo simulation samples used in the analysis to model processes of Z+jets and diboson production. The cross-sections include branching ratios and k-factors for NLO corrections. All cross-sections are rounded to two significant digits. The full ATLAS detector simulation is used in the generation process of all samples.

APPENDIX **B**

Neural Network Input Variables

B.1 Variables used in single network training

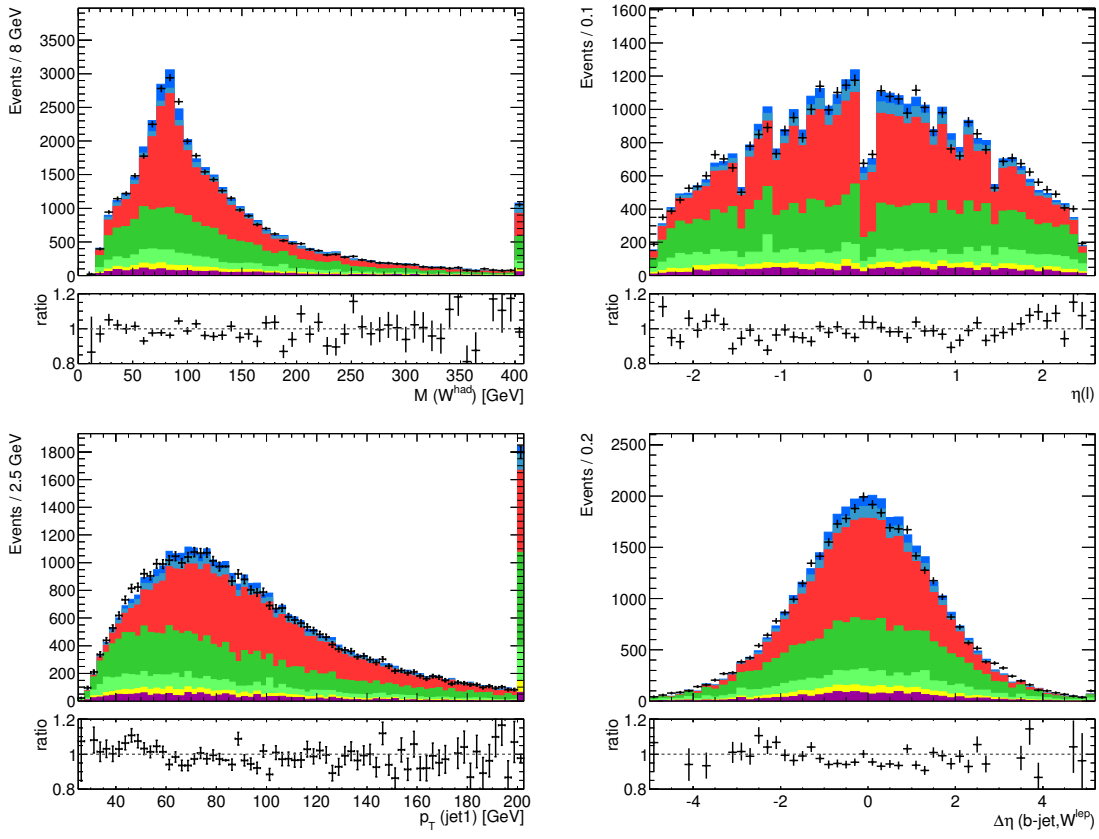


Figure B.1: Distributions of variables used in the single network training. Electron and muon channels are merged. The simulation distributions are normalized to the measured luminosity. The ratio of data and simulation is shown in the bottom plot.

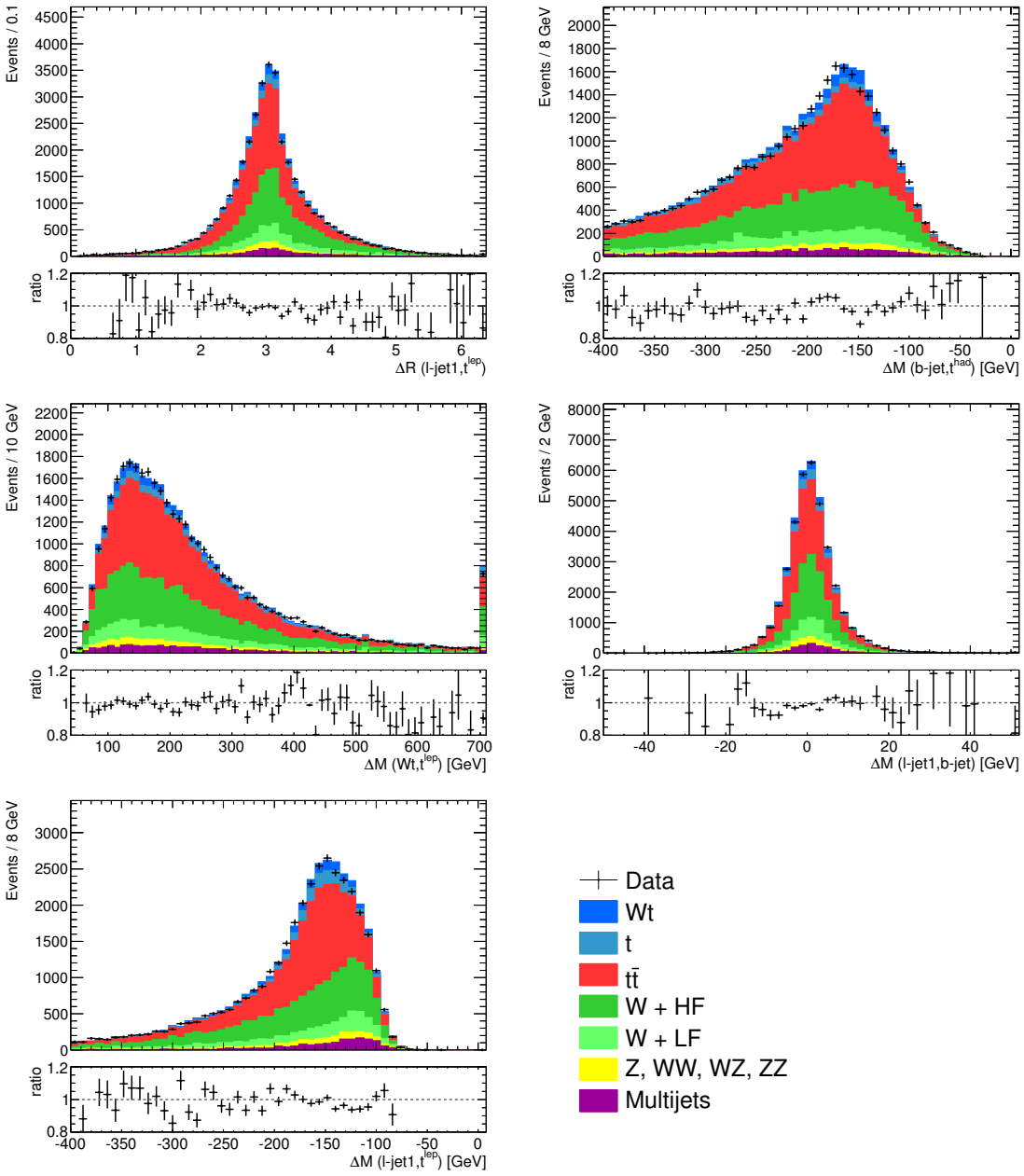


Figure B.2: Distributions of variables used in the single network training. Electron and muon channels are merged. The simulation distributions are normalized to the measured luminosity. The ratio of data and simulation is shown in the bottom plot.

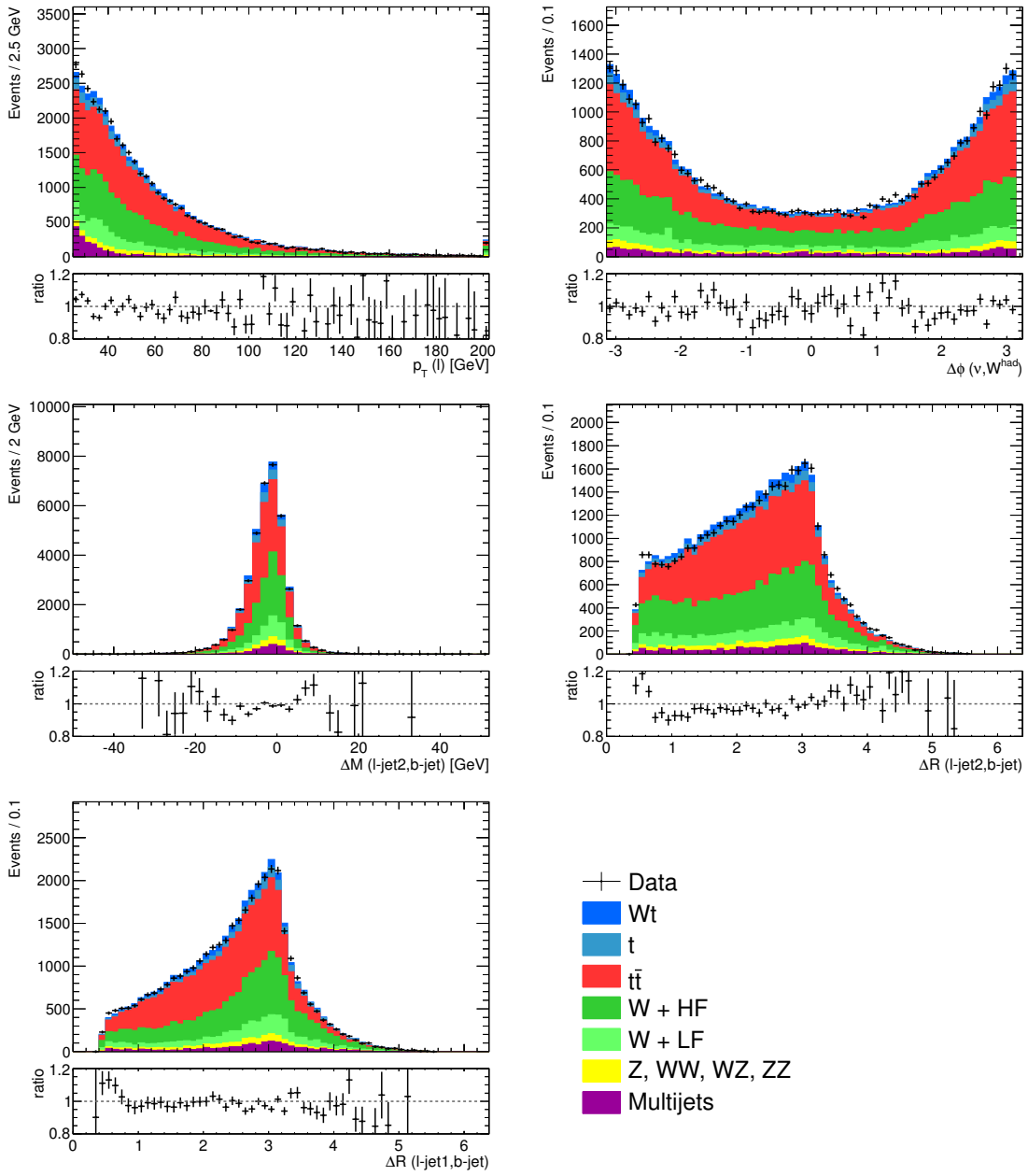


Figure B.3: Distributions of variables used in the single network training. Electron and muon channels are merged. The simulation distributions are normalized to the measured luminosity. The ratio of data and simulation is shown in the bottom plot.

B.2 Variables used in combined network trainings

B.2.1 Neural network training against top quark pair production

The variables $M(W^{\text{had}})$ and $\eta(l)$ are shown in Figure B.1, $\Delta M(\text{l-jet1}, \text{b-jet})$ is shown in Figure B.3.

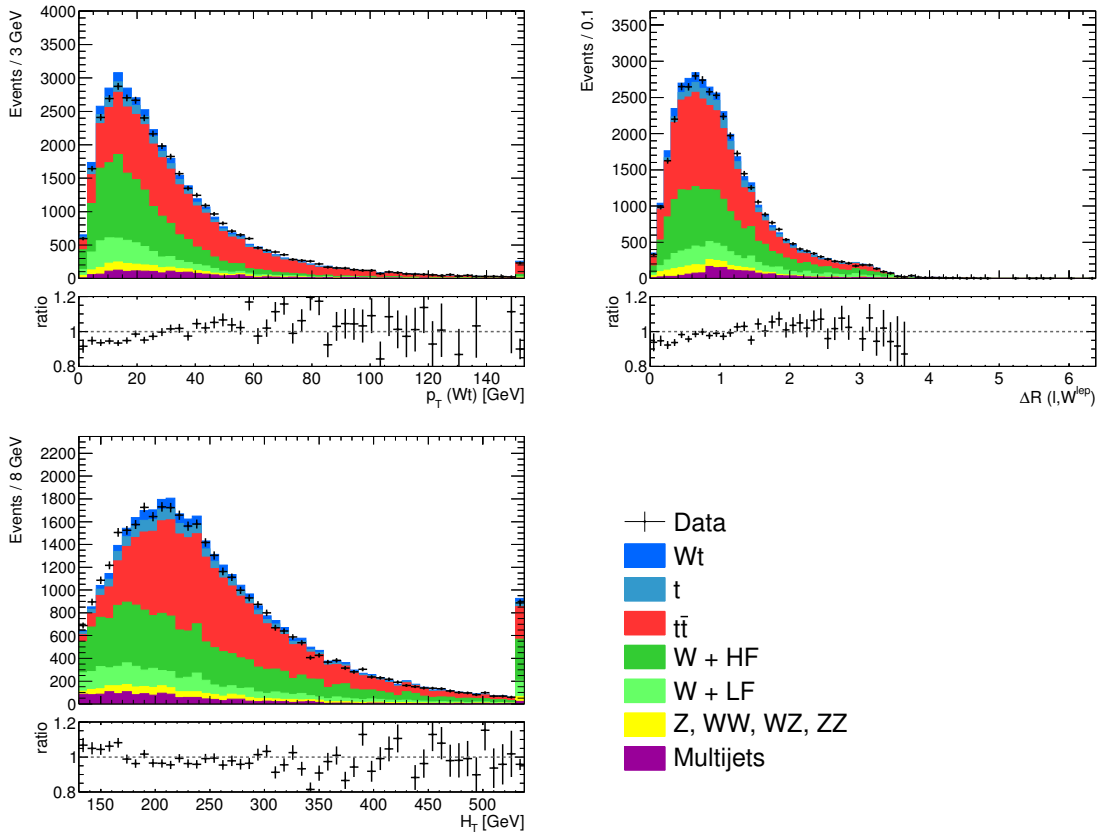


Figure B.4: Distributions of variables used in the network training against $t\bar{t}$ production. Electron and muon channels are merged. The simulation distributions are normalized to the measured luminosity. The ratio of data and simulation is shown in the bottom plot.

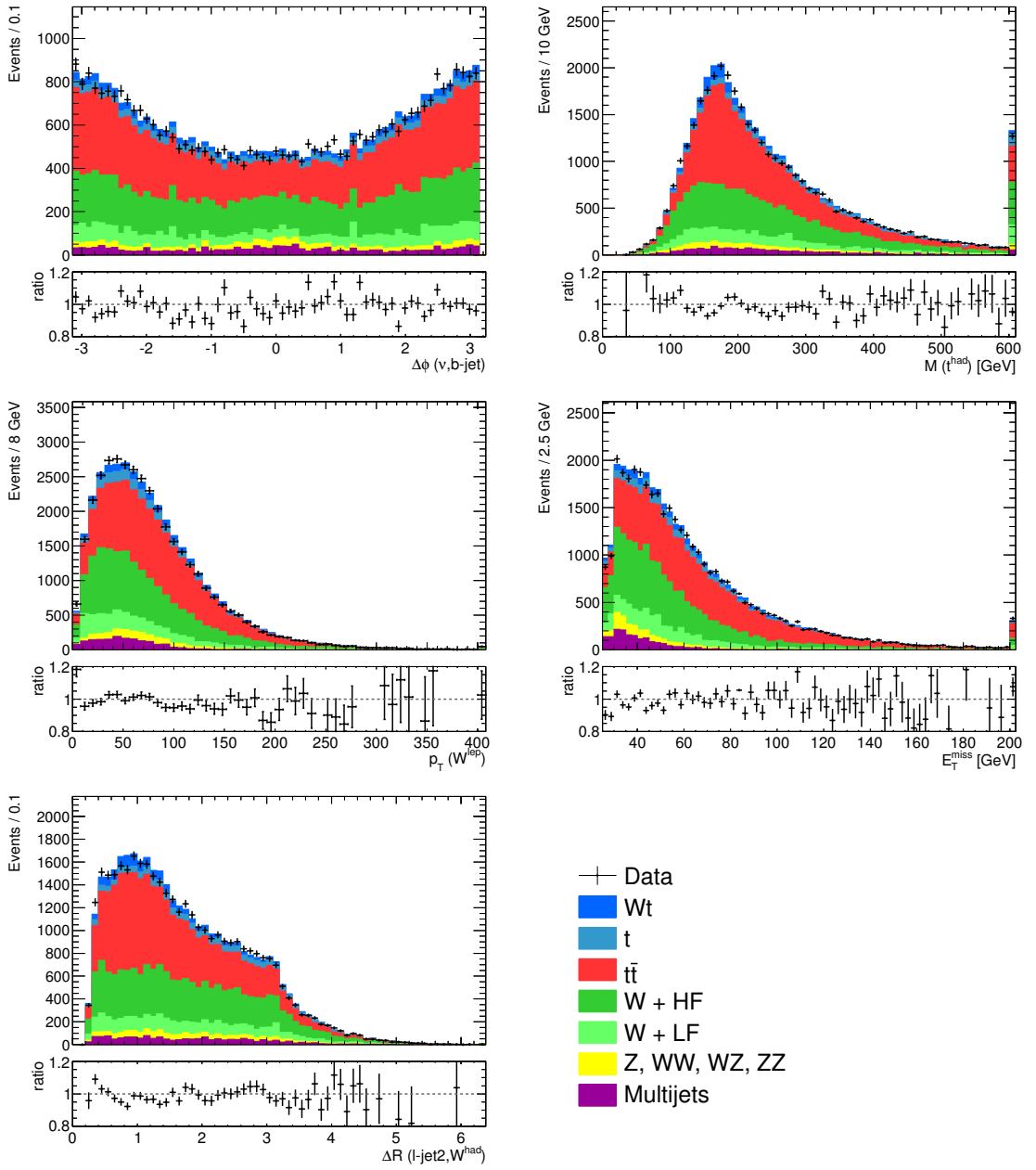


Figure B.5: Distributions of variables used in the network training against $t\bar{t}$ production. Electron and muon channels are merged. The simulation distributions are normalized to the measured luminosity. The ratio of data and simulation is shown in the bottom plot.

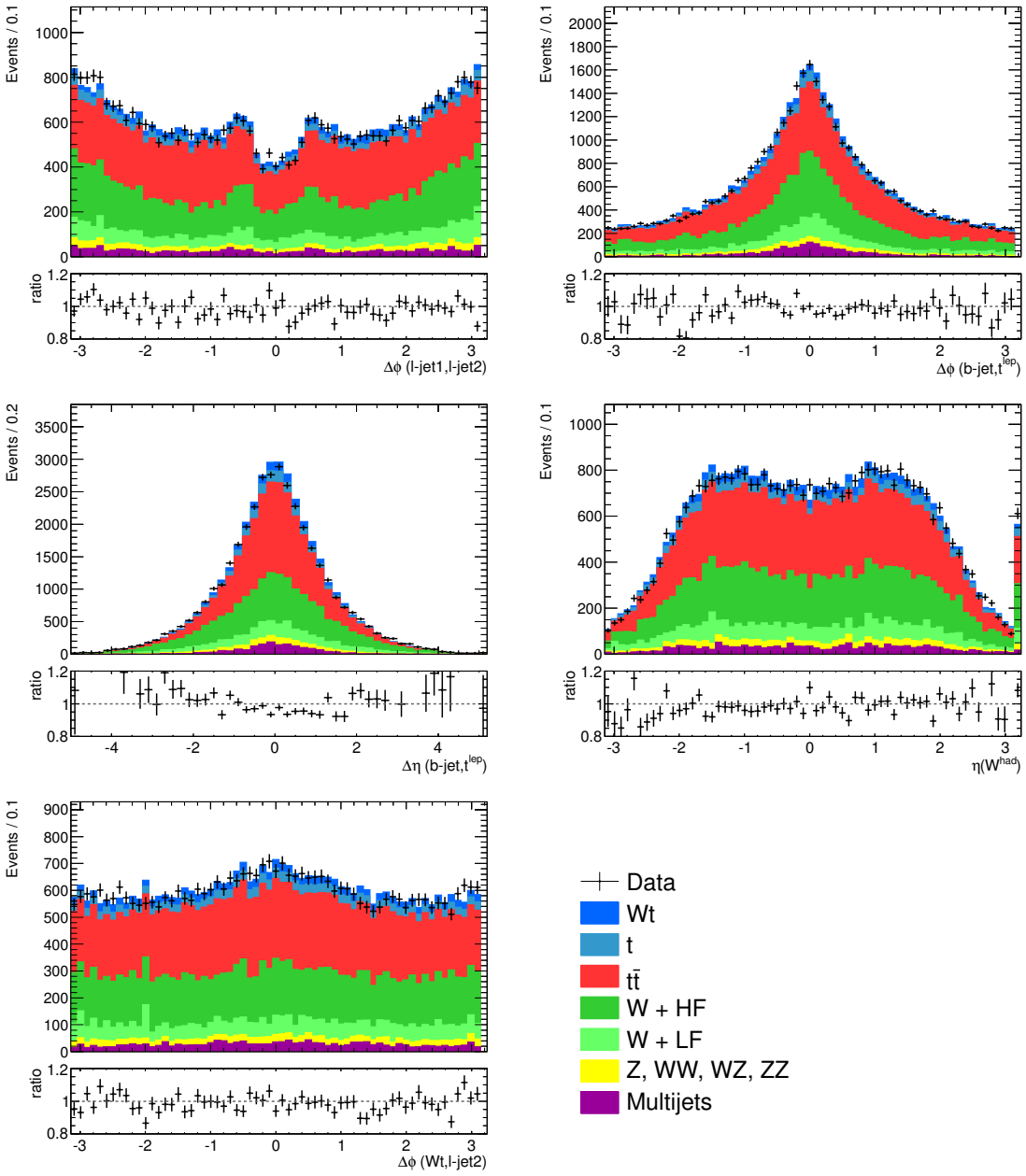


Figure B.6: Distributions of variables used in the network training against $t\bar{t}$ production. Electron and muon channels are merged. The simulation distributions are normalized to the measured luminosity. The ratio of data and simulation is shown in the bottom plot.

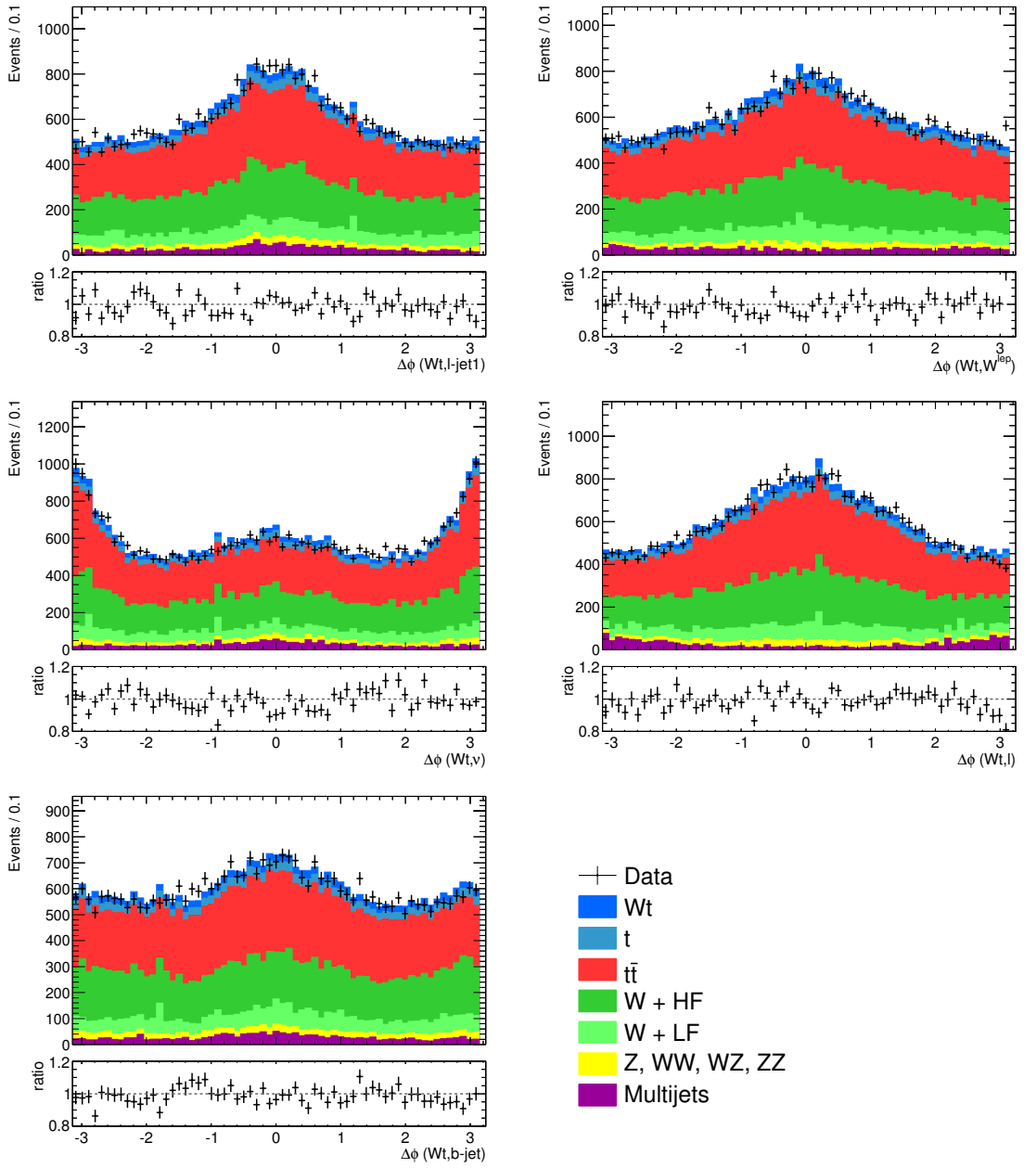


Figure B.7: Distributions of variables used in the network training against $t\bar{t}$ production. Electron and muon channels are merged. The simulation distributions are normalized to the measured luminosity. The ratio of data and simulation is shown in the bottom plot.

B.2.2 Neural network training against W+jets and other backgrounds

The variable $\eta(l)$ is shown in Figure B.1, ΔR(l-jet1, b-jet) and ΔR(l-jet2, b-jet) are shown in Figure B.3, $E_{\mathrm{T}}^{\mathrm{miss}}$ is shown in Figure B.5, $p_{\mathrm{T}}(Wt)$ is shown in Figure B.4 and ΔM(b-jet, t^{had}), ΔR(l-jet1, t^{lep}) and $\Delta M(Wt, t^{\mathrm{lep}})$ are shown in Figure B.2.

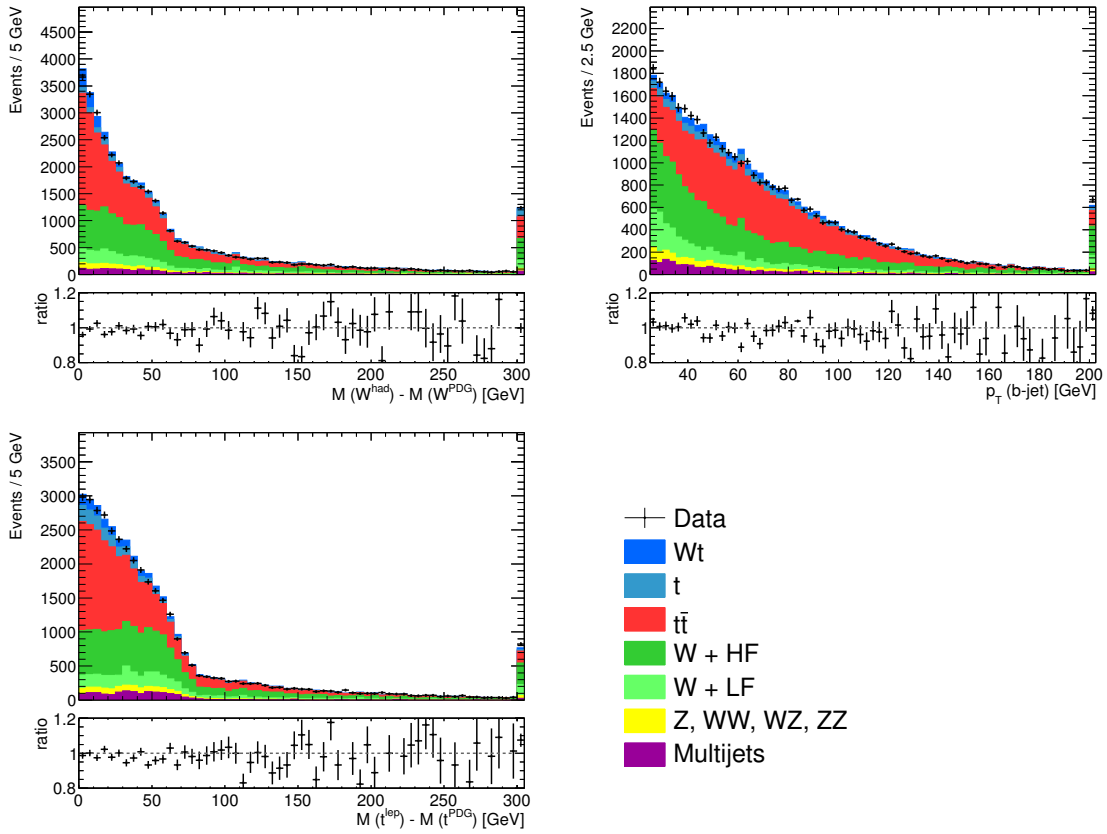


Figure B.8: Distributions of variables used in the network training against W+jets and other backgrounds. Electron and muon channels are merged. The simulation distributions are normalized to the measured luminosity. The ratio of data and simulation is shown in the bottom plot.

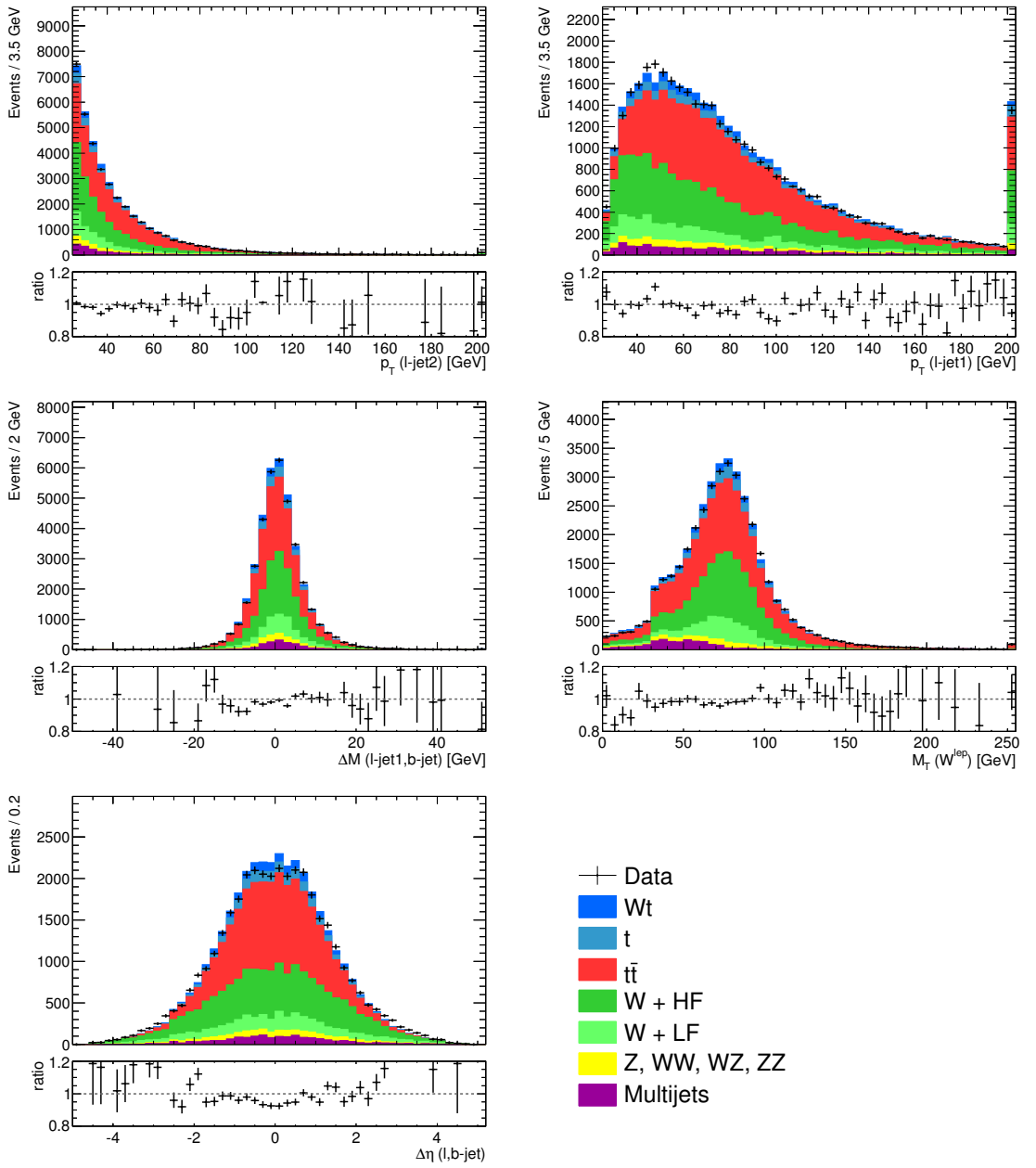


Figure B.9: Distributions of variables used in the network training against W+jets and other backgrounds. Electron and muon channels are merged. The simulation distributions are normalized to the measured luminosity. The ratio of data and simulation is shown in the bottom plot.

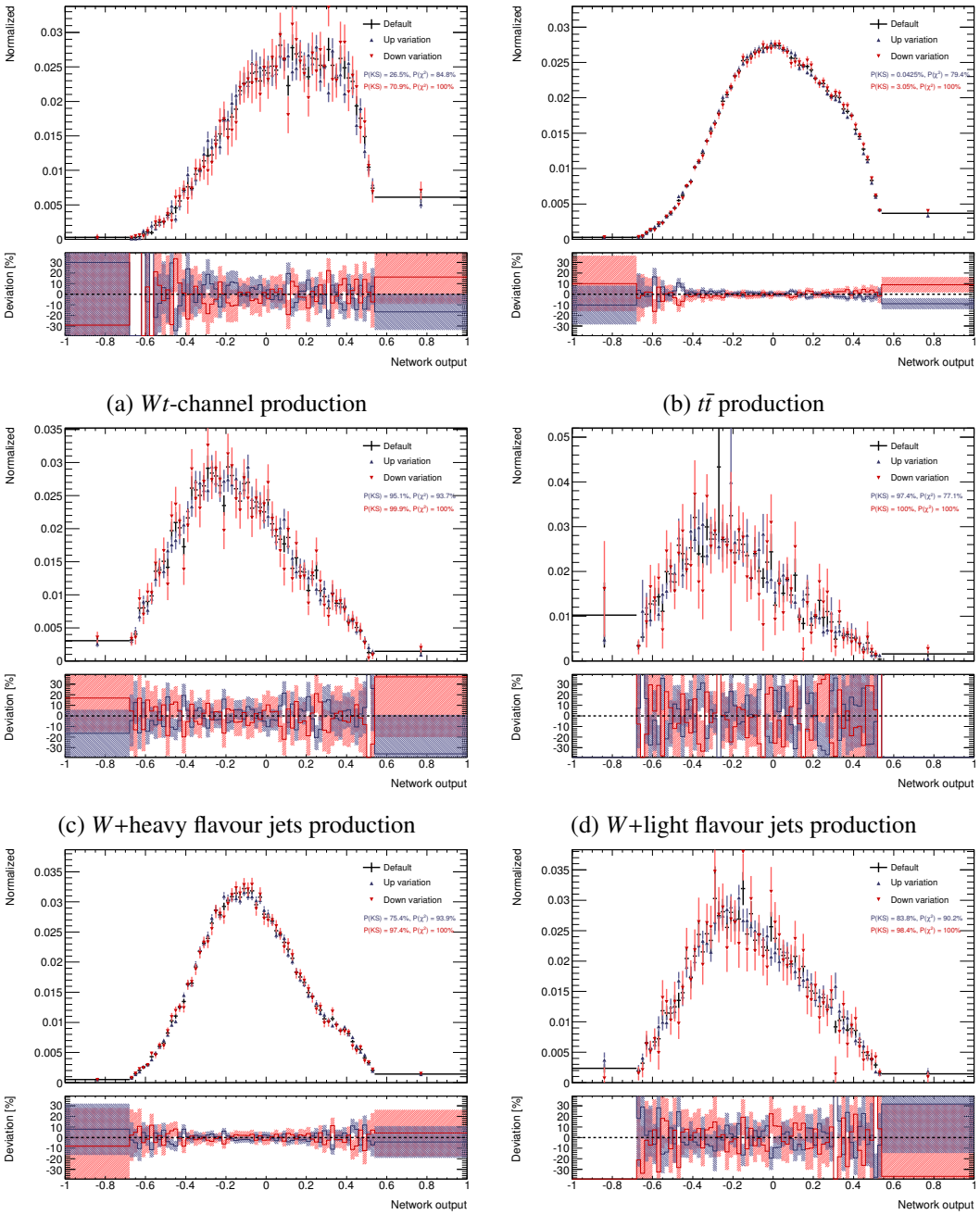
APPENDIX C

Systematic Variations of Neural Network Outputs

This appendix shows the comparisons of the neural network output distributions with the systematic variations that lead to the strongest uncertainties separately for each combination of production channels that is used as template in the binned likelihood fit.

C.1 Single neural network

C.1.1 Jet energy resolution



(a) Wt-channel production

(b) $t\bar{t}$ production

(c) W+heavy flavour jets production

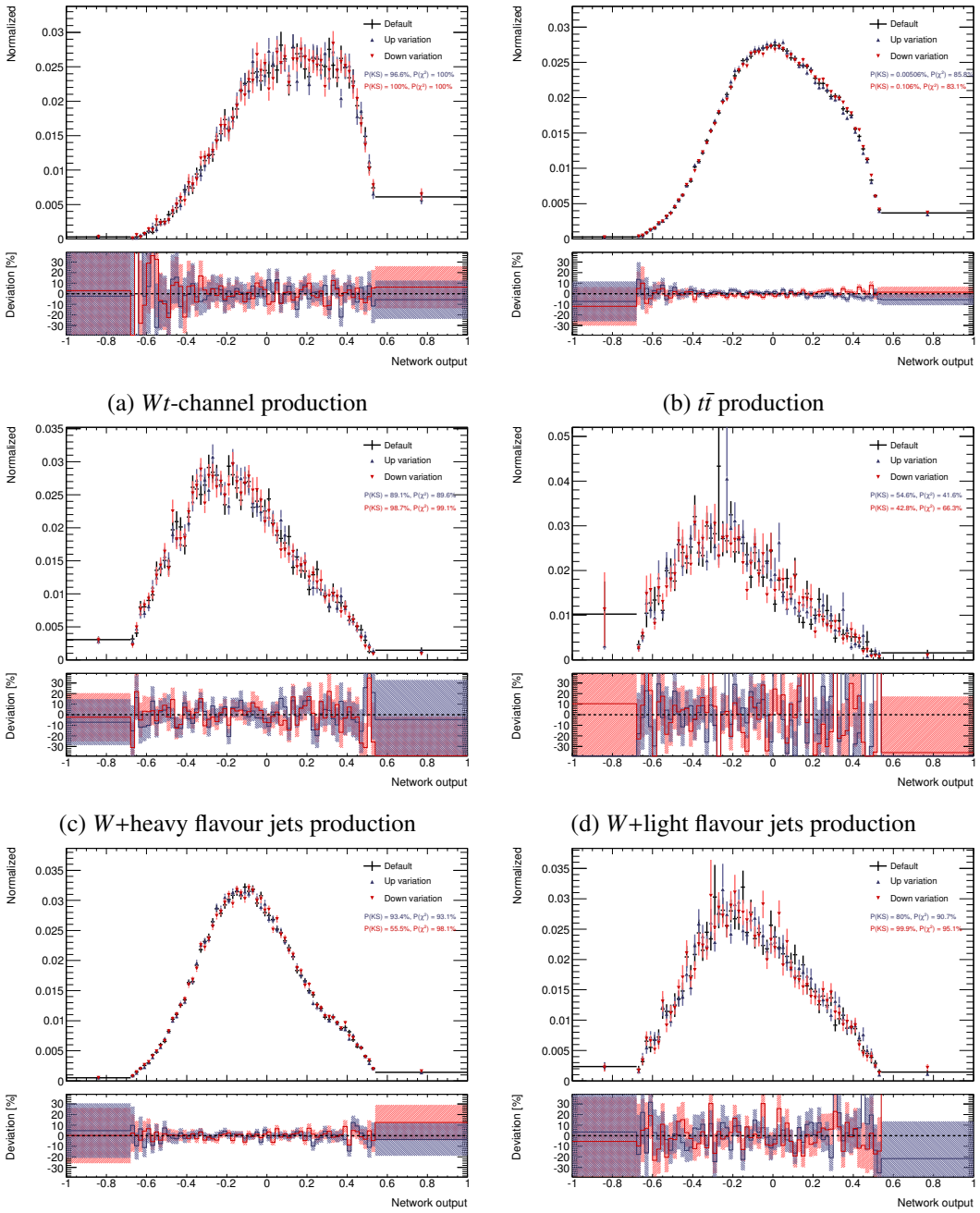
(d) W+light flavour jets production

(e) t-channel and s-channel production

(f) Z+jets, WW, WZ, ZZ

Figure C.1: Distributions of the single neural network output for different production channels normalized to unit area used as template in the fit. Each plot shows the default distribution compared to the upward and downward variations of the jet energy resolution systematic. The shown error bars correspond to the statistical uncertainty of the available Monte Carlo sample only. Probabilities resulting from a Kolmogorov-Smirnov test and a χ^2 test for the agreement between variation and default are given for both variations. The lower plot shows the deviation of the systematics including the combined statistical errors as hashed areas.

C.1.2 Jet energy scale



(a) Wt-channel production

(b) $t\bar{t}$ production

(c) W+heavy flavour jets production

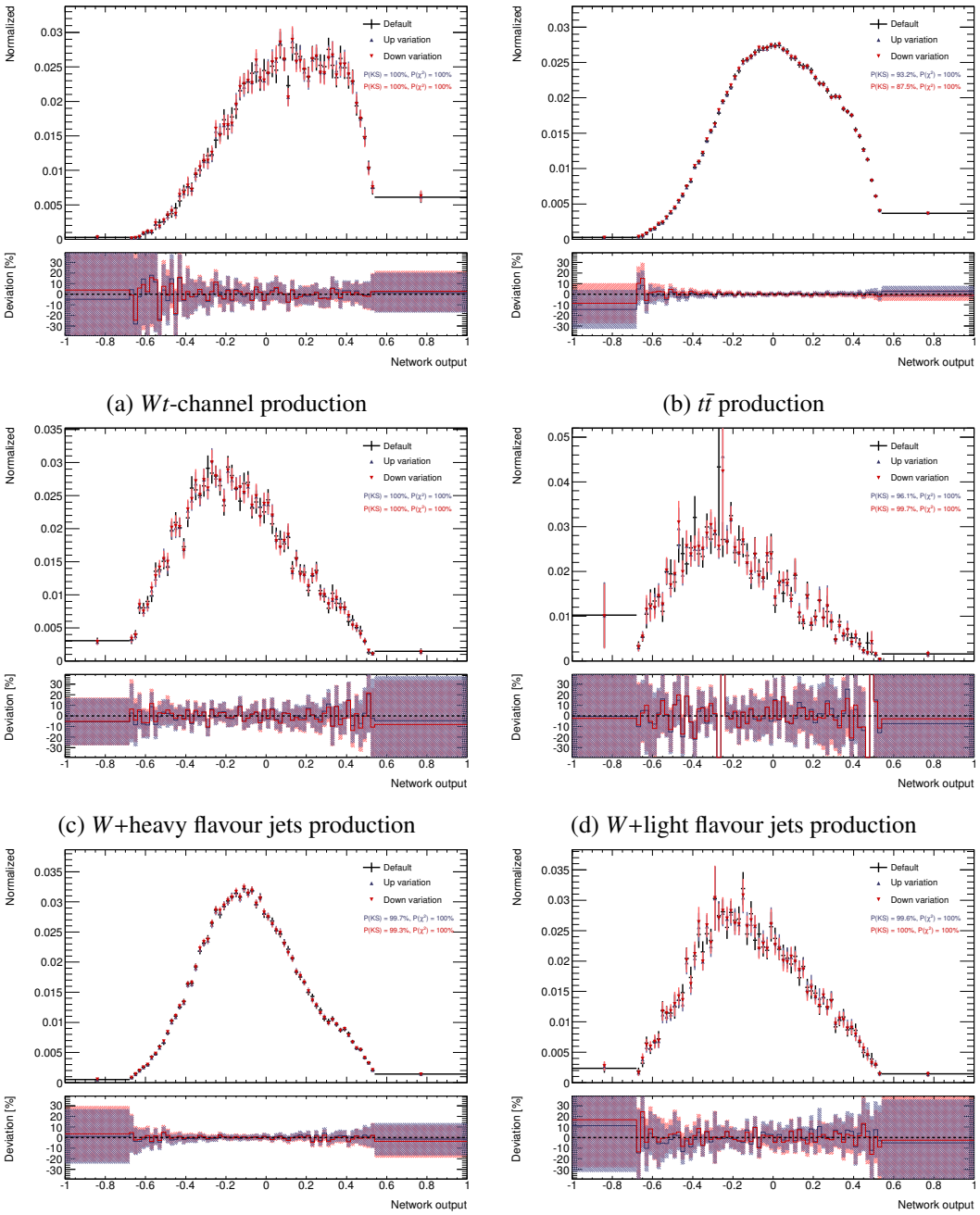
(d) W+light flavour jets production

(e) t-channel and s-channel production

(f) Z+jets, WW, WZ, ZZ

Figure C.2: Distributions of the single neural network output for different production channels normalized to unit area used as template in the fit. Each plot shows the default distribution compared to the upward and downward variations of the jet energy scale systematic. The shown error bars correspond to the statistical uncertainty of the available Monte Carlo sample only. Probabilities resulting from a Kolmogorov-Smirnov test and a χ^2 test for the agreement between variation and default are given for both variations. The lower plot shows the deviation of the systematics including the combined statistical errors as hashed areas.

C.1.3 Jet vertex fraction



(a) Wt-channel production

(b) $t\bar{t}$ production

(c) W+heavy flavour jets production

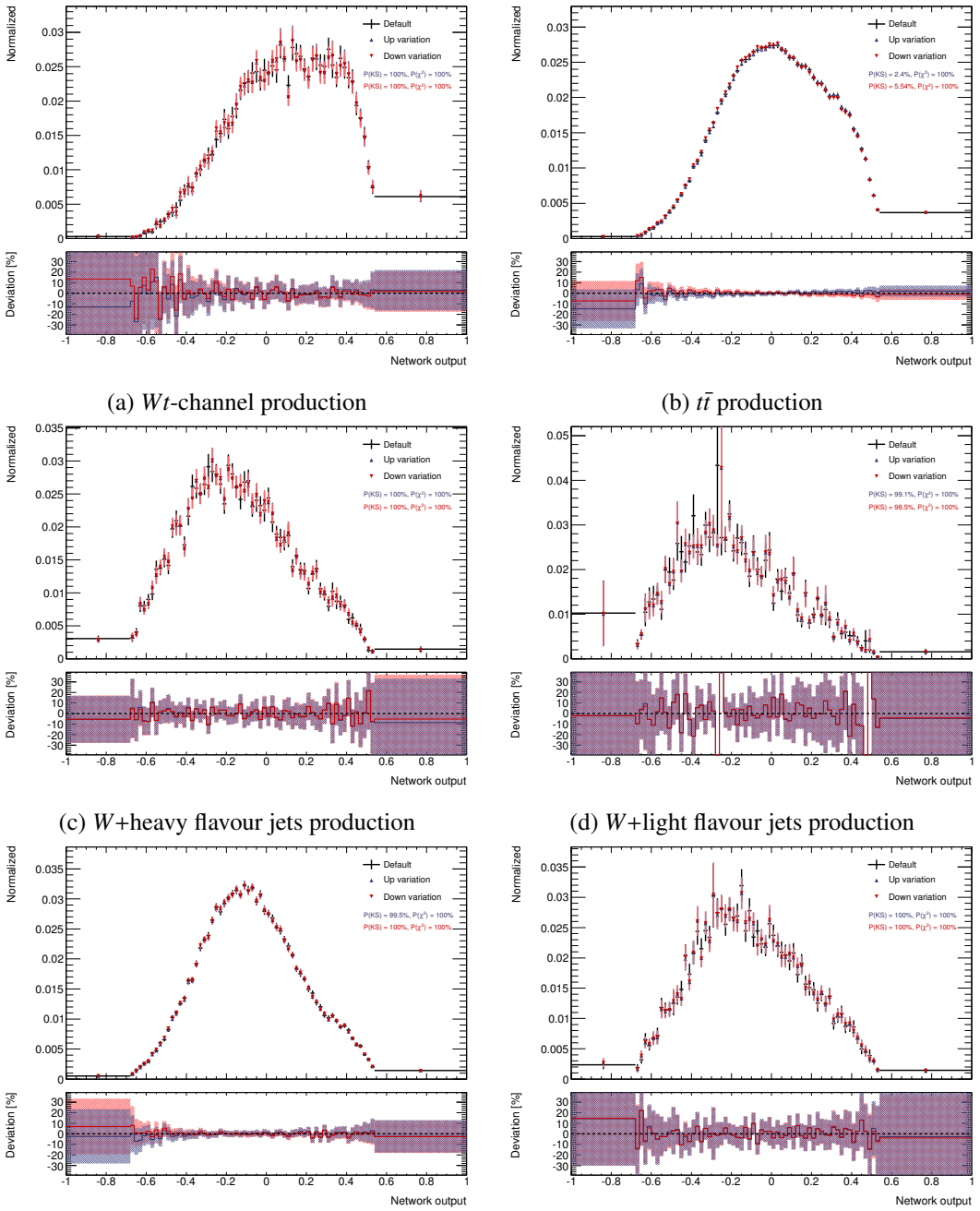
(d) W+light flavour jets production

(e) t-channel and s-channel production

(f) Z+jets, WW, WZ, ZZ

Figure C.3: Distributions of the single neural network output for different production channels normalized to unit area used as template in the fit. Each plot shows the default distribution compared to the upward and downward variations of the jet energy scale systematic. The shown error bars correspond to the statistical uncertainty of the available Monte Carlo sample only. Probabilities resulting from a Kolmogorov-Smirnov test and a χ^2 test for the agreement between variation and default are given for both variations. The lower plot shows the deviation of the systematics including the combined statistical errors as hashed areas.

C.1.4 b-tagging (b quark)



(a) Wt-channel production

(b) $t\bar{t}$ production

(c) W+heavy flavour jets production

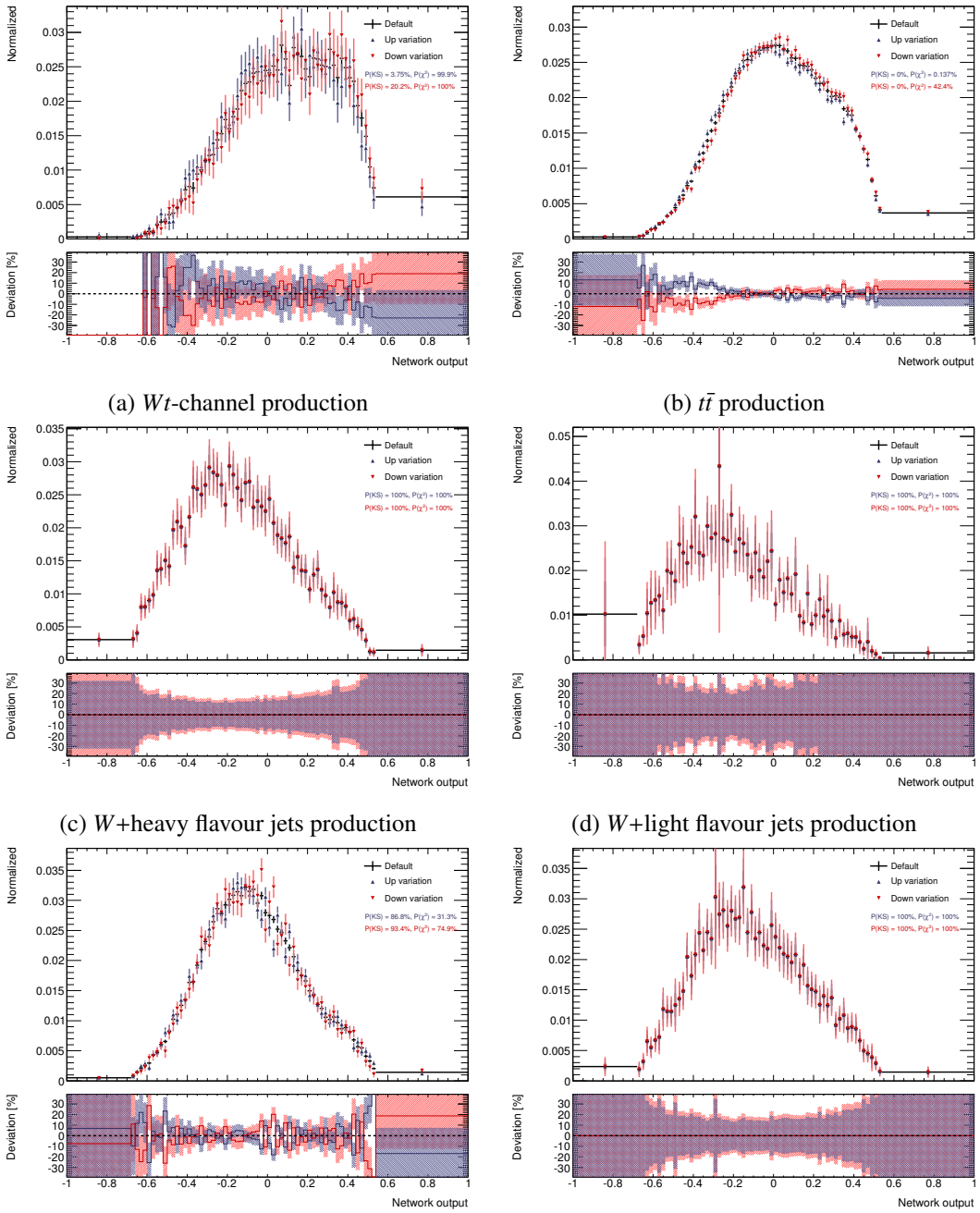
(d) W+light flavour jets production

(e) t-channel and s-channel production

(f) Z+jets, WW, WZ, ZZ

Figure C.4: Distributions of the single neural network output for different production channels normalized to unit area used as template in the fit. Each plot shows the default distribution compared to the upward and downward variations of the b-tagging systematic of real b quarks. The shown error bars correspond to the statistical uncertainty of the available Monte Carlo sample only. Probabilities resulting from a Kolmogorov-Smirnov test and a χ^2 test for the agreement between variation and default are given for both variations. The lower plot shows the deviation of the systematics including the combined statistical errors as hashed areas.

C.1.5 Initial-state and final-state radiation



(a) Wt-channel production

(b) $t\bar{t}$ production

(c) W+heavy flavour jets production

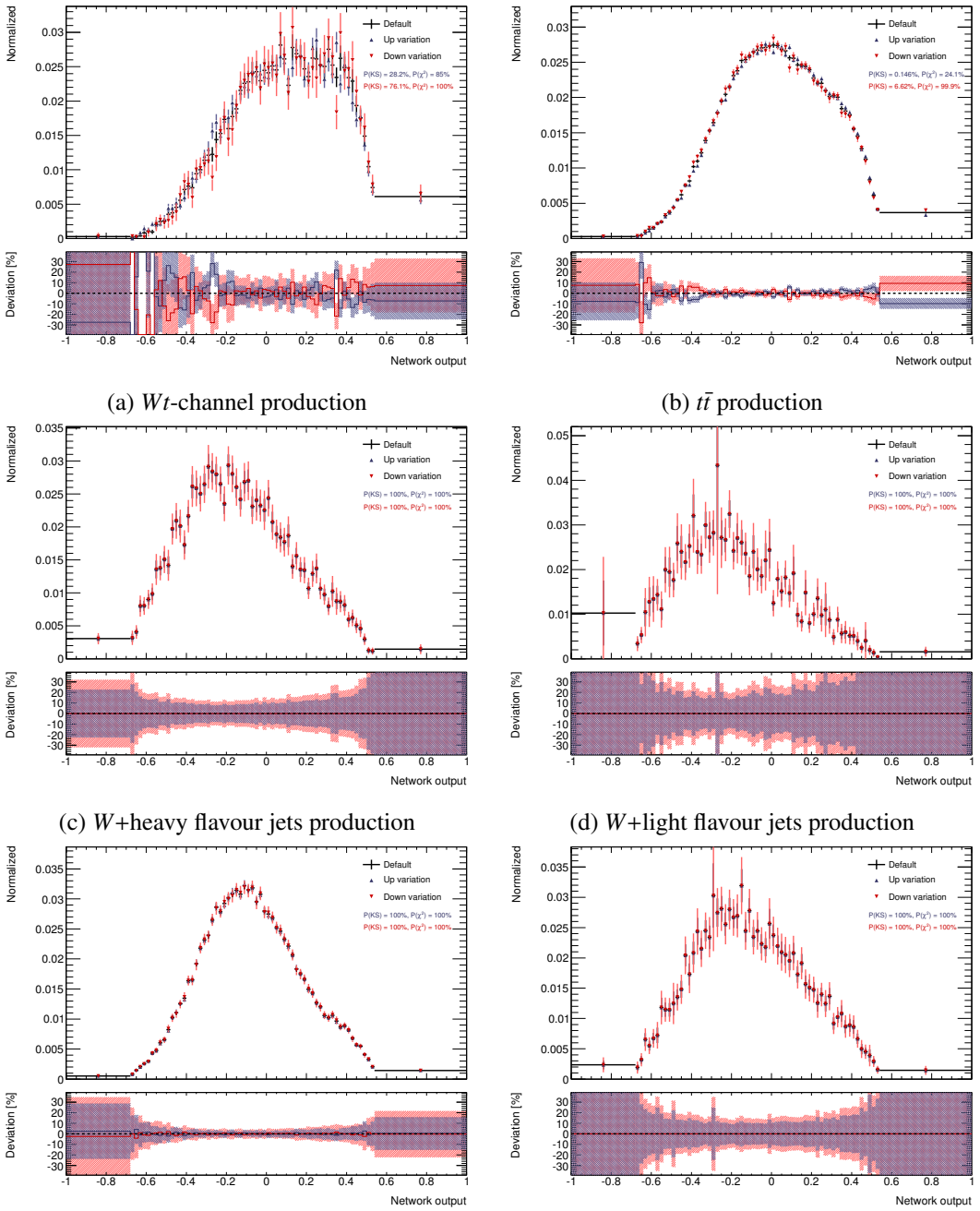
(d) W+light flavour jets production

(e) t-channel and s-channel production

(f) Z+jets, WW, WZ, ZZ

Figure C.5: Distributions of the single neural network output for different production channels normalized to unit area used as template in the fit. Each plot shows the default distribution compared to the upward and downward variations of the initial-state and final-state radiation systematic. The shown error bars correspond to the statistical uncertainty of the available Monte Carlo sample only. Probabilities resulting from a Kolmogorov-Smirnov test and a χ^2 test for the agreement between variation and default are given for both variations. The lower plot shows the deviation of the systematics including the combined statistical errors as hashed areas.

C.1.6 POWHEG+PYTHIA generator



(a) Wt-channel production

(b) $t\bar{t}$ production

(c) W+heavy flavour jets production

(d) W+light flavour jets production

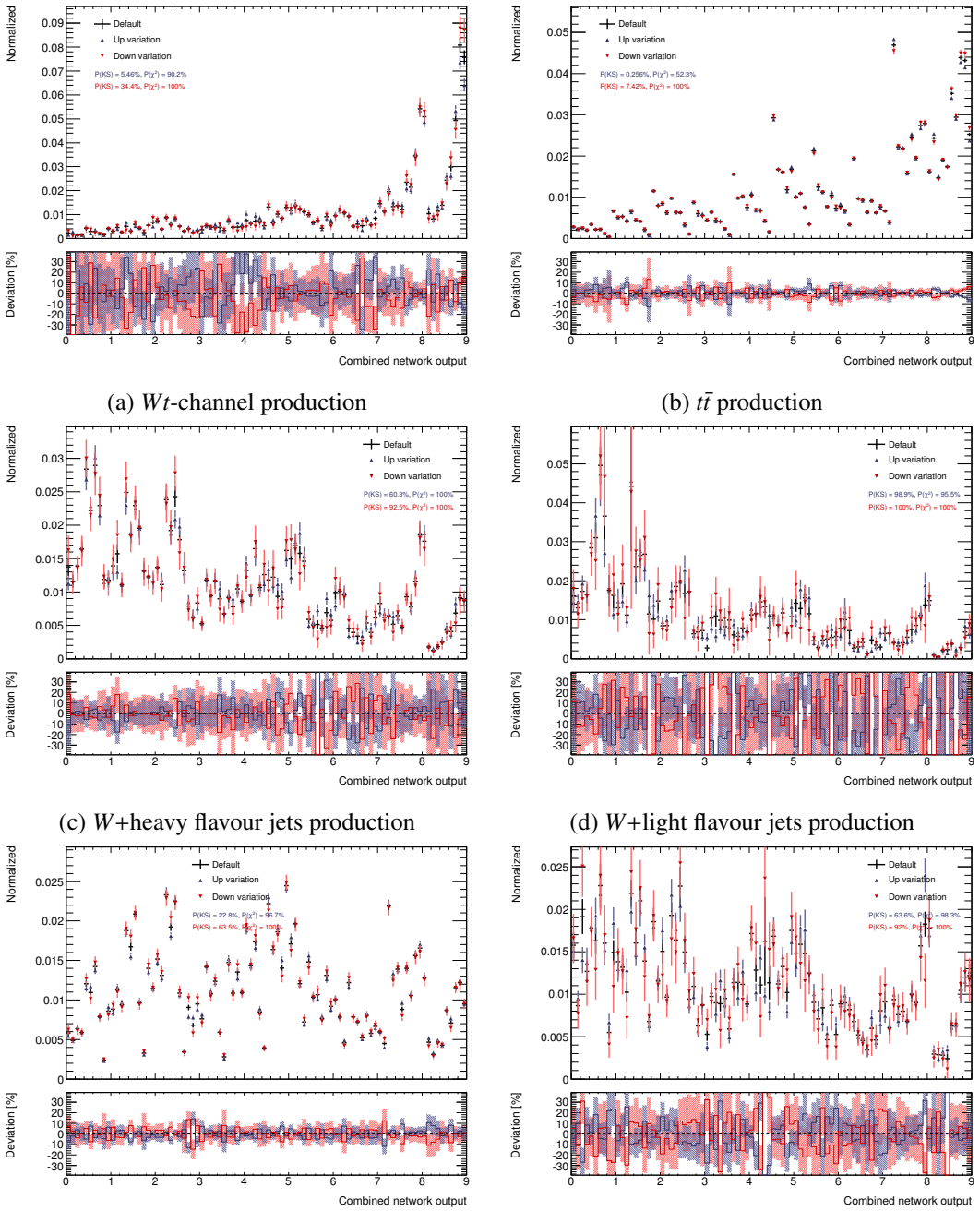
(e) t-channel and s-channel production

(f) Z+jets, WW, WZ, ZZ

Figure C.6: Distributions of the single neural network output for different production channels normalized to unit area used as template in the fit. Each plot shows the default distribution compared to the upward and downward variations of the POWHEG+PYTHIA generator systematic. The shown error bars correspond to the statistical uncertainty of the available Monte Carlo sample only. Probabilities resulting from a Kolmogorov-Smirnov test and a χ^2 test for the agreement between variation and default are given for both variations. The lower plot shows the deviation of the systematics including the combined statistical errors as hashed areas.

C.2 Combined neural network

C.2.1 Jet energy resolution



(a) Wt-channel production

(b) $t\bar{t}$ production

(c) W+heavy flavour jets production

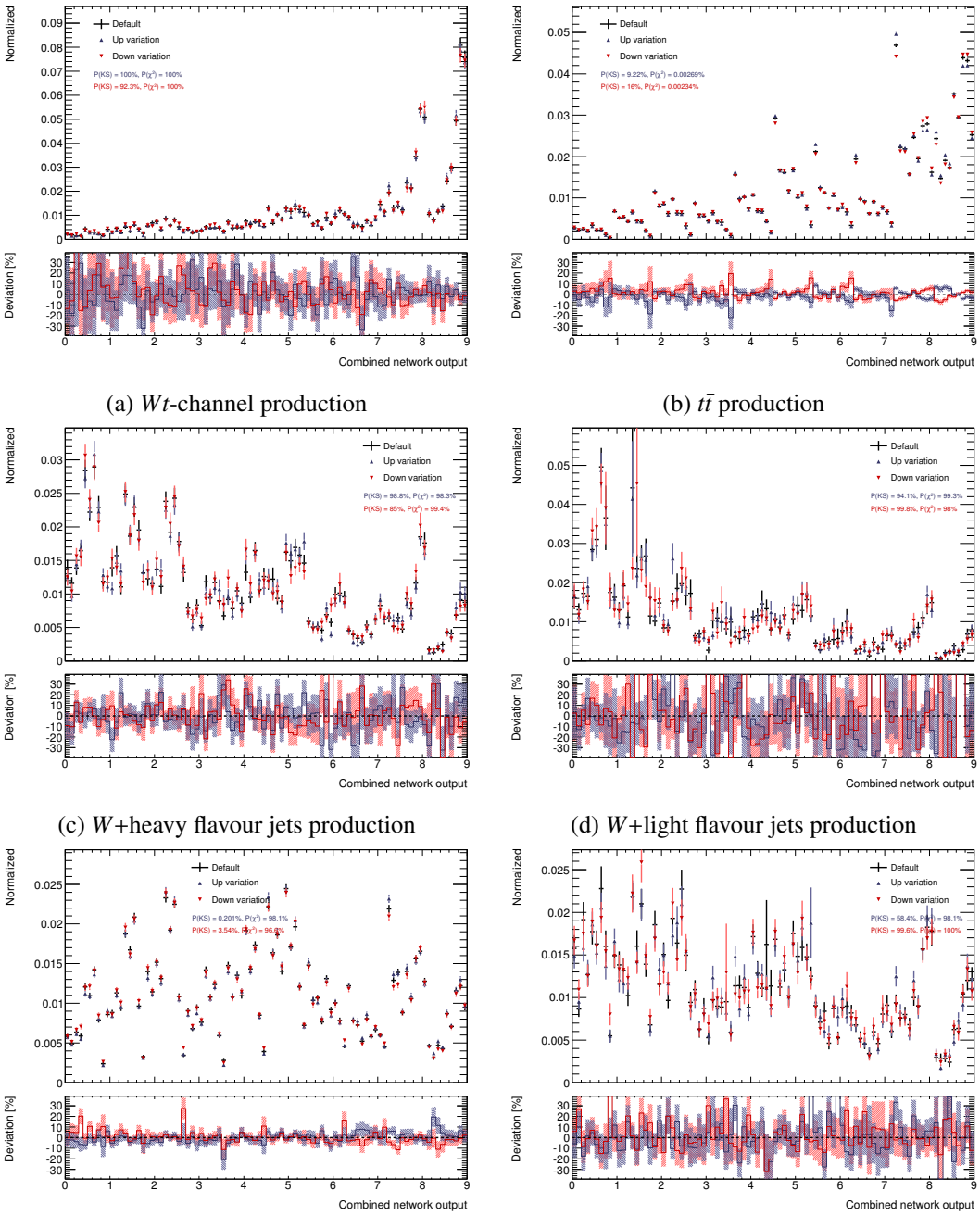
(d) W+light flavour jets production

(e) t-channel and s-channel production

(f) Z+jets, WW, WZ, ZZ

Figure C.7: Distributions of the combined neural network output for different production channels normalized to unit area used as template in the fit. Each plot shows the default distribution compared to the upward and downward variations of the jet energy resolution systematic. The shown error bars correspond to the statistical uncertainty of the available Monte Carlo sample only. Probabilities resulting from a Kolmogorov-Smirnov test and a χ^2 test for the agreement between variation and default are given for both variations. The lower plot shows the deviation of the systematics including the combined statistical errors as hashed areas.

C.2.2 Jet energy scale



(a) Wt-channel production

(b) $t\bar{t}$ production

(c) W+heavy flavour jets production

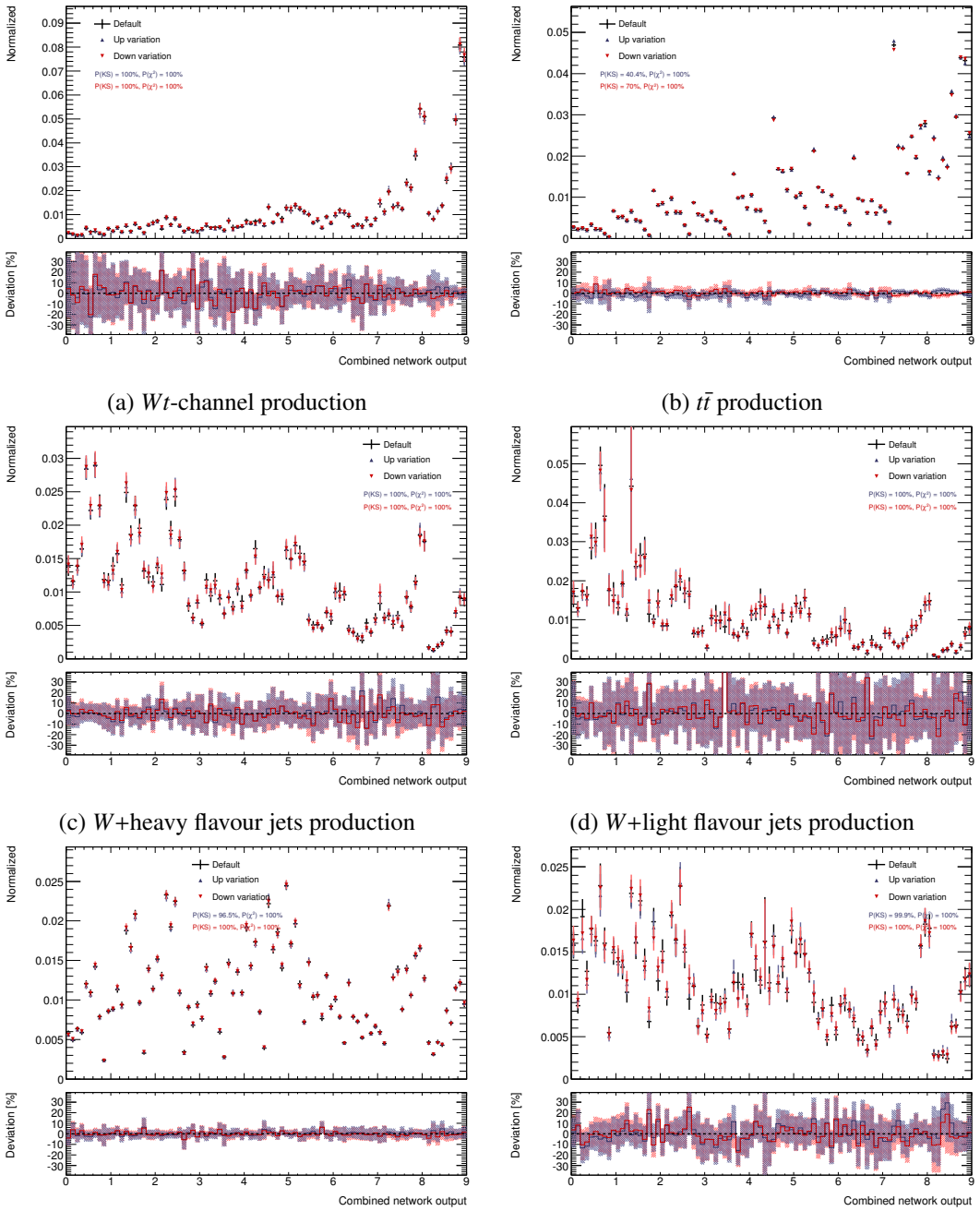
(d) W+light flavour jets production

(e) t-channel and s-channel production

(f) Z+jets, WW, WZ, ZZ

Figure C.8: Distributions of the combined neural network output for different production channels normalized to unit area used as template in the fit. Each plot shows the default distribution compared to the upward and downward variations of the jet energy scale systematic. The shown error bars correspond to the statistical uncertainty of the available Monte Carlo sample only. Probabilities resulting from a Kolmogorov-Smirnov test and a χ^2 test for the agreement between variation and default are given for both variations. The lower plot shows the deviation of the systematics including the combined statistical errors as hashed areas.

C.2.3 Jet vertex fraction



(a) Wt-channel production

(b) $t\bar{t}$ production

(c) W+heavy flavour jets production

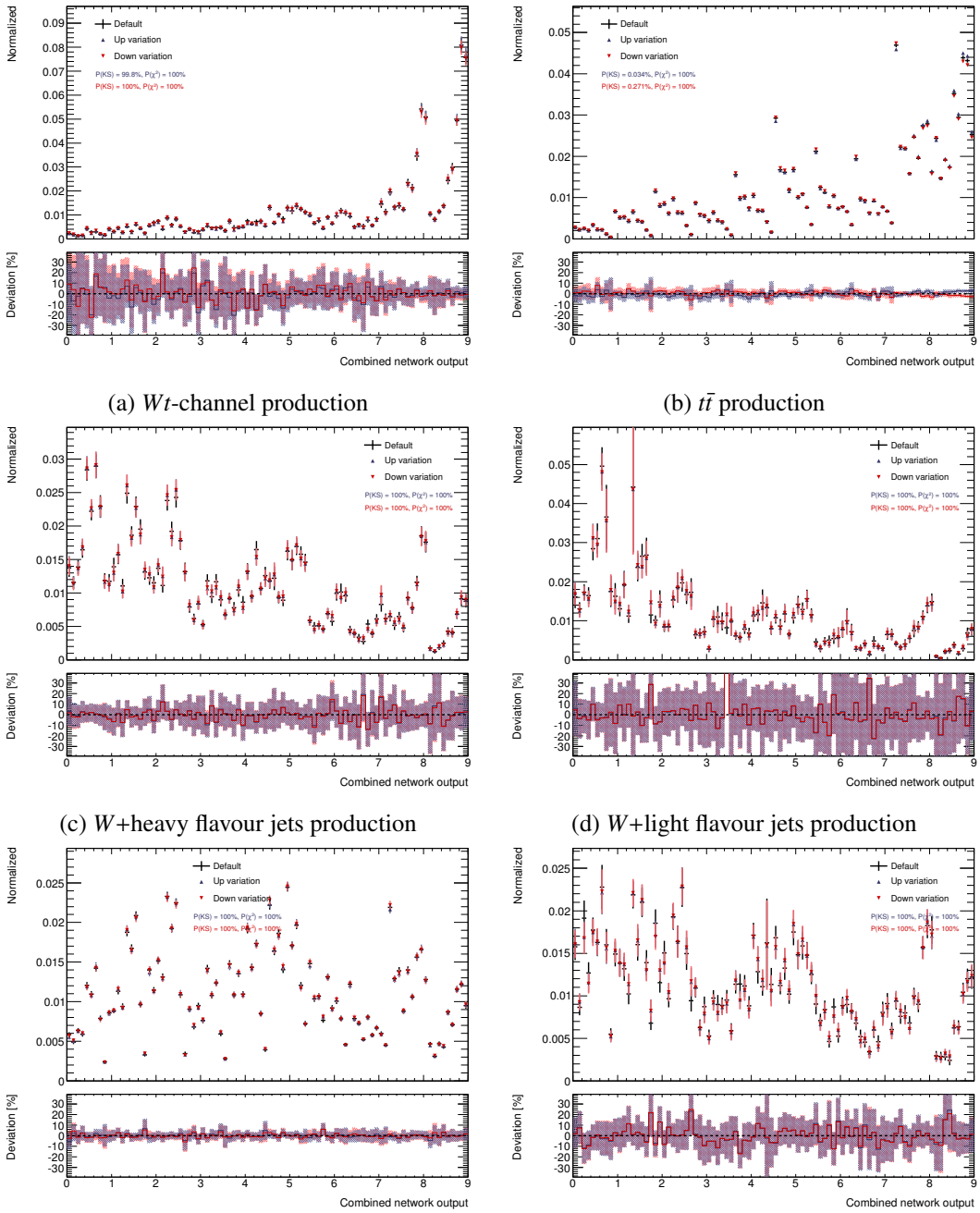
(d) W+light flavour jets production

(e) t-channel and s-channel production

(f) Z+jets, WW, WZ, ZZ

Figure C.9: Distributions of the combined neural network output for different production channels normalized to unit area used as template in the fit. Each plot shows the default distribution compared to the upward and downward variations of the jet energy scale systematic. The shown error bars correspond to the statistical uncertainty of the available Monte Carlo sample only. Probabilities resulting from a Kolmogorov-Smirnov test and a χ^2 test for the agreement between variation and default are given for both variations. The lower plot shows the deviation of the systematics including the combined statistical errors as hashed areas.

C.2.4 b-tagging (b quark)



(a) Wt-channel production

(b) $t\bar{t}$ production

(c) W+heavy flavour jets production

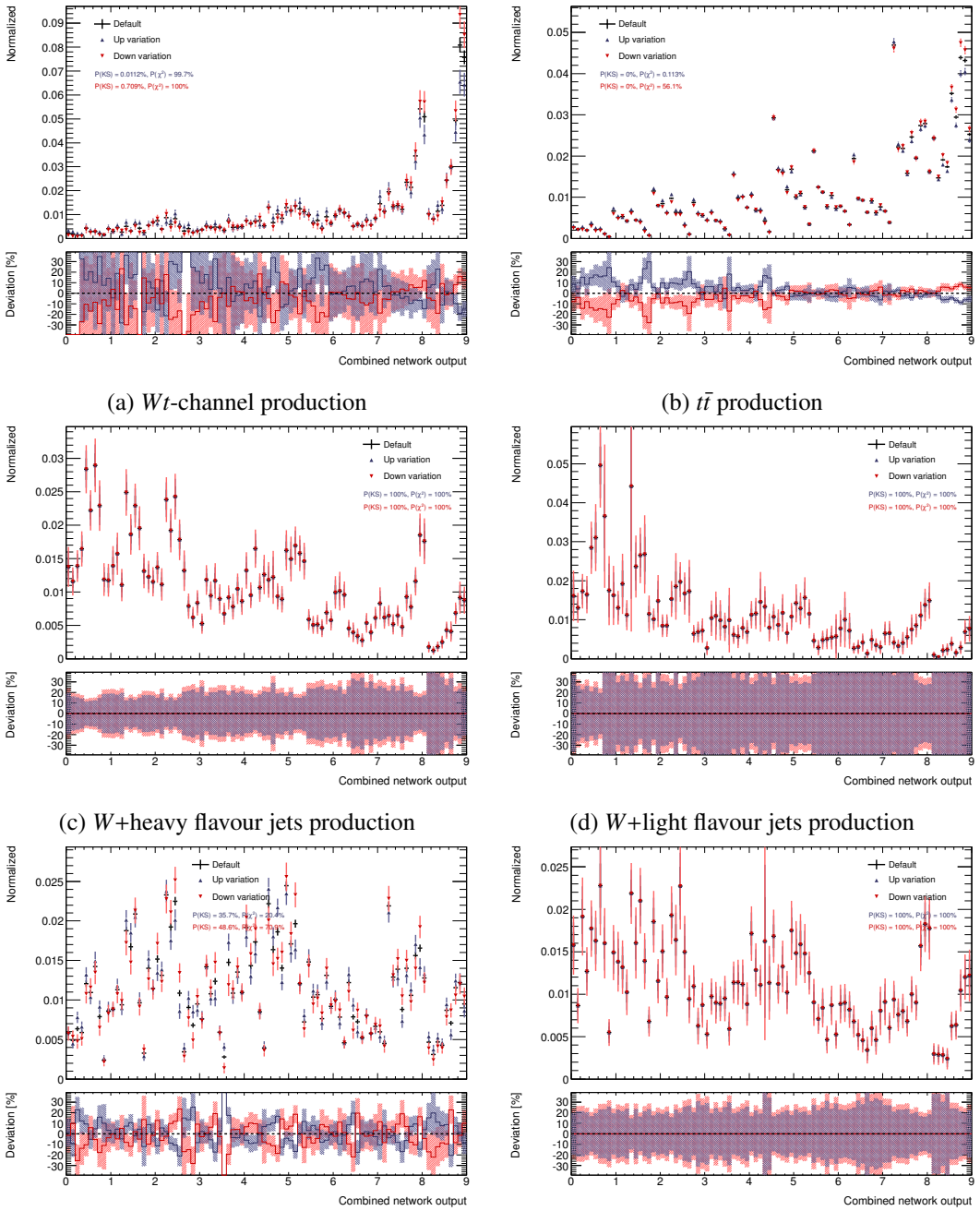
(d) W+light flavour jets production

(e) t-channel and s-channel production

(f) Z+jets, WW, WZ, ZZ

Figure C.10: Distributions of the combined neural network output for different production channels normalized to unit area used as template in the fit. Each plot shows the default distribution compared to the upward and downward variations of the b-tagging systematic of real b quarks. The shown error bars correspond to the statistical uncertainty of the available Monte Carlo sample only. Probabilities resulting from a Kolmogorov-Smirnov test and a χ^2 test for the agreement between variation and default are given for both variations. The lower plot shows the deviation of the systematics including the combined statistical errors as hashed areas.

C.2.5 Initial-state and final-state radiation



(a) Wt-channel production

(b) $t\bar{t}$ production

(c) W+heavy flavour jets production

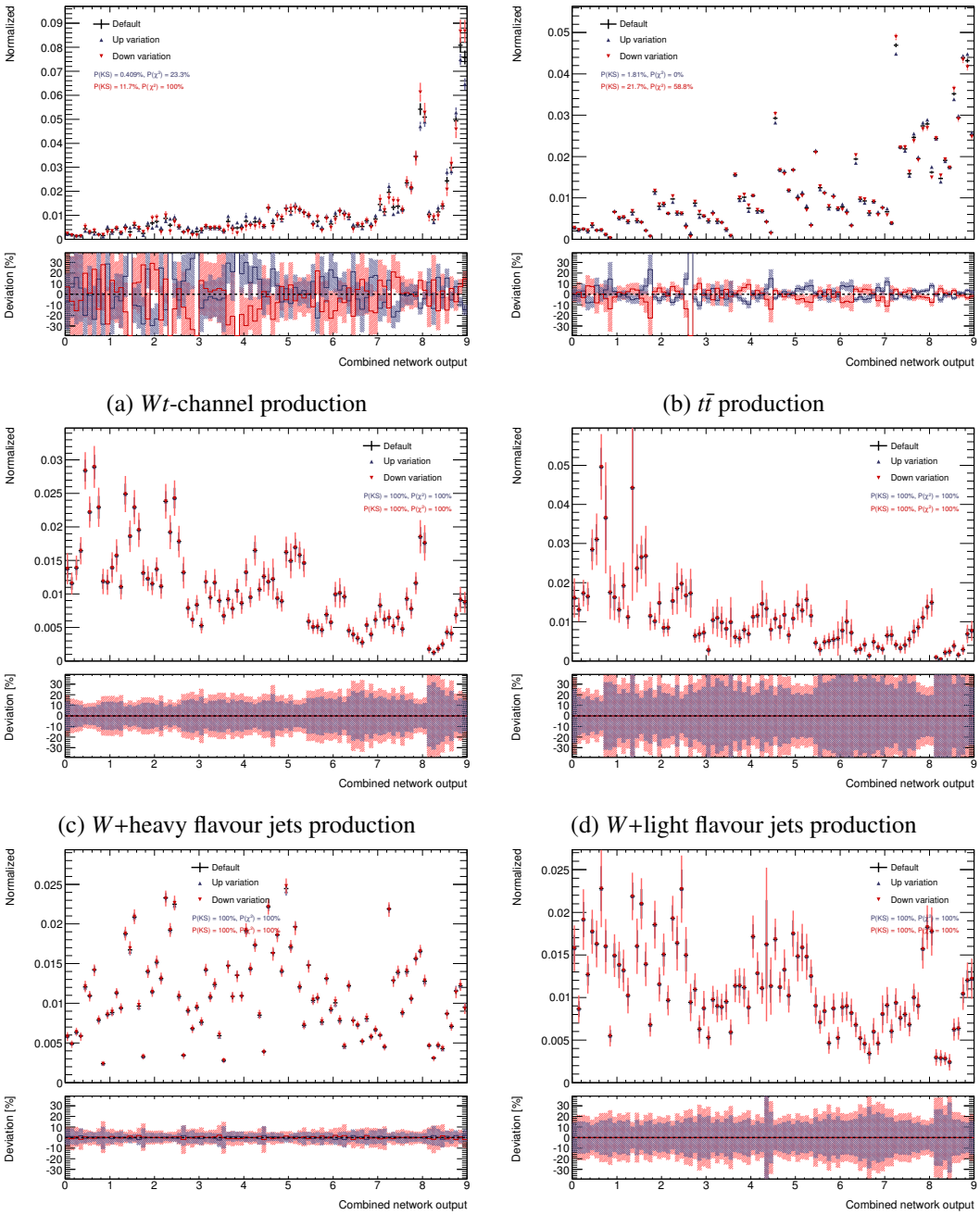
(d) W+light flavour jets production

(e) t-channel and s-channel production

(f) Z+jets, WW, WZ, ZZ

Figure C.11: Distributions of the combined neural network output for different production channels normalized to unit area used as template in the fit. Each plot shows the default distribution compared to the upward and downward variations of the initial-state and final-state radiation systematic. The shown error bars correspond to the statistical uncertainty of the available Monte Carlo sample only. Probabilities resulting from a Kolmogorov-Smirnov test and a χ^2 test for the agreement between variation and default are given for both variations. The lower plot shows the deviation of the systematics including the combined statistical errors as hashed areas.

C.2.6 POWHEG+PYTHIA generator



(a) Wt-channel production

(b) $t\bar{t}$ production

(c) W+heavy flavour jets production

(d) W+light flavour jets production

(e) t-channel and s-channel production

(f) Z+jets, WW, WZ, ZZ

Figure C.12: Distributions of the combined neural network output for different production channels normalized to unit area used as template in the fit. Each plot shows the default distribution compared to the upward and downward variations of the POWHEG+PYTHIA generator systematic. The shown error bars correspond to the statistical uncertainty of the available Monte Carlo sample only. Probabilities resulting from a Kolmogorov-Smirnov test and a χ^2 test for the agreement between variation and default are given for both variations. The lower plot shows the deviation of the systematics including the combined statistical errors as hashed areas.

APPENDIX D

Development of Physics Analysis Tools

The analysis presented in this thesis required the combination of a large variety of existing physics and statistics analysis tools and the development of new ones. In the course of the studies also data of other analysis groups has been combined on an event-by-event basis and included into the analysis chain. This appendix wants to share the experience and know-how acquired during the work on this and other projects. It is not intended to give a precise description of the technical solutions which were developed, but rather sketch these and discuss the arguments for the choice of certain paradigms. Doing so, the appendix is meant to serve as a discussion foundation and knowledge base for analysts who face such challenges or possibly the even more complex ones of the second LHC running period.

During the analysis presented in [6], which aimed at a measurement of the top-quark mass using the transverse momentum of the charged lepton in $t\bar{t}$ events, it became clear that software tools had to be developed to perform the last steps of physics analyses. The ATLAS software framework Athena [116] was developed to be used in all steps of a physics analysis. It offers the technical prerequisites to simulate, reconstruct and analyse data. By covering all these aspects of data handling of ATLAS data, it also has some downsides. The biggest problem is probably its size. The huge amount of data recorded by the LHC experiments requires data reduction methods to make analyses on specific physics channels possible. As an analyst anyway works on this topic the amount of data is often reduced with such large efficiency, that the final output dataset can be processed on single systems, like desktop computers or laptops. The code and binary size of the Athena framework and the usual overhead produced by loading and initialising all needed modules contradicts the flexibility achieved on the data side. Another reason is platform support. While Athena is built for Scientific Linux [149] systems, which is the default operating system used by CERN and all LHC collaborations, the standard analysis software in particle physics, the ROOT Data Analysis Framework [150], can be installed on many platforms. Summarising, the most commonly used analysis scheme by an ATLAS physicist is to reduce the size of the data as much as possible and finally to use the most flexible and close computing infrastructure, which is often a local institute computing cluster or even a single PC. A huge analysis framework does not fit into this scheme. The demand for such software was the main trigger for new analysis tools and drove the requirements.

This appendix first focusses on the technical parts of the developments, beginning with the requirements and sketching the chosen solutions. The second half centres more around the organisational part of the endeavour and how the software projects could be made available to a larger user group and what the consequences of such a solution are.

D.1 Requirements

The intention to create new software tools for high-energy particle physics analyses can have many root causes, which can often be generalised. The one raised above is dissatisfaction with available software. While a program to perform a certain analysis task might already exist, it does maybe not reach up to the user's expectations. Particle physics software, like the mentioned ROOT, is mostly open source software. This has the advantage of accessible code and the possibility to change it to the problem's need. But often the available software offers little to no documentation and only trivial examples that leave many questions unanswered. The potential user is obliged to read the source code to make use of the software, which is a tedious endeavour and potentially hard, given that physicists are not well-trained programmers in general. These are examples that cause frustration and lead to own developments to answer the posed question. Another reason can be the non-availability of software for the analysis task. Analyses at the LHC and high-energy particle physics often use very modern and sophisticated statistical methods, which might be implemented for the first time or only available in a way that makes it impossible to interface it to the available data formats. A more sociological reason is curiosity. Especially starters in the field want to learn now to implement a method. This often comes hand in hand with the demand for total control of the data at any runtime stage of a program, which is in fact a very important point as an analyst is fully responsible for his scientific results and, in case of the LHC, acts in the name of a larger community.

Once the motivation for the creation of the software is given, the root cause for the motivation continues to drive the creative process of developing software and defines the final package's strength. Due to the mentioned lack of education in fields that would be important for the success of the project, like software design and programming in general, the start of the software development is highly individual and based on the knowledge, experience and the scientific environment. The presented analysis led to the creation of two tools, which were used by a wider community. The creation of the main software had two base foundations, which probably also assured its success: an experienced working group and a good basis from another project called SFrame [151]. The former helped to define the requirements, while the latter was a design decision driven by the requirements.

D.1.1 Typical physics analysis

The requirements of the tool developed first are the ones of a typical particle physics analysis package, often named analysis framework although the term is probably not correctly chosen according to a computer scientist's definition. Such an analysis is organised into successively executed building blocks. They are very similar or identical for most analyses and therefore predestined to serve as a workflow for the software. At the beginning, the analysis tool has to read the data into variables. In contrast to many professional use-cases, the ATLAS data is not available in databases, but saved into files. These are organised in a table-like structure, where one event, which contains the information about the reconstructed objects from one collision in the detector, corresponds to a table row. The event is a very widely used concept to sub-divide the data, e.g. to process it in parallel without the need of sophisticated code. A multitude of structures exist in the ATLAS collaboration, which are based on the requirements of different analysis groups. This raised the desire to support different input-file contents.

The next step of the analysis is to complete the reconstruction of final state objects including the application of corrections. Not all of the selection criteria discussed in Chapter 4 are applied in the central processing of the data, but some are left to the analyser. Additionally, all the corrections have to be applied, which are described therein and often depend on the chosen selection in the analysis. These corrections include the application of weight factors, random smearing processes and recalibration of

measured quantities. The reweighting of distributions to the collision data luminosity also falls into this category. A variety of small classes, tools and functions of very diverging code quality are required to be used and have to be included into the analysis tool. After all the preprocessing, the analyser needs to apply a selection and potentially reconstruct intermediate particles from their decay products. These two steps are repeated very often and changed in an iterative optimisation procedure to find the best results adding the requirement of high repetition rate and flexible code adjustment paired with short compilation time. At the end of the typical analysis, results are calculated based on the reconstructed particles and their properties. Alternatively, the tool is meant to be used for a strong data reduction, writing out only the needed properties of finally selected events that can be used by other tools afterwards. All these steps have to be controlled by intermediate controls, which adds data visualisation and ordered formatted listing to the requirements.

Finally, the software needs to meet some technical processing requirements. It should run on as many platforms as possible, ranging from single computers to computing clusters and also include the execution on the LHC computing grid. The demand for an automatised splitting into flexible process parts is also useful in the context of sending jobs to larger facilities.

Not all of the above requirements were clearly defined before the start of the implementation but rather added once the core functionalities were finalised and tested successfully. This matches the observation of many other programming projects in particle physics. Interestingly, this can be developed into very successful methodology in agile software development. In the described environment it is driven by many factors: missing expertise for thorough requirement definition, time constraints, iterative tests of different analysis strategies to reach the desired goal but also technical constraints, like availability of computing resources.

D.1.2 Data visualisation

During the usage of the first tool for the base part of selection and reconstruction for the analysis, it became clear that the data visualisation part would better be expanded separately. The advantage of the approach is the shorter development and test times for the later parts of the presented analysis. With the evolution of an analysis the selection criteria for particles and events become more and more stable, reducing the necessity of rerunning this step. This allows to concentrate the advanced statistical techniques, like neural network trainings and the binned maximum likelihood fit presented here, on the reduced dataset. As the implementations of these techniques were separate codes from external sources and could not be integrated well into the generic analysis tool, a more modularised approach was started. This allowed a flexible adjustment of each analysis block and also the exchange of tools.

A new data visualisation tool was required to fit into this scheme. The core feature is the management and combination of distributions from different input source files. A plot like the one shown in the analysis is composed of many physics channels that are added to each other in different combinations. An example is the combination the distributions of ZZ, WZ and WW into one distribution only, the diboson distribution. A more complicated example from the analysis is the combination of the distributions of W+jets events with different number of partons into one and splitting it again into light-flavour and heavy-flavour production according to the heavy-flavour overlap removal tool's output. The main requirement is to read files, create distributions of predefined variables for each of the input files and offer the possibility of merging the initial distributions in different combinations. Additional requirements were the creation of templates for the fitting procedure, the overlay of normalized distributions and comparison plots. With the development of the neural network combination method, the support for two-dimensional distributions was also an important demand, which was added later.

A lesson learned from the first project was the consequent implementation of all features into a shared

library and the creation of example binaries which used these. This was also driven by the wish to exploit debugging tools in an easier way.

D.2 Solutions

The listed and motivated requirements were used in the creation of two tools named SFM and PlotProg. This section describes the basic design of the tools and discusses noteworthy ideas for solutions of some of the more elaborate problems. Both projects were programmed using the C++ programming language.

D.2.1 SFM analysis package

The SFM analysis package has been built on top of the SFrame data analysis framework [151], developed by members of the ATLAS collaboration and released as open-source software. It is programmed using the ROOT framework and therefore offers the further usage of ROOT also by the analyst. Its main features are the efficient reading of ROOT files, the run configuration using XML files, allowing to change parameters without code recompilation, and the support of the parallel processing tool PROOF, which is part of the ROOT framework. While only the two former features were exploited by SFM, as the PROOF execution contradicts with the scheme of one computing CPU per user job of most computing resources, SFrame also contains further useful features, like a uniform logging class with log-level support and the included reweighting of simulation events according to their theoretical cross-sections to the user-provided integrated luminosity of the collision data. With these basic prerequisites, the development of SFM could focus on more advanced concepts of the data analysis.

The first step was the creation of an object-oriented structure to hold all the particles and event properties. Reconstructed particles can be modelled using base classes that share the common properties of all particles and derived classes for every particle or object type. This includes for example classes for electrons, muons, taus, jets, tracks and more. An class was created to serve as a central storage point for all objects possibly contained in an event. It uses the vector class of the C++ standard template library (STL) to store objects of the same type and provides additional functionality for object retrieval, storage of dataset and event information and modification of objects and parameters. The idea for such an event class was mostly derived from the Z++ framework [152], which was created by former colleagues in the same working group who analysed data from the ZEUS detector at the HERA collider. While no code was transferred from the project, discussions with its developers and basic concepts were also incorporated into SFM and developed further. The idea of an event class has several benefits:

- The class defines a uniform interface for functional modules which can work with the objects. The stable interface allows the modules to be specialised, evading the need of template programming.
- It helps starting users to write analysis code in a very understandable and natural way in the sense that variables, functions and algorithms are named according to the mathematical description and can be followed by non-programmers.
- The former point also helps in the transition of coding between analysers. Existing code can be read and understood fast and easily be expanded and altered.
- As emphasised before, the event class presents a central point, where all needed data can be stored throughout the analysis code.

The first of these points was heavily exploited in the further development process. The most important modules and concepts which are included are described in the following.

Data input As outlined before, different input file structures have to be supported. The event class serves as a common basis in this respect. A factory-like structure was programmed, which calls the reader class for a specific file format, based on a variable setting in the configuration file. This allows flexible expansion and removal of support for input file formats. The reader classes have to be created by the user of a format and include the adaptation to the common event class basis. The reader class fills the event object with all needed information using the SFrame mechanism to read the file content into variables. It can also apply corrections or data conversions. All reader versions are collected in a central place, which facilitates reusability.

Generic functions A collection for generic functions acting on objects of any type were created. This included functions for the calculation of distances and angles between objects, a small set of statistics functions, functions to calculate special variables and functions for jet analysis and *b*-tagging. All these functions were programmed using template variables. The reason was to allow these function to be easily portable to other software projects. While this is of course a useful feature, it was never really used, as discussed in Section D.3. Templates add a level of complexity to the programming process, which was unnecessary in this case when looking back.

Data visualisation To match the requirement for data visualisation at any stage of the program, a plotting library was added consisting of two main elements. The internal part allows to plot many variables for all basic objects in the event, like electrons and jets. Classes for common groups of histograms are available, which can be configured by the user at object creation time. These classes create the ROOT histogram objects needed for all distributions contained in the group. Additionally, functions to fill the histogram objects with the corresponding variables are available, which are called by the user at the desired point in the analysis code control flow. Creation of multiple objects of the same type is also possible, allowing to compare distributions e.g. at different selection steps of the analysis. All histograms are written to an output file for further processing. A second library takes over the responsibility of merging certain production channels and creating stacked representations, overlaying normalized distributions or creating other visualisations. This library is not executed during SFM but meant to be used in small standalone scripts that are executed after the full processing of all simulated events and possibly also the collision data. While the module as a whole is useful in small analyses it turned out to be inflexible due to the need to reprocess the full analysis for every small change or additional investigation and was not used by the analysis at later stages but replaced with an independent library.

Event selection Applying criteria on different variables which must be fulfilled by the event to be used in analysis is a step present in all analysis. It serves the purpose of removing or reducing the impact of background events thus enhancing the signal contribution. The module consists of a collection of selection criteria defined using the event and objects classes. Examples are number of jets with more than a defined magnitude of the transverse momentum or a certain magnitude of the missing transverse momentum. These cuts can then be used by the user in the definition of a cut flow, i.e. a collection of cuts executed in a defined order. The module stores the boolean information if an event fulfils the selection criteria or not. This information can then be used by conditional statements in the analysis code. The expected number of events should match the observed number of events at each step of a cut flow. To examine this agreement, the event selection module is combined with the data visualisation one. It offers functionality to automatically draw distributions of the cutflows that are saved to the output files and stacked by the external visualisation library. These representations can also be used for comparisons, if all definitions of the cuts are implemented correctly.

Validation A very important step in the programming process is the validation of code. The fact that many analysis tools are available in each larger analysis group which has members from different institutions can be beneficial in this case. The only way to really ensure that common corrections are applied correctly and that no property is numerically changed during the execution of the program is to cross-check the components of four-vectors for a variety of reconstructed particles. Classes are available to write out these numerical values in an ordered tabular format into text files. The writing command can be called at different steps of the program to write out a certain status. Such tables can then be compared using standard tools to identify deviations which serve as a starting point to code adjustments. Additionally, many classes of reconstructed objects and the event class incorporate print functions that write out numerical values to the logging facility of SFrame in a formatted way. This is also a lesson learned from Z++ and has proven very valuable during the initial creation of above mentioned classes. Again cross-checks with several other analyses were carried out and revealed mistakes on both sides: in the SFM code and also in the comparison target analysis code.

Several other modules have been created. They are for example responsible for writing the contents of the event class into a ROOT data file again for further processing, which is very useful to provide preprocessed data to starting analysers or undergraduate students. Another important module for the presented analysis allows the creation of reconstructed particles. It has the benefit of storing the objects which were combined in the reconstruction process. This can be exploited later in the analysis if automatic procedure were used in the reconstruction. All these modules use the initially created event class and classes for physics objects, which shows the power of this approach. While the fixed interfaces of these classes show a large potential for further developments, it is the main drawback at the same time. As analyses evolve and more information must often be included, the interface might need to be changed, which is often a cumbersome process when it needs to be propagated into all modules.

D.2.2 PlotProg library

Valuable expertise was gained from the development of SFM and influenced the development of PlotProg, the second larger program created for the analysis. Due to the purpose of data visualisation as main requirement, the classes for data were completely dropped together with the SFrame base. For this project it was not necessary to process a huge number of input files but rather only one file per simulation production channel. Allowing to start from scratch, the first decision was to employ a proper build system. The open-source build system CMake [153] was chosen as it is much easier to use than the plain Makefiles from SFM and allows to easily include other libraries, like ROOT, using publicly available macros. The configuration files are also much easier to understand for beginners.

After defining the build tools, the basic design of the library was fixed including the strategy to fulfil the requirements defined above. The central concepts are the definition of variables to plot, the management of histograms and the merging process. The variable definition in this sense includes the name of a variable as it is defined in the input file followed by histogram properties like axis labels and binning. This definition is added to a central tool: the ChannelPlotter. This component creates a ROOT formula object in C++ syntax that is used for the filling of a histogram. Apart from the variable name, this formula object contains input file parameters and can also include a base selection, which the user can define using functions of the ChannelPlotter. Additionally, the ChannelPlotter holds a container for the histograms that are created for each input file that corresponds to a simulation production channel or to collision data. The ChannelPlotter accepts weight factors for each defined channel, which is useful to apply scale factors that result from a fit. It also stores the desired input location of the files. The file names are built from a list of channels to be read, as each channel corresponds to one input file.

Once the configuration is finalised, the ChannelPlotter creates histograms for each defined channel and containers for each variable to be plotted. It reads the input files and fills the histograms, supporting one-dimensional and two-dimensional distributions. The user can then obtain the containers of histograms and start a cascaded merging procedure. A class ChannelMerger was designed to merge the histograms created for all variables based on the user's merging definitions that consists of the source channels and the target channels. This merging can for instance combine electron and muon channels and different simulation samples. The creation and deletion of the containers storing different merging stages is controlled by the user for flexibility reasons.

To accomplish the actual creation of combined data visualisations like stack plots of simulation samples or normalized distribution overlays, the classes of the SFM histogram module have been reused and a new interface has been written to connect them. This is a good example how well-designed classes should be reused in new projects to reduce the development time drastically while assuring code quality. Only the interface had to be coded and tested, while the migrated classes remained unchanged. At a later stage, the class which stores the histograms for one variable has been extended to meet the additional requirements. Functionality was added to store the single histograms of all contained channels in a directory of an output file, as single files in different graphics file formats and even applying the very special naming convention that was needed for the fit templates in the analysis. Such naming schemes can be added easily to one class and are then available library-wide.

The library has proven its power and flexibility in the presented analysis. During the evolution of the analysis it could be extended without major reorganisation of the code basis despite the non-perfect designed interdependencies of classes.

D.3 Collaborative code development in inhomogeneous environments

The development of the two tools and other code was always started as a project for the presented analysis. The foundations were laid and the second project was created entirely for the analysis only. One of the prerequisites of experiments and analyses in particle physics at the highest reachable energies is collaboration. This is a result of the complexity of the experiments and of the event contents. Many analysis projects at the large detectors would not be possible to be finished by one PhD student in the time-frame of a thesis project. Precision measurements and the isolation of channels with large irreducible backgrounds, as seen in this analysis, add so many challenges that teamwork becomes more and more essential. Fortunately, this development is also reflected in the organisation of working groups at the institute level which allow sharing of tasks which are of common interest. Given the observations pointed out in the beginning of this appendix, it was already clear in the development phase of SFM that the program would be used by several people. By advertising the existence of a development project, the demand rose quickly, especially by people who planned to start an analysis project.

In this section, the sociological side of the software development will be discussed and explained how the project could be enlarged to a larger number of users, which are developers at the same time. The scaling of such a project opens up opportunities but also creates new challenges. The largest challenge is the different level of expertise of the involved people. A working group is a very inhomogeneous environment, composed of people from different countries with different cultures and also with different programming knowledge. These facts have to be realised and taken into account.

D.3.1 Project scaling and demand management

The object-oriented code base of the SFM project was created from the beginning with the intention to be reused by other analysts or in similar projects. The initial development targeted the top-quark mass measurement in [6] and aimed at including the search for the single top-quark production in the Wt-channel at a later stage. This analysis already included a second analyst, who worked on the same project as a diploma thesis project. It was therefore concluded that sharing the code was the best way to combine the effort and reduce the amount of work. The decision was to use a version control system (VCS) with a central repository to remove the risk of code loss while working on it. The Bazaar VCS [154] was chosen as one of the most modern systems at that time. It has the advantage that no large infrastructure is needed to provide a central repository, it is available for all major platforms and easy to use but at the same time very powerful. Version control systems are nowadays widespread in the development of code in high-energy physics because the huge advantages were realised. Even local usage for small code development projects is worthwhile as modern systems support local repositories and branching even on this small scale. An analyst can try out different solutions for a problem at the same time and merge parts of the code between them. The PlotProg program was also started using Bazaar but later migrated to Git [155], which is the current most used version control system. Both support similar workflows and can support small projects up to very large ones with hundreds of developers. The transition between the tools is possible with preservation of the full commit history.

Already at this early stage of the project, the less experienced team member could contribute significantly, reducing the workload and freeing more time for the development of the core classes. It also allows starters to see advancements in the code. Reading code is the best way to learn new programming concepts and solutions. It has immediate effects on the coding style of less experienced analysts and also makes the code more uniform. As the beginning experience was very positive, the project was generally opened to the working group and other working groups at the same institute. This step is a very drastic one. A development for a small team hardly needs any management, the requirements are clearly defined, communication is very direct, decision can be taken fast and agreement is reached easily. This is very different for a larger group of developers and users with very different expectations.

A first aspect is the documentation of the project. This point is ignored in most projects as it has several negative effects: it needs time that is preferably used for further development, it is often considered unnecessary as users and developers are the same people and it is often ignored as many larger developments start from very small programs that are initially not meant to grow. The positive effects of a good documentation are often not acknowledged in the field. When an analyst starts to use a program and contribute to the code, it is very important that the first experience is positive. In this case, the analyst accepts the program and is inclined to contribute to it if asked for. A good documentation can support this positive first experience and create confidence. The analyst can manage the code with the provided documentation even in the case that the main developers are not there to provide additional support. This enhances the first binding to the project. A wiki page was created for SFM [156] that was nicely received by starting users and also expanded by contributors with additional information of general character or about specialisations. Together with the in-line comments in the code, this provided an acceptable basis. The time gain by answering questions with a short hint to the documentation clearly outweighed the effort to create it. Of course insufficiencies in the documentation had to compensated with personal support.

The personal support is very important. Giving audience to a user with a short response time can clarify problems when they arise. This has positive consequences like the prevention of conflicts, a higher identification with the software and often the growing wish to return to the project or to the helping individuals. For the skilled developers that are able to help, it offers the opportunity to get an impression of

common problems of less experienced contributors and also to get to know the expectations and requirements. On a longer time scale it can also help to identify strengths and weaknesses of team members opening up possibilities for project management. In general, these relations are not seen and therefore not exploited enough. They need to be taken into account when the decision is pending if a project should be scaled up and opened or not. Triggering positive effects on a project from a larger scale of contributors requires a lot of work in the beginning and also constant attention in the development of a project. The expectations and requirements of all parties have to be matched as well as possible, which demands considerable time investment and a very flexible code basis.

D.3.2 Limitations of shared development

The scaling of SFM and especially PlotProg to a larger user basis was initially successful but not sustainable. The reasons for these developments are different for both projects. SFM became a very successful project with more than ten users and about five active developers that contributed important modules to SFM. Examples are the data visualisation model, a build system that supported remote compilation and job splitting, submission and merging on the Bonn Analysis Facility, many improvements of already previously existing modules and specialised classes or modules for concrete analysis tasks. Several diploma [157–159], master [139, 160, 161] and PhD theses projects [138, 162, 163] were successfully completed using the SFM software and various papers from the ATLAS collaboration were published [164–167]. The most obvious reason for the little usage at the present time is a change of generation. Most SFM contributors have finalised the projects and the next generation of analysts seem to have found other solutions or are not yet at a stage where the demand for such a tool is imminent. This goes in line with the fact that there were no contributors from the higher scientific ranks who would have invested on a longer time-scale.

For the second run of the LHC, the ATLAS Collaboration also developed new analysis software, which fulfils many of the requirements stated initially. This is a very desired development as for instance the application of corrections is a common task for all analyses with a similar selection and should be shared in the framework of the collaboration. The sort of code sharing achieved in the SFM framework across different physics topics is still not available though, as it heavily relied on direct personal communication and the consequently fast decisions and flexible adjustments.

The PlotProg development project is a nice confirmation of the argumentation for the success of SFM. Due to time constraints of finalising the analysis it was not possible to invest the effort in documentation and also in personal support. While it served its purpose in this analysis and two master thesis projects in the field of the search for the Wt-channel of single top-quark production, it could never reach a larger number of contributors that would be interested in a long-term investment. It demonstrates that the evolution of such a project and the possible exploitation of synergy effects from teamwork in the field of high-energy particle physics heavily depends on the strong commitment towards a project scaling in the early phase by a few individuals. Of similar importance is a large demand for the provided software solution.

In conclusion the sector should learn from non-scientific projects in terms of project management and software development methodology. The teaching of programming knowledge and the training of usage of standard tools for version control, debugging, profiling and the consequent usage of integrated development environments would be very beneficial for the evolution of such projects and free more time to develop analysis strategies and focus on the fundamental physics in all steps of the analysis.

Bibliography

[1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B 716 (Sept. 2012) 1, DOI: 10.1016/j.physletb.2012.08.020.

[2] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B 716 (Sept. 2012) 30, DOI: 10.1016/j.physletb.2012.08.021.

[3] I. Brock (ed.) and T. Schoerner-Sadenius (ed.), *Physics at the Terascale*, Wiley-VCH, 2011.

[4] D. Griffiths, *Introduction to Elementary Particles*, Wiley-VCH, 2008.

[5] F. Halzen and A. D. Martin, *Quarks and Leptons: An Introductory Course in Modern Particle Physics*, John Wiley & Sons, 1984.

[6] J. A. Stillings, *Top Quark Mass Measurement using Lepton Transverse Momenta at ATLAS*, BONN-IB-2009-06, Diplomarbeit: Rheinische Friedrich-Wilhelms-Universität Bonn, July 2009, URL: http://brock.physik.uni-bonn.de/files/pubs/bonn-ib-2009-06.pdf.

[7] J. Thomson, *Cathode Rays*, Phil. Mag. 44 (1897) 293.

[8] C. D. Anderson and S. H. Neddermeyer, *Cloud Chamber Observations of Cosmic Rays at 4300 Meters Elevation and Near Sea-Level*, Phys. Rev. 50 (Aug. 1936) 263, DOI: 10.1103/PhysRev.50.263.

[9] M. L. Perl et al., *Evidence for Anomalous Lepton Production in $e^+ - e^-$ Annihilation*, Phys. Rev. Lett. 35 (Dec. 1975) 1489, DOI: 10.1103/PhysRevLett.35.1489.

[10] M. L. Perl et al., *Properties of anomalous $e\mu$ events produced in e^+e^- annihilation*, Phys. Lett. B 63 (Aug. 1976) 466, DOI: 10.1016/0370-2693(76)90399-3.

[11] M. L. Perl et al., *Properties of the proposed τ charged lepton*, Phys. Lett. B 70 (Oct. 1977) 487, DOI: 10.1016/0370-2693(77)90421-X.

[12] L. M. Lederman, M. Schwartz, J. Steinberger et al., *Observation of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos*, Phys. Rev. Lett. 9 (July 1962) 36, DOI: 10.1103/PhysRevLett.9.36.

[13] DONUT Collaboration, *Observation of tau neutrino interactions*, Phys. Lett. B 504 (Apr. 2001) 218, DOI: 10.1016/S0370-2693(01)00307-0.

[14] C. L. Cowan, F. Reines et al., *Detection of the free neutrino: A Confirmation*, Science 124.3212 (July 1956) 103, DOI: 10.1126/science.124.3212.103.

[15] Q. Ahmad et al., *Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory*, Phys. Rev. Lett. 89 (June 2002) 011301, DOI: 10.1103/PhysRevLett.89.011301, arXiv: nucl-ex/0204008 [nucl-ex].

[16] K2K Collaboration, *Measurement of Neutrino Oscillation by the K2K Experiment*, Phys. Rev. D 74 (2006) 072003, DOI: 10.1103/PhysRevD.74.072003, arXiv: hep-ex/0606032.

[17] G. Zweig, *An SU_3 model for strong interaction symmetry and its breaking; Version II*, CERN-TH-412 (Feb. 1964), URL: http://cds.cern.ch/record/570209.

[18] G. Zweig, *An SU_3 model for strong interaction symmetry and its breaking; Version I*, CERN-TH-401 (Jan. 1964), URL: http://cds.cern.ch/record/352337.

[19] M. Gell-Mann, *A schematic model of baryons and mesons*, Phys. Lett. 8 (Feb. 1964) 214, DOI: 10.1016/S0031-9163(64)92001-3.

[20] R. E. Taylor, *Deep inelastic scattering: The early years*, Rev. Mod. Phys. 63 (July 1991) 573, DOI: 10.1103/RevModPhys.63.573.

[21] H. W. Kendall, *Deep inelastic scattering: Experiments on the proton and the observation of scaling*, Rev. Mod. Phys. 63 (July 1991) 597, DOI: 10.1103/RevModPhys.63.597.

[22] J. I. Friedman, *Deep inelastic scattering: Comparisons with the quark model*, Rev. Mod. Phys. 63 (July 1991) 615, DOI: 10.1103/RevModPhys.63.615.

[23] J.-E. Augustin et al., *Discovery of a Narrow Resonance in e^+e^- Annihilation*, Phys. Rev. Lett. 33 (Dec. 1974) 1406, DOI: 10.1103/PhysRevLett.33.1406.

[24] G. S. Abrams et al., *Discovery of a Second Narrow Resonance in e^+e^- Annihilation*, Phys. Rev. Lett. 33 (Dec. 1974) 1453, DOI: 10.1103/PhysRevLett.33.1453.

[25] S. W. Herb et al., *Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions*, Phys. Rev. Lett. 39 (Aug. 1977) 252, DOI: 10.1103/PhysRevLett.39.252.

[26] CDF Collaboration, *Observation of Top Quark Production in $\bar{p}p$ Collisions with the Collider Detector at Fermilab*, Phys. Rev. Lett. 74 (Apr. 1995) 2626, DOI: 10.1103/PhysRevLett.74.2626.

[27] DØ Collaboration, *Search for High Mass Top Quark Production in $p\bar{p}$ Collisions at $\sqrt{s} = 1.8$ TeV*, Phys. Rev. Lett. 74 (Mar. 1995) 2422, DOI: 10.1103/PhysRevLett.74.2422.

[28] Particle Data Group, *The Review of Particle Physics*, Chin. Phys. C 38 (2014) 090001, URL: http://pdg.lbl.gov/.

[29] M. Battaglia et al., *The CKM Matrix and the Unitarity Triangle*, Proceedings of the First Workshop on the CKM Unitarity Triangle, ed. by M. Battaglia et al., CERN, (Apr. 2003), arXiv: hep-ph/0304132.

[30] R. P. Feynman, *Very High-Energy Collisions of Hadrons*, Phys. Rev. Lett. 23 (Dec. 1969) 1415, DOI: 10.1103/PhysRevLett.23.1415.

[31] J. C. Collins, D. E. Soper and G. F. Sterman, *Factorization of Hard Processes in QCD*, Adv. Ser. Direct. High Energy Phys. 5 (1988) 1, arXiv: hep-ph/0409313 [hep-ph].

[32] *LHAPDF PDF sets*,
URL: http://lhapdf.hepforge.org/pdfsets.html (visited on 09/10/2014).

[33] G. Altarelli and G. Parisi, *Asymptotic freedom in parton language*,
Nucl. Phys. B 126 (Aug. 1977) 298, DOI: 10.1016/0550-3213(77)90384-4.

[34] H1 and ZEUS Collaborations,
QCD Analysis of the Inclusive $e^{\pm}p$ Scattering Cross Sections at HERA,
H1prelim-10-141, ZEUS-prel-10-017 (Apr. 2014), URL: https://www.desy.de/h1zeus/combined_results/index.php?do=proton_structure.

[35] M. Kobayashi and T. Maskawa,
CP-Violation in the Renormalizable Theory of Weak Interaction,
Prog. Theor. Phys. 49.2 (Feb. 1973) 652, DOI: 10.1143/PTP.49.652.

[36] ATLAS Collaboration, *Measurement of the top quark-pair production cross section with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV*, Eur. Phys. J. C 71 (Mar. 2011) 1577,
DOI: 10.1140/epjc/s10052-011-1577-6.

[37] CMS Collaboration, *First Measurement of the Cross Section for Top-Quark Pair Production in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV*, Phys. Lett. B 695 (Jan. 2011) 424,
DOI: 10.1016/j.physletb.2010.11.058, arXiv: 1010.5994 [hep-ex].

[38] ATLAS Collaboration, CDF Collaboration, CMS Collaboration and DØ Collaboration,
First combination of Tevatron and LHC measurements of the top-quark mass,
ATLAS-CONF-2014-008, CDF-NOTE-11071, CMS-PAS-TOP-13-014, D0-NOTE-6416 (2014), arXiv: 1403.4427 [hep-ex].

[39] ALEPH, DELPHI, L3, OPAL and SLD Collaborations,
Precision electroweak measurements on the Z resonance, Phys. Rept. 427 (May 2006) 257,
DOI: 10.1016/j.physrep.2005.12.006, arXiv: hep-ex/0509008 [hep-ex].

[40] CDF Collaboration, *Direct Measurement of the Total Decay Width of the Top Quark*,
Phys. Rev. Lett. 111 (Nov. 2013) 202001, DOI: 10.1103/PhysRevLett.111.202001.

[41] ATLAS Collaboration, *Measurements of spin correlation in top-antitop quark events from proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*,
Phys. Rev. D. 90 (Dec. 2014) 112016, DOI: 10.1103/PhysRevD.90.112016.

[42] CMS Collaboration, *Measurements of $t\bar{t}$ Spin Correlations and Top-Quark Polarization Using Dilepton Final States in pp Collisions at $\sqrt{s} = 7$ TeV*,
Phys. Rev. Lett. 112 (May 2014) 182001, DOI: 10.1103/PhysRevLett.112.182001.

[43] ATLAS Collaboration,
Measurement of the top quark charge in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector,
JHEP 11 (Nov. 2013) 031, DOI: 10.1007/JHEP11(2013)031.

[44] ATLAS Collaboration, *Measurement of the $t\bar{t}$ production cross-section using $e\mu$ events with b-tagged jets in pp collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS detector*,
Eur. Phys. J. C 74 (Oct. 2014) 3109, DOI: 10.1140/epjc/s10052-014-3109-7.

[45] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, Phys. Rev. D 83 (May 2011) 091503(R),
DOI: 10.1103/PhysRevD.83.091503.

[46] CDF Collaboration, *Observation of Electroweak Single Top-Quark Production*,
Phys. Rev. Lett. 103 (Aug. 2009) 092002, DOI: 10.1103/PhysRevLett.103.092002.

[47] DØ Collaboration, *Observation of Single Top Quark Production*, Phys. Rev. Lett. 103 (Aug. 2009) 092001, DOI: 10.1103/PhysRevLett.103.092001.

[48] N. Kidonakis, *Next-to-next-to-leading logarithm resummation for s-channel single top quark production*, Phys. Rev. D 81 (Mar. 2010) 054028, DOI: 10.1103/PhysRevD.81.054028.

[49] CDF and DØ Collaborations, *Observation of s-Channel Production of Single Top Quarks at the Tevatron*, Phys. Rev. Lett. 112 (June 2014) 231803, DOI: 10.1103/PhysRevLett.112.231803.

[50] ATLAS Collaboration, *Comprehensive measurements of t-channel single top-quark production cross sections at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Phys. Rev. D 90 (Dec. 2014) 112006.

[51] CMS Collaboration, *Measurement of the single-top-quark t-channel cross section in pp collisions at $\sqrt{s} = 7$ TeV*, JHEP 12 (Dec. 2012) 035, DOI: 10.1007/JHEP12(2012)035.

[52] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-*, Phys. Rev. D 82 (Sept. 2010) 054018, DOI: 10.1103/PhysRevD.82.054018.

[53] T. M. Tait, *The tW^- mode of single top quark production*, Phys. Rev. D 61 (Dec. 1999) 034001, DOI: 10.1103/PhysRevD.61.034001, arXiv: hep-ph/9909352 [hep-ph].

[54] A. Belyaev and E. Boos, *Single top quark tW + X production at the CERN LHC: A Closer look*, Phys. Rev. D 63 (Jan. 2001) 034012, DOI: 10.1103/PhysRevD.63.034012, arXiv: hep-ph/0003260 [hep-ph].

[55] J. M. Campbell, R. K. Ellis and F. Tramontano, *Single top-quark production and decay at next-to-leading order*, Phys. Rev. D 70 (Nov. 2004) 094012, DOI: 10.1103/PhysRevD.70.094012, arXiv: hep-ph/0408158 [hep-ph].

[56] J. M. Campbell and F. Tramontano, *Next-to-leading order corrections to Wt production and decay*, Nucl. Phys. B 726 (Oct. 2005) 109, DOI: 10.1016/j.nuclphysb.2005.08.015, arXiv: hep-ph/0506289 [hep-ph].

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

[58] C. D. White et al., *Isolating Wt production at the LHC*, JHEP 11 (Nov. 2009) 074, DOI: 10.1088/1126-6708/2009/11/074, arXiv: 0908.0631 [hep-ph].

[59] ATLAS Collaboration, *Search for W+t single-top events in the dileptonic channel at ATLAS*, ATLAS-CONF-2011-104 (July 2011), URL: http://cds.cern.ch/record/1369829.

[60] CMS Collaboration, *Search for single top tW associated production in the dilepton decay channel in pp collisions at $\sqrt{s} = 7$ TeV* (Sept. 2011), URL: http://cds.cern.ch/record/1385552.

[61] ATLAS Collaboration, *Evidence for the associated production of a W boson and a top quark in ATLAS at $\sqrt{s} = 7$ TeV*, Phys. Lett. B 716 (Sept. 2012) 142, DOI: 10.1016/j.physletb.2012.08.011.

[62] J.A. Stillings for the ATLAS Collaboration,
Evidence for the associated production of a W boson and a top quark in ATLAS,
EPJ Web Conf. 60 (Nov. 2013) 20005, DOI: 10.1051/epjconf/20136020005.

[63] CMS Collaboration, *Evidence for Associated Production of a Single Top Quark and W Boson in pp Collisions at* $\sqrt{s} = 7$ TeV, Phys. Rev. Lett. 110 (Jan. 2013) 022003,
DOI: 10.1103/PhysRevLett.110.022003.

[64] ATLAS Collaboration, *Measurement of the cross-section for associated production of a top quark and a W boson at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*,
ATLAS-CONF-2013-100 (Sept. 2013), URL: http://cds.cern.ch/record/1600799.

[65] CMS Collaboration, *Observation of the Associated Production of a Single Top Quark and a W Boson in pp Collisions at* $\sqrt{s} = 8$ *TeV*, Phys. Rev. Lett. 112 (June 2014) 231802,
DOI: 10.1103/PhysRevLett.112.231802.

[66] ATLAS and CMS Collaborations,
Combination of cross-section measurements for associated production of a single top-quark and a W boson at $\sqrt{s} = 8$ *TeV with the ATLAS and CMS experiments*,
ATLAS-CONF-2014-052, CMS-PAS-TOP-14-009 (Sept. 2014),
URL: http://cds.cern.ch/record/1951032.

[67] ATLAS Collaboration, *Summary plots from the ATLAS Top physics group web page*, URL:
https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/TOP/
(visited on 19/11/2014).

[68] W. Stirling, *private communication*, 2014.

[69] LEP Collaboration, ed., *LEP Design Report, Volume 2: The LEP main ring*, Geneva, 1984,
URL: http://cds.cern.ch/record/102083.

[70] C. Wyss, ed., *LEP Design Report, Volume 3: LEP2*, Geneva, 1996,
URL: http://cds.cern.ch/record/314187.

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

[72] CMS Collaboration, *The CMS experiment at the CERN LHC*, JINST 3 (Aug. 2008) S08004,
DOI: 10.1088/1748-0221/3/08/S08004.

[73] LHCb Collaboration, *The LHCb Detector at the LHC*, JINST 3 (Aug. 2008) S08005,
DOI: 10.1088/1748-0221/3/08/S08005.

[74] ALICE Collaboration, *The ALICE experiment at the CERN LHC*,
JINST 3 (Aug. 2008) S08002, DOI: 10.1088/1748-0221/3/08/S08002.

[75] TOTEM Collaboration, *The TOTEM experiment at the CERN Large Hadron Collider*,
JINST 3 (Aug. 2008) S08007, DOI: 10.1088/1748-0221/3/08/S08007.

[76] LHCf Collaboration, *The LHCf detector at the CERN Large Hadron Collider*,
JINST 3 (Aug. 2008) S08006, DOI: 10.1088/1748-0221/3/08/S08006.

[77] MoEDAL Collaboration, *Technical Design Report of the MoEDAL Experiment*,
CERN-LHCC-2009-006, MoEDAL-TDR-001, Geneva, June 2009,
URL: http://cds.cern.ch/record/1181486.

[78] O. S. Brüning et al., *LHC Design Report, Vol. 1, The LHC Main Ring*,
CERN-2004-003-V-1 (2004), URL: http://cds.cern.ch/record/782076.

[79] O. S. Brüning et al., *LHC Design Report, Vol. 2, The LHC Infrastructure and General Services*, CERN-2004-003-V-2 (2004), URL: http://cds.cern.ch/record/815187.

[80] M. Benedikt et al., *LHC Design Report, Vol. 3, The LHC Injector Chain*, CERN-2004-003-V-3 (2004), URL: http://cds.cern.ch/record/823808.

[81] CERN Communication Group, *LHC the guide*, Geneva: CERN, Feb. 2009, URL: http://cds.cern.ch/record/1165534/files/CERN-Brochure-2009-003-Eng.pdf.

[82] ATLAS Collaboration, *ATLAS Photos*, URL: http://atlas.ch/photos/index.html (visited on 12/01/2015).

[83] ATLAS Collaboration, *Expected Performance of the ATLAS Experiment – Detector, Trigger and Physics* (Jan. 2009), arXiv: 0901.0512.

[84] ATLAS Collaboration, *Performance of the ATLAS Electron and Photon Trigger in p-p Collisions at $\sqrt{s}$ = 7 TeV in 2011*, ATLAS-CONF-2012-048 (May 2012), URL: http://cds.cern.ch/record/1450089.

[85] ATLAS Collaboration, *ATLAS Trigger Operation Public Results*, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/TriggerOperationPublicResults (visited on 12/01/2015).

[86] T. Cornelissen et al., *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*, ATL-SOFT-PUB-2007-007 (Mar. 2007), URL: https://cds.cern.ch/record/1020106.

[87] ATLAS Collaboration, *Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in the High Pile-Up LHC Environment*, ATLAS-CONF-2012-042 (Mar. 2012), URL: http://cds.cern.ch/record/1435196.

[88] P. Billoir, *Track fitting with multiple scattering: A new method*, Nucl. Instr. and Meth. 225 (Aug. 1984) 352, DOI: 10.1016/0167-5087(84)90274-6.

[89] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instr. and Meth. 262 (Dec. 1987) 444, DOI: 10.1016/0168-9002(87)90887-4.

[90] ATLAS Collaboration, *Performance of primary vertex reconstruction in proton-proton collisions at $\sqrt{s}$ =7 TeV in the ATLAS experiment*, ATLAS-CONF-2010-069 (July 2010), URL: http://cds.cern.ch/record/1281344.

[91] S. Gibson et al., *The Multi-channel High Precision ATLAS SCT Alignment Monitoring System: A Progress Report*, eConf, 9th International Workshop on Accelerator Alignment, C06092511, (2006) WEPO08, URL: http://www.slac.stanford.edu/econf/C06092511/papers/WEPO08.PDF.

[92] ATLAS Collaboration, *Study of alignment-related systematic effects on the ATLAS Inner Detector tracking*, ATLAS-CONF-2012-141 (Oct. 2012), URL: http://cds.cern.ch/record/1483518.

[93] ATLAS Collaboration, *Alignment of the ATLAS Inner Detector Tracking System with 2010 LHC proton-proton collisions at $\sqrt{s}$ = 7 TeV*, ATLAS-CONF-2011-012 (Mar. 2011), URL: http://cds.cern.ch/record/1334582.

[94] ATLAS Collaboration, *Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data*, Eur. Phys. J. C 72 (Mar. 2012) 1, DOI: 10.1140/epjc/s10052-012-1909-1.

[95] T. Lagouri et al., *A muon identification and combined reconstruction procedure for the ATLAS detector at the LHC at CERN*, IEEE Transactions on Nuclear Science 51 (Dec. 2004) 3030, DOI: 10.1109/TNS.2004.839102.

[96] ATLAS Collaboration, *Identification of muon candidates in pp collisions at $\sqrt{s}$ = 900 GeV with the ATLAS detector*, ATLAS-CONF-2010-015 (July 2010), URL: http://cds.cern.ch/record/1276549.

[97] ATLAS Collaboration, *Muon reconstruction efficiency in reprocessed 2010 LHC proton-proton collision data recorded with the ATLAS detector*, ATLAS-CONF-2011-063 (Apr. 2011), URL: http://cdsweb.cern.ch/record/1345743.

[98] ATLAS Collaboration, *Muon Momentum Resolution in First Pass Reconstruction of pp Collision Data Recorded by ATLAS in 2010*, ATLAS-CONF-2011-046 (Mar. 2011), URL: http://cds.cern.ch/record/1338575.

[99] ATLAS Collaboration, *Jet energy scale and its systematic uncertainty in proton-proton collisions at sqrt(s)=7 TeV with ATLAS 2011 data*, ATLAS-CONF-2013-004 (Jan. 2013), URL: https://cds.cern.ch/record/1509552.

[100] ATLAS Collaboration, *Properties of Jets and Inputs to Jet Reconstruction and Calibration with the ATLAS Detector Using Proton-Proton Collisions at $\sqrt{s}$ = 7 TeV*, ATLAS-CONF-2010-053 (July 2010), URL: http://cds.cern.ch/record/1281310.

[101] M. Cacciari, G. P. Salam and G. Soyez, *The Anti-k_t jet clustering algorithm*, JHEP 04 (Apr. 2008) 063, DOI: 10.1088/1126-6708/2008/04/063, arXiv: 0802.1189 [hep-ph].

[102] ATLAS Collaboration, *Pile-up corrections for jets from proton-proton collisions at $\sqrt{s}$ = 7 TeV in ATLAS in 2011*, ATLAS-CONF-2012-064 (July 2012), URL: https://cds.cern.ch/record/1459529.

[103] ATLAS Collaboration, *Selection of jets produced in proton-proton collisions with the ATLAS detector using 2011 data*, ATLAS-CONF-2012-020 (Mar. 2012), URL: https://cds.cern.ch/record/1430034.

[104] ATLAS Collaboration, *Jet energy resolution in proton-proton collisions at $\sqrt{s}$ = 7 TeV recorded in 2010 with the ATLAS detector*, Eur. Phys. J. C 73 (Mar. 2013) 2306, DOI: 10.1140/epjc/s10052-013-2306-0.

[105] ATLAS Collaboration, *Measurement of the b-tag Efficiency in a Sample of Jets Containing Muons with 5 fb^{-1} of Data from the ATLAS Detector*, ATLAS-CONF-2012-043 (Mar. 2012), URL: http://cds.cern.ch/record/1435197.

[106] ATLAS Collaboration, *Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data*, ATLAS-CONF-2011-102 (July 2011), URL: https://cds.cern.ch/record/1369219.

[107] ATLAS Collaboration, *BTaggingBenchmarks ATLAS Internal TWiki page*, URL: https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/BTaggingBenchmarks (visited on 12/01/2015).

[108] ATLAS Collaboration, *Calibrating the b-Tag Efficiency and Mistag Rate in 35 pb^{-1} of Data with the ATLAS Detector*, ATLAS-CONF-2011-089 (June 2011), URL: http://cds.cern.ch/record/1356198.

[109] ATLAS Collaboration, *Measuring the b-tag efficiency in a top-pair sample with 4.7 fb^{-1} of data from the ATLAS detector*, ATLAS-CONF-2012-097 (July 2012), URL: http://cds.cern.ch/record/1460443.

[110] ATLAS Collaboration, *Measurement of the Mistag Rate with 5 fb^{-1} of Data Collected by the ATLAS Detector*, ATLAS-CONF-2012-040 (Mar. 2012), URL: https://cds.cern.ch/record/1435194.

[111] ATLAS Collaboration, *Performance of Missing Transverse Momentum Reconstruction in ATLAS with 2011 Proton-Proton Collisions at $\sqrt{s}$ = 7 TeV*, ATLAS-CONF-2012-101 (July 2012), URL: https://cds.cern.ch/record/1463915.

[112] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s}$ = 7 TeV with ATLAS*, Eur. Phys. J. C 72 (Jan. 2012) 1844, DOI: 10.1140/epjc/s10052-011-1844-6.

[113] ATLAS Collaboration, *Improved luminosity determination in pp collisions at $\sqrt{s}$ = 7 TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C 73 (Aug. 2013) 2518, DOI: 10.1140/epjc/s10052-013-2518-3.

[114] S. van der Meer, *Calibration of the effective beam height in the ISR*, CERN-ISR-PO-68-31 (1968), URL: https://cds.cern.ch/record/296752.

[115] C. Rubbia, *Measurement of the luminosity of $p\overline{p}$ collider with a (generalized) Van der Meer Method*, CERN-$p\overline{p}$-Note-38 (Nov. 1977), URL: https://cds.cern.ch/record/1025746.

[116] ATLAS Collaboration, *Atlas Computing: Technical Design Report*, ATLAS-TDR-17, CERN-LHCC-2005-022 (2005), URL: http://cds.cern.ch/record/837738.

[117] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, JHEP 06 (June 2002) 029, DOI: 10.1088/1126-6708/2002/06/029.

[118] G. Corcella et al., *HERWIG 6: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, JHEP 01 (Jan. 2001) 010, DOI: 10.1088/1126-6708/2001/01/010.

[119] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP 11 (Nov. 2004) 040, DOI: 10.1088/1126-6708/2004/11/040.

[120] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, JHEP 11 (Nov. 2007) 070, DOI: 10.1088/1126-6708/2007/11/070.

[121] S. Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP 6 (June 2010) 043, DOI: 10.1007/JHEP06(2010)043.

[122] M. L. Mangano et al., *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP 07 (July 2003) 001, DOI: 10.1088/1126-6708/2003/07/001.

[123] Borut Paul Kersevan and Elzbieta Richter-Was, *The Monte Carlo event generator AcerMC versions 2.0 to 3.8 with interfaces to PYTHIA 6.4, HERWIG 6.5 and ARIADNE 4.1*, Comput. Phys. Comm. 184 (Mar. 2013) 919, DOI: 10.1016/j.cpc.2012.10.032.

[124] T. Sjöstrand, S. Mrenna and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP 05 (May 2006) 026, DOI: 10.1088/1126-6708/2006/05/026.

[125] J. Butterworth, J. Forshaw and M. Seymour, *Multiparton interactions in photoproduction at HERA*, Zeitschrift für Physik C: Particles and Fields 72 (Dec. 1996) 637, DOI: 10.1007/s002880050286.

[126] *Official web page for TAUOLA PHOTOS and universal interface for TAUOLA*, URL: http://wasm.home.cern.ch/wasm/f77.html (visited on 12/01/2015).

[127] H.-L. Lai et al., *New parton distributions for collider physics*, Phys. Rev. D 82 (Oct. 2010) 074024, DOI: 10.1103/PhysRevD.82.074024.

[128] A. Sherstnev and R. Thorne, *Different PDF approximations useful for LO Monte Carlo generators*, Proceedings of 16th International Workshop on Deep Inelastic Scattering and Related Subjects (DIS 2008), (London, UK), ed. by R. Devenish and J. Ferrando, Science Wise Publishing, (Apr. 2008) 149, arXiv: 0807.2132 [hep-ph].

[129] J. Pumplin et al., *New Generation of Parton Distributions with Uncertainties from Global QCD Analysis*, JHEP 07 (July 2002) 012, DOI: 10.1088/1126-6708/2002/07/012.

[130] S. Agostinelli et al., *Geant4 – a simulation toolkit*, Nucl. Instr. and Meth. A 506 (July 2003) 250, DOI: 10.1016/S0168-9002(03)01368-8.

[131] ATLAS Collaboration, *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*, ATL-PHYS-PUB-2010-013 (Oct. 2010), URL: https://cds.cern.ch/record/1300517.

[132] ATLAS Collaboration, *Characterisation and mitigation of beam-induced backgrounds observed in the ATLAS detector during the 2011 proton-proton run*, JINST 8 (July 2013) P07004, DOI: 10.1088/1748-0221/8/07/P07004.

[133] ATLAS Collaboration, *New ATLAS event generator tunes to 2010 data*, ATL-PHYS-PUB-2011-008 (Apr. 2011), URL: http://cds.cern.ch/record/1345343.

[134] P. Z. Skands, *Tuning Monte Carlo generators: The Perugia tunes*, Phys. Rev. D 82 (Oct. 2010) 074018, DOI: 10.1103/PhysRevD.82.074018.

[135] R. Gavin et al., *W physics at the LHC with FEWZ 2.1*, Comput. Phys. Commun. 184 (Jan. 2013) 208, DOI: 10.1016/j.cpc.2012.09.005, arXiv: 1201.5896 [hep-ph].

[136] J. M. Campbell, R. K. Ellis and C. Williams, *MCFM web page*, URL: http://mcfm.fnal.gov/ (visited on 12/01/2015).

[137] ATLAS Collaboration, *ATLAS Public Results*, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic (visited on 12/01/2015).

[138] 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: Rheinische Friedrich-Wilhelms-Universität Bonn, 2012, URL: http://hss.ulb.uni-bonn.de/2012/2967/2967.htm.

[139] P. Seema,
Search for single top-quark production in the Wt-channel with 1 lepton and 2 jets at ATLAS,
BONN-IB-2013-01, Master Thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2013,
URL: http://brock.physik.uni-bonn.de/files/pubs/Seema_MasterThesis.pdf.

[140] ATLAS Collaboration,
Search for FCNC single top-quark production at $\sqrt{s} = 7$ TeV with the ATLAS detector,
Phys. Lett. B 712 (June 2012) 351, DOI: 10.1016/j.physletb.2012.05.022.

[141] 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. B 717 (Oct. 2012) 330,
DOI: 10.1016/j.physletb.2012.09.031.

[142] H. Voss, "Classification", *Data Analysis in High Energy Physics*, Wiley-VCH, (2013) p. 153,
DOI: 10.1002/9783527653416.ch5.

[143] M. Feindt, *A Neural Bayesian Estimator for Conditional Probability Densities*,
IEKP-KA/04-05 (Feb. 2004), arXiv: physics/0402093.

[144] M. Feindt and U. Kerzel, *The NeuroBayes neural network package*,
Nucl. Instr. and Meth. A 559 (Apr. 2006) 190, DOI: 10.1016/j.nima.2005.11.166.

[145] Phi-T Physics Information Technologies GmbH, *The NeuroBayes® User's Guide*, Apr. 2010,
URL: http://neurobayes.phi-t.de/nb_doc/NeuroBayes-HowTo.pdf.

[146] P. Sturm, *Measurement of the t-Channel Single Top-Quark Production Cross-Section with the ATLAS Detector at $\sqrt{s} = 7$ TeV*, CERN-THESIS-2012-211,
PhD Thesis: Bergische Universität Wuppertal, 2012,
URL: https://cds.cern.ch/record/1501083/.

[147] CDF Collaboration,
Observation of single top quark production and measurement of $|V_{tb}|$ with CDF,
Phys. Rev. D 82 (Dec. 2010) 112005, DOI: 10.1103/PhysRevD.82.112005.

[148] F. James and M. Roos, *Minuit - a system for function minimization and analysis of the parameter errors and correlations*, Comp. Phys. Comm. 10 (Dec. 1975) 343,
DOI: 10.1016/0010-4655(75)90039-9.

[149] *Scientific Linux*, URL: https://www.scientificlinux.org (visited on 13/01/2015).

[150] *ROOT Data Analysis Framework*, URL: https://root.cern.ch (visited on 13/01/2015).

[151] *SFrame – A ROOT data analysis framework*,
URL: http://sourceforge.net/projects/sframe/ (visited on 13/01/2015).

[152] O. M. Kind, *Production of Heavy Flavours with Associated Jets at HERA*, BONN-IR-06-01,
PhD Thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2006,
URL: http://hss.ulb.uni-bonn.de/2007/1105/1105.htm.

[153] *CMake – cross-platform, open-source build system*,
URL: http://www.cmake.org (visited on 13/01/2015).

[154] *Bazaar version control system*,
URL: http://bazaar.canonical.com (visited on 13/01/2015).

[155] *Git version control system*, URL: http://git-scm.com (visited on 13/01/2015).

[156] *SFM wiki page*,
URL: https://wiki.physik.uni-bonn.de/atlas/public/index.php/SFM (visited on 13/01/2015).

[157] P. Mehnert, *Bestimmung der Top-Quark-Masse anhand der Transversalimpulse der Zerfallsleptonen im Lepton+Jets-Kanal mit dem ATLAS-Detektor*, BONN-IB-2011-10, Diplomarbeit: Rheinische Friedrich-Wilhelms-Universität Bonn, 2011,
URL: http://brock.physik.uni-bonn.de/files/pubs/BONN-IB-2011-10.pdf.

[158] M. Müller, *Anwendung der System8-Methode zur Bestimmung von b-Tagging-Effizienzen am ATLAS-Detektor*, BONN-IB-2011-11,
Diplomarbeit: Rheinische Friedrich-Wilhelms-Universität Bonn, 2011,
URL: http://brock.physik.uni-bonn.de/files/pubs/BONN-IB-2011-11.pdf.

[159] E. Nickel, *Messung der Top-Quark-Masse mittels der Lepton-Transversalimpuls-Methode mit einem Datensatz von 5 fb^{-1} am ATLAS-Detektor*, BONN-IB-2014-04,
Diplomarbeit: Rheinische Friedrich-Wilhelms-Universität Bonn, 2012,
URL: http://brock.physik.uni-bonn.de/files/pubs/BONN-IB-2014-04.pdf.

[160] I. Cioară, *Optimization of the single top-quark production analysis in the Wt channel at ATLAS using kinematic fitting and neural networks*, BONN-IB-2014-08,
Master Thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2014,
URL: http://brock.physik.uni-bonn.de/files/pubs/BONN-IB-2014-08.pdf.

[161] M. Hansen, *Application of the matrix element method to top-quark physics at the LHC*, BONN-IR-2014-06, Master Thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2013,
URL: http://brock.physik.uni-bonn.de/files/pubs/BONN-IB-2011-11.pdf.

[162] S. Zimmermann,
A search for stable massive particles carrying electric charges in the range of 2e to 6e in proton-proton collisions at $\sqrt{s} = 7$ TeV recorded with the ATLAS detector at the LHC,
BONN-IR-2013-12, PhD Thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2013,
URL: http://hss.ulb.uni-bonn.de/2013/3323/3323.htm.

[163] P. Kövesárki, *Multivariate methods and the search for single top-quark production in association with a W boson in ATLAS*, BONN-IR-2012-13,
PhD Thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2012,
URL: http://hss.ulb.uni-bonn.de/2013/3188/3188.htm.

[164] ATLAS Collaboration, *Search for long-lived, multi-charged particles in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, Phys. Lett. B 722 (May 2013) 305,
DOI: 10.1016/j.physletb.2013.04.036.

[165] ATLAS Collaboration, *Search for events with large missing transverse momentum, jets, and at least two tau leptons in 7 TeV proton-proton collision data with the ATLAS detector*,
Phys. Lett. B 714 (Aug. 2012) 180, DOI: 10.1016/j.physletb.2012.06.055.

[166] ATLAS Collaboration,
Search for supersymmetry in events with large missing transverse momentum, jets, and at least one tau lepton in 7 TeV proton-proton collision data with the ATLAS detector,
Eur. Phys. J. C 72 (Nov. 2012) 2215, DOI: 10.1140/epjc/s10052-012-2215-7.

[167] ATLAS Collaboration,
Search for supersymmetry in events with large missing transverse momentum, jets, and at least one tau lepton in 20 fb^{-1} of $\sqrt{s}$ = 8 TeV proton-proton collision data with the ATLAS detector,
JHEP 09 (Sept. 2014) 103, DOI: 10.1007/JHEP09(2014)103.

List of Figures

List of Tables

Acknowledgements

First of all I want to express my deepest gratitude to my Ph.D. advisor Professor Ian C. Brock for the financial, scientific and personal support of the thesis project itself as well as for giving me the freedom to develop ideas beyond the thesis topic, like software programming, creation of the STYX experiment and learning high-performance computing cluster administration. I would also like to thank Professor Jochen Dingfelder, Professor Bernard Metsch and Professor Mariele Evers for their interest on the topic of the thesis and the interesting discussions during the disputation. I want to thank the colleagues involved in the *Wt* analysis in Bonn, namely Thomas Velz, Sebastian Mergelmeyer, Pienpen Seema and Irina Cioară, for their teamwork and their efforts to verify the existence of this challenging production mechanism despite all the setbacks. Thomas and Sebastian also contributed considerably to the software development part and without them, the SFM software certainly would not have been successful on a larger scale. Additionally, I want to thank the single top analysis group of the University of Wuppertal: Wolfgang Wagner, Dominic Hirschbühl, Philipp Sturm and Kathrin Becker. Thank you for answering all those questions about NeuroBayes, the Bill tool for likelihood fitting, the JetElectron multijet estimation and countless ideas for solving the problems on the analysis path.

Special thanks go to Markus Jüngst and Steffen Schaepe for supporting the thesis project and myself in countless aspects, which I can only name a few. Thank you both for the time at CERN we spent together on several occasions, the foundation and realisation of the STYX experiment, the support of the SFM software, the great hiking trips, the endless discussions about physics and data analysis, sailing on the lake Geneva and of course for the good friendship.

I'm very grateful to have worked together with many great colleagues in the Brock group during the time of my studies, in particular with Muhammad Alhroob, Ozan Arslan, Detlef Bartsch, Peter Kövesárki, Verena Schönberg, Michał Własenko and Professor Ewald Paul. Thanks also go to the students Andrea Wagner, Philipp Mehnert, Ismail Nasser, Pienpen Seema and Dustin Hebecker who I supervised during their bachelor and master thesis times. It was great working with all of you and I hope you learned as much as you taught me. I'm also very thankful for the opportunity to work at CERN for more than thirteen months and for spending time together with Götz Gaycken, Marc Lehmacher, Kristof Schmieden, Jan Therhaag, Mathias Uhlenbrock and Simone Zimmermann. Christoph Rembser I want to thank for the great support of the STYX experiment and during my work for the ATLAS TRT group.

I want to acknowledge the good collaboration with the entire working group of Professor Klaus Desch and want to thank the people who helped me operating the computing infrastructure for all ATLAS groups in Bonn. This wouldn't have been possible without Peter Wienemann, the advice of the former administrator Gizo Nanava and the help of the group admins Irina Cioară, David Hohn, Oliver Ricken, Steffen Schaepe, Markus Voge, Stephanie Yuen and Rui Zhang.

Last but not least very special thanks go to my wife Miryam for her tireless support, great understanding and all the love she gave me and to my parents. The finalization of this thesis wouldn't have been possible without all of you!