Universität Bonn

Physikalisches Institut

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

Thomas Loddenkötter

In order to provide discrimination between the *Wt*-channel signal and its backgrounds for analyses that try to measure single top-quark production in the *Wt*-channel, a kinematic fit to the lepton+jets decay mode of the *Wt*-channel has been implemented using the KLFitter package. The fit has been validated by studying its performance in terms of the efficiency of the fit to correctly assign the final-state quarks of the fit model to the measured jets as a function of various parameters, as well as the improvement of the energy resolutions of the fitted particles due to the fit. By combining the output variables of the kinematic fitter using neural networks, it has been shown that the fit results are suitable to identify the decay mode of the top quark in *Wt* events and to identify whether the kinematic fit succeeded in correctly assigning the final-state quarks to the measured jets. In order to demonstrate the value of the kinematic fit for analysis, another neural network – again using strictly results of the kinematic fit as input – has been trained to separate to the *Wt*-channel signal from its backgrounds. A separation power comparable to a conventional neural-network-based *Wt*-channel analysis has been achieved.

Physikalisches Institut der
Universität Bonn
Nußallee 12
D-53115 Bonn

BONN-IR-2012-06
August 2012
ISSN-0172-8741

Universität Bonn

Physikalisches Institut

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

Thomas Loddenkötter

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

1. Gutachter:	Prof. Dr. Ian C. Brock
2. Gutachter:	Prof. Dr. Klaus Desch
Angenommen am:	30.01.2012
Tag der Promotion:	09.03.2012

Preface

The aim of elementary particle physics is to identify and study the fundamental building blocks of matter and the interactions between them. The theory that comprises our current knowledge about these topics is the Standard Model of elementary particle physics.

The heaviest among the known elementary particles that are described by the Standard Model is the top quark. Its very high mass of about 173 GeV makes it very special among all other quarks. For example, it plays an important role in higher-order corrections to many other processes; furthermore it is the only quark that decays before it can form bound states, which makes it possible to measure the top-quark mass much more precisely than that of any other quark. Such measurements can be used in order to put indirect constraints on the mass of the Higgs boson, the only particle of the Standard Model that has not been found experimentally yet. For these and other reasons, top-quark physics is a very interesting research field. Present measurements of the top quark have been performed at two hadron colliders: the proton–antiproton collider Tevatron, where the top quark was discovered in 1995, and the proton–proton collider LHC, which started colliding protons at a center-of-mass energy of 7 TeV in 2010 and yielded first top-quark measurements already in the same year. At these colliders, top quarks can be produced either in pairs of a top quark and an antitop quark via the strong interaction or as single top quarks via the weak interaction. Single top-quark production happens much more rarely than top quark pair production and is more difficult to measure. It was discovered only in 2009 at the Tevatron. First measurements of single top-quark production at the LHC followed in the first half of 2011.

Single top-quark production is an interesting field on its own. Measurements of single top-quark production cross sections are the only way to directly determine a top-quark-related element of the Cabibbo-Kobayashi-Maskawa (CKM) matrix. Furthermore different kinds of new physics beyond the Standard Model are expected to show up in single top-quark production. Single top-quark production in the Standard Model can happen in three ways, which are called t-channel production, s-channel production and Wt-channel production. While a combination of t-channel and s-channel production was used for the first single top-quark measurement at the Tevatron and pure t-channel production could be measured at the LHC, the Wt-channel has not been seen experimentally yet. At the Tevatron the Wt-channel cross section is too small to be measured at all. But at the LHC it is predicted to give the second largest contribution to the total single top-quark production cross section (after t-channel production). As such it is expected to be measurable as soon as a sufficient amount of data is available.

This thesis deals with the measurement of Wt-channel production in the lepton+jets decay mode in ATLAS, which is one of the four major experimental collaborations at the LHC. One general problem of Wt-channel analyses is that it is very hard to find variables that provide good separation from background. The point is that the two main backgrounds, W+jets production and top quark pair production, have signatures that are very similar to that of Wt-channel production. In particular, top quark pair production contains the full Wt-channel signature and may even interfere with Wt-channel production.

The idea that is followed in this thesis is to employ a kinematic fit to the signal topology. Wt-channel production is characterized by the associated production of a top-quark and an on-shell W boson. In the lepton+jets decay mode this leads to one b-quark jet, two light-quark jets, one charged lepton and a neutrino in the final state. The final state is measured with the ATLAS detector. The kinematic fit then tries to reconstruct the top quark and the W boson from the measured objects. In doing so, it provides

some output variables that are directly sensitive to how much an event looks like *Wt*-channel signal. For the final measurement these variables can then be combined with other variables in a neural network. The focus of this thesis is on the implementation of the kinematic fit using the KLFitter package, its validation and the evaluation of its performance. Furthermore a neural network taking solely output from the kinematic fit as input variables is used to demonstrate that these are indeed useful for analysis.

The outline of the thesis is as follows: Chapter 1 is mostly aimed at readers that are not familiar with particle physics. It comprises a general introduction to the Standard Model and its underlying theoretical concepts as well as to some further theoretical and experimental concepts that are prevalent in hadron collider physics. In Chapter 2 the LHC and the ATLAS detector with its main components are introduced. Chapter 3 deals with the phenomenology of the *Wt*-channel. It starts very generally with an introduction to the top quark itself, its production and its decay, and it ends with the definition of the signal final state that is considered for analysis and the list of relevant background processes. Chapter 4 describes how the relevant physical objects for the analysis are reconstructed by the ATLAS software. The topic of Chapter 5 is the basic analysis setup. It contains the descriptions of all used datasets, the precise definitions of the physics objects that are used for event selection, the event selection itself as well as a description of the general analysis strategy. In Chapter 6 the implementation of the kinematic fit to the *Wt*-channel topology is detailed, before Chapter 7 deals with the validation of this implementation and the study of the performance of the fit. In Chapter 8 the output variables of the kinematic fit are combined by neural network, first in order to distinguish different subsets of the simulated signal dataset, then for the actual analysis, i.e. in order to discriminate signal from background. Chapter 9 finally summarizes the results of the thesis.

Contents

Chapter 1

Theoretical and experimental basics

This chapter is mainly addressed to readers that are not familiar with particle physics in general and physics at hadron colliders in particular. It introduces the fundamental theoretical and experimental basics of this field. The first section briefly introduces the overall situation, setting the scene for the rest of the chapter: what is particle physics all about, what kind of experiments are conducted, what kind of quantities do they measure and which theories exist that are used to predict the measurement results? The next two sections deal with the theoretical side: the first of them introduces some general theoretical concepts that are used in particle physics, the latter describes how the Standard Model, which is the central theory in particle physics to date, is built from these concepts. The next section then explains how the Standard Model is used in order to calculate predictions for hadron collider experiments. Finally some general experimental concepts that are used (not only) at hadron colliders are introduced.

1.1 Setting the scene

1.1.1 Experiments in particle physics

Elementary particle physics tries to understand the fundamental building blocks of matter and the interactions between them. The most important type of experiments that are conducted in order to reach this goal are scattering experiments. In scattering experiments, particles are accelerated and shot at other particles. The target particles can be either at rest (fixed-target experiments) or accelerated as well (collider experiments). The collision products are registered by detectors that are built close to or around the collision points and the data are analyzed statistically. One important type of experimental setup are hadron colliders, in which protons collide with either protons or antiprotons. The two most energetic hadron colliders to date are the Tevatron, a proton–antiproton collider located near Chicago, and the LHC, a proton–proton collider, which is described in Chapter 2. At these colliders one is mostly interested in *inelastic* collisions, in which the initial-state protons[1] break up and various types of new particles are created (in contrast to *elastic* collisions, in which the protons are only deflected but stay intact).

1.1.2 Basic observables

Cross section and luminosity. The most important type of physical observables that are measured in scattering experiments are *cross sections*, σ, of the processes that occur. A cross section is a measure for the probability of a given process to happen in a collision. It can be thought of as the effective size of an area around the center of a target particle that a probe particle has to hit in order to trigger that process. The unit that is used to express cross sections is called *barn* (b), where 1 b = 10^{-22} cm^{-1}. In

[1] Here and in for the rest of this chapter, "proton" may denote either proton or antiproton where applicable.

practice, the cross sections for many relevant processes are in the order of picobarn or femtobarn. Cross sections are calculated from theories using the general relation

$$\sigma \propto |M_{if}|^2 \cdot \rho \,, \tag{1.1}$$

where all the relevant physics is encoded in the so-called the *matrix element* or *transition amplitude* M_{if} for the transition of the given initial state (i) to the final state (f) of the process of interest, and ρ is a phase space factor that shall not be of further interest here. The way how the matrix elements are calculated from theory is sketched in Section 1.2.3

The average rate $\mathrm{d}N/\mathrm{d}t$ with which a given reaction occurs is directly proportional to its cross section:

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

The factor $\mathcal{L}$ is called the (instantaneous) *luminosity*. The luminosity is the general measure for the collision rate at a collider. It can either be calculated from beam parameters, such as the particle flow and density in each beam and the beam sizes, or by measuring the rate of a process that has a known cross section. The integral over time of the luminosity, $\int \mathcal{L}\, \mathrm{d}t$, is called the *integrated luminosity*. which is the measure for the amount of data that was taken. Integrated luminosities are usually expressed in inverse barn (or any derivatives thereof, i.e. pb^{-1}, fb^{-1} etc.), instantaneous luminosities are expressed in $\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ or as the amount of total luminosity that is collected per year.

Decay width and branching fraction. The analog of the cross section for the decay of a particle, is called the particle's *decay width*, Γ. Decay widths are expressed in electron-volts (eV), which is the standard energy unit in particle physics. Each individual decay mode has its own corresponding (partial) width, Γ_i; and sum of the widths of all possible decay modes is called the total width. The total width of a particle is directly related to its lifetime, τ, via the relation

$$\Gamma = \frac{1}{\tau} \,.$$

That is, the larger the width, the shorter the lifetime of a particle is. The ratio of the partial width Γ_i of a given decay mode to the total decay width is called the *branching fraction*, $\mathcal{B}$, of that decay mode:

$$\mathcal{B} = \frac{\Gamma_i}{\Gamma} \,.$$

Natural units. In the previous paragraph, the relation $\Gamma = \frac{1}{\tau}$ was used. In terms of physical dimensions this means [energy] = [time]$^{-1}$, which is wrong in standard SI units. In fact, in these units the relation contains the reduced Planck constant, $\hbar$, in addition: $\Gamma = \frac{\hbar}{\tau}$, which fixes the physical dimensions. In particles physics one instead uses so-called *natural units*, i.e. any occurrence not only of $\hbar$, but also of the speed of light, c, is dropped from all equations and relations; phrased more properly, one defines $c = \hbar = 1$. As a consequence, not only time can be expressed in terms of inverse energy units, but also masses and momenta are expressed in terms of energy units.

1.1.3 Theory in particle physics

The most important theory that is used to predict and compare the results of experiments is the so-called *Standard Model* of elementary particle physics [1–11]. Is was formulated already back in the 1970s, but

it still successfully describes all relevant measurements that have been conducted so far. Some of the particles that are predicted by the Standard Model were discovered by experiment only decades later. One last particle, the Higgs boson, is just about to be discovered at the LHC, if it exists. The basic theoretical concepts upon which the Standard Model is built are introduced in Section 1.2, the Standard Model itself is topic of Section 1.3.

1.2 Basic theoretical concepts

1.2.1 General ideas

Quantum field theory. Theories in particle physics are mostly quantum field theories. In such theories, particles are represented by fields (i.e. wave-functions), interactions between particles are mediated by the exchange of other particles (which are themselves represented by fields). Theories are fully defined by what is called their *Lagrangian density* (also simply called the Lagrangian), which is a scalar function of the fields of all particles that are described by the respective theory. Physical properties can be derived from the Lagrangian density by applying general theoretical principles, which lead to equations that the Lagrangian density must satisfy and which then form physics laws.

The Lagrangian. The form of the Lagrangian is not completely arbitrary, but it consists of several terms, each of which has a well-defined physical interpretation: a product of three or more fields denotes an interaction between the corresponding particles (interaction terms); other terms denote the kinetic energy of a particle (kinetic term) or the mass of a particle (mass terms). When a Lagrangian density contains an interaction term that comprises a certain combination of particles, one says that these particles "couple" to each other. A Lagrangian that does not contain any interactions is called a *free* Lagrangian.

Parameters of theories. Any constants that occur in a theory, whose values are not predicted by the theory itself, are called the free *parameters* of the theory. When the structure of a theory – defined by the precise shape of its Lagrangian – is known, the full behavior of the theory is fixed by the values of these parameters. In particular the knowledge of the parameter values is necessary in order to calculate predictions. As these are not predicted by the theory itself, they need to be determined experimentally. It turns out that this is in fact not always as trivial as it seems at first glance (see Section 1.2.4).
Mainly two types of parameters occur: particle masses, which are the coefficients of the mass terms in the Lagrangian density, and *coupling(strength)s*,[2] which are the coefficients of the interaction terms and determine how strongly the involved particles couple to (i.e. interact with) each other.

1.2.2 Gauge theory

Global gauge invariance. Particle fields, being wave-functions, contain phases. The physics that is encoded in a Lagrangian density is, however, supposed to be independent of the overall phase of the Lagrangian. Consequently one is in general free to arbitrarily apply any phase transformation – also referred to as gauge transformation – to the Lagrangian density without changing the physics that is described by it. This is called (global) gauge invariance or gauge symmetry of the theory. A theory may contain several such symmetries. These are typically invariances under some generalized rotations

[2] The coupling strengths are also often referred to as *coupling constants*, although in a strict sense they are not necessarily constant (see Section 1.2.4).

in higher spaces – just like the given example of a simple phase transformation represents a rotation in the one-dimensional complex plane. The symmetry is called global because the phase transformation is the same at each point in space-time. Each such symmetry of the Lagrangian refers to a *symmetry group* (also called *gauge group*), which is basically the set of all transformations that correspond to that symmetry (e.g. all rotations around a particular axis by any angle).

Local gauge invariance. In contrast to global transformations, which apply the same transformation to each point in space-time, in a local gauge transformation the phase by which the whole Lagrangian gets shifted is a function of space-time. At first, a naively constructed Lagrangian density is not invariant under such transformations. This can be resolved by a trick. By replacing all derivatives that occur in a free Lagrangian by what is called the "covariant derivative", the Lagrangian can be made invariant also under local gauge transformations. This covariant derivative contains one or more additional fields, called gauge fields. Evaluating this modified Lagrangian, one finds that by demanding local gauge invariance, one has introduced interactions between the formerly free particles, mediated by the particles that correspond to the newly introduced gauge fields. Interactions that are created this way are often called gauge interactions and the particles that correspond to the gauge fields are called gauge bosons. This mechanism of creating interactions by local gauge invariance is considered to be particularly elegant, as the interaction does not need to be inserted into the theory by hand, but arises automatically from a general principle, by "localizing" a global symmetry. In particular, the structure of the interaction is completely determined by the gauge group, i.e. by the symmetry which was enforced to become local. Theories that are constructed this way are called gauge theories.

Coupling strengths and charges. The coupling strength of a gauge interaction appears in the Lagrangian as the coefficient to (each of) the gauge fields in the covariant derivative. In particular, even for gauge interactions that have multiple gauge fields, there is only one single coupling strength, which appears in all the interaction terms of these gauge fields.

In the interaction term of a particle with a gauge boson, the coupling strength is always multiplied by a factor that is specific to the particle type. This factor is called the *charge* of the particle regarding this interaction. By having different charges, different particle types can couple differently the same gauge boson, despite the universal nature of the coupling strength. In particular, only particles that have non-zero charge participate in the corresponding interaction. Gauge bosons may be charged as well, in this case they also interact with themselves. This is called self-interaction. Whether or not gauge bosons carry charge depends on the type of the symmetry that mediates the interaction. In case these symmetry transformations commute with each other (abelian gauge group), the gauge bosons do not carry charge. In case the gauge bosons do not commute (non-abelian gauge group), they do.

As an example for these concepts, the elementary charge is the coupling strength of the electromagnetic interaction, while the charge of the electron in the sense defined above is -1. The symmetry that generates the electromagnetic interaction is a simple phase transformation, which is abelian. Therefore the photon, which is the gauge boson of the interaction, is electrically neutral.

1.2.3 Perturbation theory

Concept. The equations that arise from the Lagrangian density in order to calculate the matrix elements for scattering processes can normally not be solved analytically. The usual approach to overcome this problem when calculating cross sections is to approximate the squared matrix element by a power series in the coupling strength(s). As even with this approximation typically only few terms of the power series can be calculated, this is only efficient when the power series converges fast. A crucial

requirement for this approach is therefore that the coupling strengths are sufficiently small, i.e. $\ll 1$, which can be interpreted such that the interactions are only small perturbations of the free motion of particles. This motivates the term *perturbation theory* for this approach; one also says that quantities are calculated *perturbatively*, and the power series in the coupling strengths is called a *perturbation series* or *perturbative expansion*. The individual terms of a perturbation series are called *orders*, the first order of a series is called *leading order* (LO), the second term is called *next-to-leading order* (NLO) etc.

Feynman diagrams. The individual terms of a perturbation series are usually still hard to calculate. A very useful tool for calculating them is provided by Feynman diagrams. A Feynman diagram (or Feynman graph) is a pictorial representation of a certain reaction as well a precise representation of a corresponding mathematical term both at the same time. Feynman graphs consist of external lines, representing incoming and outgoing particles, internal lines, representing intermediate particles, and vertices, which are the points at which lines meet. All these components have precise associations to mathematical expressions. These associations are called Feynman rules, and they arise uniquely from the Lagrangian. The full diagram then corresponds to the product of its components

All possible Feynman diagrams that contribute to a given process (i.e. that have this initial and final state) add up to the full matrix element for the process. For the calculation of a cross section up to a given order, one only needs to include all the diagrams that contribute to the squared matrix element up to this order. These can be identified by by counting the vertices in a Feynman diagram, as the vertices are related to the interaction terms of the Lagrangian and hence their Feynman rules contain the respective coupling strength (and of course the relevant charge). Obviously, the cross section in general contains not only the squares of the individual diagrams, but also interference terms between the diagrams (though these may be zero). While individual diagrams look like a precise sketch of the way some interaction may take place, this is not the case for the interference terms at all. This makes clear that the interpretation of Feynman graphs as a pictorial representation of a certain process is limited and should be taken with care.

1.2.4 Renormalization

To make a theoretical prediction means to calculate a measurable value from the parameters of the theory. As stated in Section 1.2.1, this requires that first the parameters themselves need to be determined from experiment. To experimentally determine a parameter means that a measurement is performed and the parameter value is calculated from the outcome of this measurement. The necessary relation between the parameter and the outcome of the measurement is, of course, again obtained by calculating a prediction for the measurement as a function of the parameter. An example for this would be the mass parameter of a particle that needs to be related to the experimentally determined mass. So far this all sounds trivial. The point that make things complicated is that predictions are calculated in perturbation theory, including the relations that are used to determine the parameters. This means that these relations inherently are only valid at a certain order. Assume a theory with only three parameters and one determines these parameters from three measurements, calculating the required relations at a certain order of perturbation theory. Using these parameter values for predictions at any other order will then lead to wrong results, in particular one could not even correctly predict the measured values from which the parameters were determined! Instead, when moving to a different order, the parameters need to be *renormalized*, i.e. recalculated from measurements using relations at the same order at which one wants to calculate predictions. Renormalization therefore means nothing else than that for a prediction at a certain order one always has to use parameters that have been determined at the same order.

Phrased like this, renormalization sounds quite intuitive and natural. In practice, intuition is spoiled by two points. Firstly, some parameters normally have a somewhat universal notion, such that one is tempted to identify them one-to-one with the corresponding theory parameters. A particle mass being different at LO and NLO, for example, in fact seems quite unnatural. Here one has to keep in mind that only the theory parameter changes in order to preserve the prediction for the measured mass, as it should. The second point, which is even much more disturbing, is that at NLO the relations between the parameters and their measured values may even contain singularities. As a consequence the parameters end up having infinite values. Though this indeed looks like a problem, it turns out that in practice this is fine as long as the number of such singularities is finite. Each singularity is then absorbed into the renormalization of one parameter,[3] and in the prediction of any measurable quantity these infinities cancel, such that the predictions remain finite. When this is the case, a theory is called *renormalizable*.[4]

Renormalization involving infinities has one serious consequence, however: the precise definition of the theory parameters is not unique any more. Rather there is some freedom in how and how much of the divergences to absorb into the renormalized parameters. The exact prescription of the "how" is called a *renormalization scheme*. Usually a renormalization scheme involves a free scale parameter which adjusts the "how much", and which is called the *renormalization scale*. This way, the theory parameters become functions of the scale, one speaks e.g. about "running coupling(strength)s". For any prediction therefore a (value for the) scale has to be chosen, which is normally done such that is corresponds to what one considers to be the natural scale of the relevant process, e.g. the mass of the heaviest involved particle or the center-of-mass energy of a collision. Though any physical quantity of course should not depend on such an arbitrary scale, predictions at any finite order generally do. This is called the *(renormalization) scale uncertainty*, which is supposed to decrease toward higher orders. The scale uncertainty is typically one of the main uncertainties of theoretical predictions.

1.2.5 The Higgs mechanism

By construction, gauge theories as introduced above only describe massless gauge bosons. In reality we know that some of the gauge bosons are in fact among the heaviest particles of the Standard Model. Without further measures, explicit mass terms for them in the Lagrangian would not only break down gauge symmetry, but the attempt of accepting the mass terms, discarding gauge invariance, would even destroy renormalizability, rendering the theory useless.

To rescue gauge theory even in the face of massive gauge bosons, a mechanism of *spontaneous symmetry breaking* is introduced, the Higgs mechanism. The Higgs mechanism adds an additional field to the theory, the Higgs field. The Higgs field comes along with interactions with the gauge bosons and with a corresponding Higgs potential (i.e. a scalar function of the Higgs field which enters the Lagrangian). This potential is chosen such that its vacuum expectation value (VEV), i.e. the position of its minimum, is non-zero. Furthermore the requirement to put the Higgs potential to that minimum only fixes some of the degrees of freedom of the Higgs field, while it is degenerate in the remaining ones. When conducting perturbation theory, the perturbation series always has to be expanded around the ground state, i.e. the vacuum. In terms of the Higgs field this refers to the VEV. This gives rise to what is called spontaneous (gauge) symmetry breaking: though the full Lagrangian and hence the overall theory remains gauge invariant, this symmetry is not visible at the ground state. The symmetry is rather

[3] It may in fact happen that new parameters need to be introduced into the theory, e.g. if more singularities occur than the theory has parameters. For the Standard Model, this is not the case, however.

[4] For a non-renormalizable theory, an infinite number of parameters is needed in order to renormalize all singularities, and hence an infinite number of measurements is needed in order to determine all these parameters, before any prediction can be made. Apparently, such a theory is rather useless.

implicitly "hidden" in the arbitrary choice of the VEV. In fact, any state of the Higgs field that belongs to the VEV can be expressed as a gauge transformation containing those degrees of freedom of the Higgs field in which the VEV is degenerate, acting on one arbitrary state that belongs on the VEV. The full Higgs field can then be expressed as that general parametrization of the VEV plus some fluctuations that depend only on the remaining degrees of freedom. By choosing an appropriate gauge, those degrees of freedom that parametrize the VEV disappear from the Lagrangian. Instead, mass terms for as many gauge bosons appear as parameters are in the VEV; one may say that these degrees of freedom are "eaten up" by giving masses to the gauge bosons. The remaining degrees of freedom of the Higgs field, which parametrize the fluctuations around the vacuum, show up in the Lagrangian as additional scalar bosons, the Higgs bosons. Furthermore, interaction terms between the Higgs bosons and the gauge bosons show up. Any gauge bosons that remain massless correspond to remaining symmetries which are still intact in the vacuum state.

1.3 The Standard Model

This section describes how the Standard Model of elementary particle physics (SM) is built up using the general theoretical concepts that were introduced in the previous section. Furthermore, two aspects of the Standard Model are introduced that arise from this particular implementation and that are relevant for this thesis. The descriptions given this section are not meant to be complete. The focus is rather put on the conceptual structure of the theory. Many details that are not directly relevant for this thesis, such as the concept of handedness and all the subtleties that come along with it as well as the weak isospin and hypercharge quantum numbers, are completely left out, trying to still keep the remaining descriptions as precise and consistent as possible.

1.3.1 The ingredients of the Standard Model

Gauge groups, interactions and gauge bosons

The Standard Model is a renormalizable gauge theory with the gauge group $SU(3)_C \times SU(2)_L \times U(1)_Y$. The $SU(3)_C$ symmetry group describes the strong interaction. This part of the SM is called Quantum ChromoDynamics (QCD). The charge of this force is called "color", which is motivated by the three different types of color charge that exist, denoted "red", "green" and "blue" (plus the respective anticolors), which all together add up to being color neutral, like the three elementary colors add up to white. Accordingly, the sum of any two color charges results in the anticharge of the third color. The gauge bosons of $SU(3)_C$, which mediate the strong interaction, are called gluons. The $SU(2)_L \times U(1)_Y$ group describes the electroweak force, which is the unification of the electromagnetic and the weak force. The gauge bosons of both groups mix with each other, giving rise to the W and the Z boson as the mediators of the weak interaction and the photon as the mediator of the electromagnetic interaction.

Particle content

The particle content of the SM consists of twelve fermions, six of which are called quarks and six are called leptons. The quarks and leptons are arranged in three generations, each consisting of a pair of quarks and a pair of leptons. Each generation of quarks consists of one "up-type" quark, carrying electric charge +2/3, and a "down-type" quark, carrying electric charge -1/3. The six different quark types are also referred to as the quark "flavors". Each lepton generation consists of one charged lepton with electric charge -1 and an associated electrically neutral particle called neutrino. All these particles

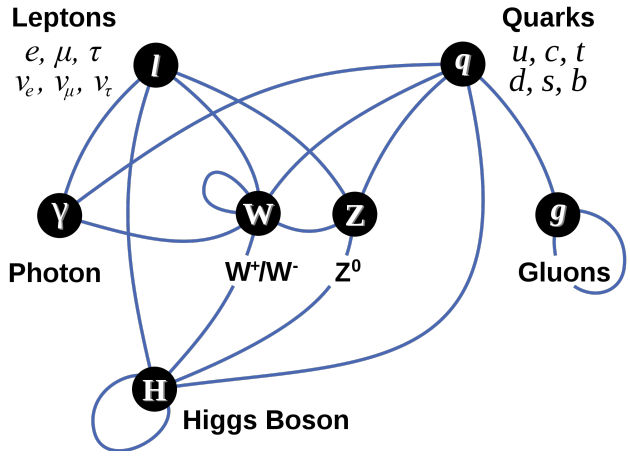


Figure 1.1: Overview of the particle content of the Standard Model and the interactions between the particles. The names of the charged leptons are electron (e), muon (μ) and tauon (or tau lepton, τ); the neutrinos are the electron neutrino (ν_e), muon neutrino (ν_μ) and tau neutrino (ν_τ). The three up-type quarks are called up (u), charm (c) and top (t), the down-type quarks are called down (d), strange (s) and bottom (b). Particles that are connected by a line interact with each other (the line between the photon and the leptons only holds for the charged leptons).

have been verified experimentally. Besides these particles, the SM contains one Higgs boson, which could not be discovered yet. An overview of the particle contents of the SM along with the interactions between them is given in figure 1.1.

Antiparticles. Each particle in the Standard Model has a corresponding antiparticle. Antiparticles have many of the same properties as the corresponding particles; in particular they have the same masses and lifetimes, while some attributes, such as the electric charge, have opposite sign. Unless stated explicitly, antiparticles are therefore generally not referred to as distinct particles, but when talking about a certain particle, its antiparticle is referred to implicitly as well. Antiparticles are denoted by putting a bar on top of the symbol of the corresponding particle, e.g. the antiparticle of X would be $\bar{X}$.

Particles and their interactions

The different particles behave differently under the three interactions. The weak interaction acts on all quarks and all leptons, while the electromagnetic interaction acts on all particles that carry charge, i.e. all quarks and leptons except for the neutrinos, which are electrically neutral. The strong interaction only acts on quarks which are the only fermions that carry a color charge. To be more precise, a quark always carries one color charge (antiquarks carry anticolor accordingly), either red or green or blue, and each quark flavor exists in all three colors, which in some contexts are treated as three distinct particles. Putting things the other way around, one can summarize that the quarks participate in all three interactions, the charged leptons participate in the electromagnetic and the weak force and the neutrinos participate only in the weak interaction.

1.3.2 The Higgs sector of the Standard Model

Gauge boson masses

Among the gauge bosons of the Standard Model, the photon and the gluons are massless, while the W and Z bosons are heavy, with masses of 80.4 GeV and 91.2 GeV, respectively. Therefore a Higgs

sector is necessary in the Standard Model. The Higgs field of the Standard Model is a complex doublet, i.e. it contains four degrees of freedom. Its vacuum expectation value is degenerate in three of them. From these, the W and Z bosons obtain their masses (the W boson needs two degrees of freedom as it is actually two gauge bosons, the W^+ and the W^-). The remaining degree of freedom gives rise to the Higgs boson of the Standard Model. The Higgs boson interacts with itself and with the W and the Z boson. The photon stays massless in contrast to the W and the Z due to a remaining $U(1)$ symmetry – the $U(1)$ symmetry of Quantum ElectroDynamics (QED), which is different from the original $U(1)_Y$.

Fermion masses

In the Standard Model, the Higgs mechanism is also used to generate the masses for the fermions. This is necessary, as explicit mass terms for the fermions would break the electroweak gauge symmetry due to the structure of the electroweak theory. The mass generation for the fermions is done in a similar way as for the gauge bosons. For each fermion, a (gauge invariant) interaction term with the Higgs field is added to the Lagrangian. After spontaneous symmetry breaking (i.e. rewriting the Higgs field in the normal way) two terms remain for each fermion. One can be identified as a mass term for the fermion, the other one describes the coupling of the fermion to the Higgs boson, with a coupling strength that is proportional to the fermion mass.

1.3.3 Quark mixing and the CKM matrix

When W bosons couple to quarks, they always couple an up-type quark to a down-type quark. In a naive construction of the theory, this coupling would always take place only within one quark generation. In contrast, one experimentally observes also couplings between different quark generations take place, though they are generally suppressed with respect to the couplings within one generation. The interpretation of this observation is that weak eigenstates of the quarks, i.e. the actual states to which the W couples, are not identical to the physical quark states, i.e. to the mass eigenstates. Rather each up- and down-type weak eigenstate contains also small admixtures of both other flavors of up- and down-type mass eigenstates, respectively. This behavior is accordingly also reflected in the theory. Mathematically, it is conventionally written such that only the down-type quarks are mixed, while the mass and weak eigenstates of the up-type quarks identical. More precisely, the weak eigenstates (d', s', b') are defined as linear combinations of the mass eigenstates (d, s, b), expressed by the CKM matrix V_{CKM}:

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

The entries of the CKM matrix then appear in the couplings of the W boson to the respective quarks and accordingly in the Feynman rules of the corresponding vertices (e.g. V_{tb} appears in the Wtb vertex).

In general V_{CKM} can be complex. The most important requirement that restricts its parametrization is that it has to be unitary. Furthermore the five relative phases between the six quarks can be chosen arbitrarily. At the end of the day, four independent parameters are left, such that the most general parametrization of V_{CKM} can be chosen as three real rotations – each mixing two quark generations – plus one complex phase that is added to one of these rotations. In particular this means that some matrix elements have an imaginary part, which can be used to explain CP violation in the Standard Model,[5]

[5] In fact, this was precisely the motivation of Kobayashi and Maskawa when they first proposed the CKM matrix in its today-known shape back in 1973,[12] when only three quark flavors were known experimentally.

but this is not topic of this thesis. The current best experimental values for the magnitudes of all CKM matrix elements, obtained by combining the results of a large variety of experiments in a global fit, including theoretical constraints, are:[13]

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

One observes that the CKM matrix has a strongly hierarchical structure. The entries on the diagonal are close to one, while the off-diagonal elements are generally much smaller. The physical interpretation is that mixing in general is small: as mentioned, couplings within the same quark generation are highly preferred over couplings between the generations. Furthermore the individual mixings are again clearly ordered. The mixing between the first and second generation is much larger than the mixing between the second and third generation, which is again much larger than the mixing between the first and third generation.

1.3.4 Confinement and asymptotic freedom in the strong interaction

Due to the non-abelian structure of the $SU(3)_C$ gauge group in conjunction with its gauge bosons, the gluons, being massless, some intriguing features of the strong interaction arise. The non-abelian structure of $SU(3)_C$ implies that gluons themselves carry color charge, which enables them to undergo self-interactions, i.e. gluons can couple to gluons. This leads to the unique feature of QCD that there is "antiscreening", i.e. the force between two color charges *increases* when they are moved apart, in contrast to intuition and in contrast to e.g. the electric force, which gets smaller at longer distances. In principle in the weak force, there is self-interaction, too. But, unlike the gluons in QCD, here the gauge bosons are massive, making the weak interaction inherently short-ranged, while in the electromagnetic interaction the gauge bosons (the photons) are massless but don't undergo self-interaction, as they don't carry charge.

In terms of the running coupling this antiscreening effect implies that the coupling strength of QCD increases at low energies (and long distances correspondingly), while it decreases at high energies (and small distances correspondingly). This has two main consequences. Firstly, due the increasing coupling strength the energy that is stored in the color field between two colored particles increases rapidly with distance. At some point this energy gets high enough to create two new (colored) particles from the vacuum between the two original particles. As this reduces the distances between the color charges, this configuration is energetically preferable, even though some energy went into the masses of the newly generated particles. The final consequence of this effect is that colored particles (like quarks) can never exist freely, but only within color-neutral composite particles. This phenomenon is called *confinement*. The resulting color-neutral particles are called hadrons. In a simple picture, hadrons consist of either a quark and an antiquark (mesons) or three quarks or three antiquarks (baryons). The process of hadron formation, that takes place whenever colored particles are produced in some interaction, is called hadronization. Due to the large coupling strength that drives hadronization, as well as any other long-distance/low-energy QCD processes, these cannot be calculated perturbatively, but rather models must be employed in order to describe them.

At small distances/high energies, where the strong coupling strength gets weak, the opposite phenomenon arises. As a consequence of the small coupling at these distances, quarks can move almost freely within hadrons. It allows in that some sense hadrons can be described as a bag of quasi-free quarks and gluons. This phenomenon is known as *asymptotic freedom*.

1.4 Calculating predictions for hadron colliders

In this section it is discussed how predictions for hadron collider experiments are calculated. First the concept of *factorization* is explained, which is the general underlying concept based on which such predictions are calculated. Then the simulation of physics processes using Monte Carlo event generators is discussed.

1.4.1 The factorization approach

One of the main consequences of asymptotic freedom (see Section 1.3.4) is that the inelastic scattering processes between protons can be described as the scattering between their constituents, i.e. either quarks or gluons, which are collectively called *partons* in the following. In the simplified view of a proton as a bag of quasi-free partons, a collision between two protons looks like two bags of quasi-free partons that are thrown at each other, where one parton from one bag hits one parton from the other bag. This parton-parton collision is denoted as the *hard scattering/collision/interaction*. The cross section for the whole process can then be described as the cross section for the hard scattering process between the two partons, multiplied by the probability that exactly these two partons are the ones that interact.

The precise theoretical manifestation of this idea is called *factorization*. The theoretical calculation of any cross section at proton colliders is based on this approach. For this purpose, the momentum of a parton that takes part in the hard scattering is denoted by its fraction, x, of the full proton momentum:

$$x = \frac{p_{\text{parton}}}{p_{\text{proton}}} .$$

The squared center-of-mass energy, $\hat{s}$, of the hard scattering between two partons with momentum fractions x_1 and x_2 is given by

$$\hat{s} = x_1 x_2 s$$

for a collision of two protons with center-of-mass energy $\sqrt{s}$. Factorization then states that the cross section to produce a final state X in the collision of two protons at a center-of-mass energy $\sqrt{s}$ can be expressed as

$$\sigma(p_1 p_2 \to X) = \sum_{i,j=q,\bar{q},g} \int \mathrm{d}x_i \mathrm{d}x_j \, f_{i/p_1}(x_i, \mu^2) f_{j/p_2}(x_j, \mu^2) \cdot \hat{\sigma}_{ij}(ij \to X; \hat{s}, \mu^2) . \tag{1.2}$$

In this equation, $f_{i/p_k}(x_i, \mu^2)$ is the *parton density function* (PDF) of the proton, denoting the probability that a parton of type i (where type means either gluon or a particular quark flavor) from proton k and with a momentum fraction x_i takes part in the hard scattering; and $\hat{\sigma}_{ij}(ij \to X; \hat{s}, \mu^2)$ is the partonic cross section to produce the final state X in the collision of two partons i and j with a squared partonic center-of-mass energy of $\hat{s} = x_i x_j s$. The product of the PDFs for each proton and the partonic cross section is integrated over all possible momentum fractions x_i and x_j. This integral is calculated for all possible combinations of parton types that may produce the final state X and the sum of all these combinations then yields the full cross section.

The only quantity in equation 1.2 that was not explained yet is the *factorization scale* μ. In order to understand this concept, the simplified picture of bags of quasi-free partons must be refined a bit. The interactions of the partons inside the proton, which have been neglected so far, must now enter the picture. While factorization means that the interactions inside the protons can be decoupled from the hard scattering process, it is in fact not unique how and where to draw the line between them. Instead,

similar to renormalization (see Section 1.2.4), a particular factorization scheme and a factorization scale must be chosen, where the scheme precisely defines the "how" and the scale defines the "where" to draw the line. The factorization scheme defines e.g. which quark flavors are considered to be contained in the proton and therefore have a PDF and which must be produced in the hard scattering. The factorization scale can be regarded as a resolution parameter that determines how much substructure of the proton is resolved and can therefore become part of the hard scattering. Like the renormalization scale it is an arbitrary scale, which no physical quantity should depend on, while any finite order predictions generally do. The factorization scale is popularly chosen equal to the renormalization scale, and like the latter it is typically one of the main sources of uncertainties of theoretical predictions, supposed to decrease toward higher orders. An exemplary situation in which both the factorization scheme and scale play a role is discussed in Section 3.2.2.

Parton density functions

As mentioned, PDFs denote the probability to find a certain parton with a certain momentum fraction inside the proton. The PDFs are universal properties of the proton. However, they cannot be calculated from theory, but must be measured experimentally. A variety of different experiments has performed measurements of PDFs, and several collaborations exist, such as the CTEQ or the MRST/MSTW, that regularly perform global fits to these experimental data in order to extract sets of PDFs that can then be used for cross-section calculations.

Parton luminosities. Instead of integrating over x_i and x_j independently in equation 1.2, one can perform one integral over all combinations of x_i and x_j for which $\hat{s} = x_i x_j s$ is constant, and the other integral over $\hat{s}$. This is done by substituting $x_j = \hat{s}/x_i s$. As a consequence the integral over x_i becomes independent of the partonic cross section. One obtains:

$$\begin{aligned}\sigma(p_1 p_2 \to X) &= \sum_{i,j=q,\bar{q},g} \int \mathrm{d}x_i \mathrm{d}x_j \; f_{i/p_1}(x_i,\mu^2) f_{j/p_2}(x_j,\mu^2) \cdot \hat{\sigma}_{ij}(ij \to X;\hat{s},\mu^2) \\ &= \sum_{i,j=q,\bar{q},g} \int \mathrm{d}x_i \frac{\mathrm{d}\hat{s}}{x_i s} \; f_{i/p_1}(x_i,\mu^2) f_{j/p_2}(\tfrac{\hat{s}}{x_i s},\mu^2) \cdot \hat{\sigma}_{ij}(ij \to X;\hat{s},\mu^2) \\ &= \sum_{i,j=q,\bar{q},g} \int \mathrm{d}\hat{s} \; L_{ij}(\hat{s};s,\mu^2) \cdot \hat{\sigma}_{ij}(ij \to X;\hat{s},\mu^2)\,.\end{aligned}$$

In the last line, the *parton luminosity* $L_{ij}(\hat{s};s,\mu^2)$ has been introduced, which is given by:

$$L_{ij}(\hat{s};s,\mu^2) = \frac{1}{s} \int \frac{\mathrm{d}x_i}{x_i} f_{i/p_1}(x_i,\mu^2) f_{j/p_2}(\tfrac{\hat{s}}{x_i s},\mu^2)\,.$$

The parton luminosity is a measure for the probability that the hard collision takes place between two partons of types i and j at a partonic center-of-mass energy of $\sqrt{\hat{s}} = x_i x_j \sqrt{s}$. It is a useful quantity in order to compare the same production process at different colliders. An exemplary situation is discussed in Section 3.2.1

The partonic cross section

The second ingredient that is needed besides the PDFs in order to calculate cross sections for hadron collisions is the partonic cross section $\hat{\sigma}_{ij}(ij \to X;\hat{s},\mu^2)$. It is obtained according to equation 1.1, the

necessary matrix element is calculated perturbatively using Feynman diagrams as explained in Section 1.2.3. When precise predictions are of interest, higher-order calculations (i.e. at NLO or beyond) in QCD are required. In the following, some concepts and issues that occur in such calculations are discussed.

Higher-order calculations in QCD. Higher-order corrections in QCD have the problem that they give rise to different types of singularities that need to be regularized. Regularization can happen in different ways, depending on the process and the type of singularity. Some singularities that are related to gluons emissions from initial-state quarks are simply absorbed in the PDFs. As a result, PDFs become dependent on the order of the calculation as well (i.e. different PDF sets are needed for LO, NLO etc.). In other situations, different types of singularities just cancel each other.

It may happen that after the regularization of the occurring singularities some large logarithmic terms remain, which lead to large theoretical uncertainties of the calculation. However, these terms are often universal in the sense that at each order the same term occurs with a higher power; in this case it is possible to just sum them up to all orders. This technique is called resummation, and the precision of such calculations is called "Leading Logarithm"(LL), "Next-to-Leading Logarithm"(NLL) etc. depending on the order of the calculation the resummation is applied to.

1.4.2 Simulation of physics processes

The simulation of physics processes is a crucial tool for data analysis. For the generation of simulated events, first the actual collision is simulated by a Monte Carlo (MC) event generator, which provides a list of particles that have been produced in the simulated collision as output. This is called event generation. The simulated collisions are then run through a detector simulation, which models the response of the detector to the generated particles. After that, the same reconstruction can be applied that is also used for real data, such that in the end, the simulated events are available in the same data format as real events. For a data analysis, all physics processes that are supposed to be relevant after event selection are simulated, added up and compared to real data. From this comparison one can then extract how many signal events are contained in the real data sample. In the following some selected aspects of event generation are sketched.

Event generation using Monte Carlo generators

Like the plain calculation of cross sections, also MC generators make use of the factorization theorem. Using a set of PDFs and the matrix element – usually at LO – for the process that is to be simulated, they first create a raw event that contains only the immediate final-state particles of the process. This does not provide a realistic description yet; further steps are necessary. In any processes that involve colored particles, as is always the case at hadron colliders, these particles may radiate gluons at any time, gluons may also split into a quark-antiquark pair. For the simulation of these processes, so-called *parton shower* models are employed. Furthermore, not only the actual hard scattering may take place in a proton collision, but additional, softer interactions between the remaining partons can occur as well. This is called *multi-parton interactions*, which are simulated using appropriate models as well. Finally, all colored particles that are in the "final" state at that stage need to be hadronized. For this purpose again dedicated models are employed.

Various different MC generators are on the market. Except for the PDFs, which are not considered part of the generator, they may differ in any of the discussed aspects. Every generator has its particular strengths and weaknesses, therefore usually different generators are used for the simulation of different processes even within one single data analysis. MC generators are generally modular regarding the

different steps of event generation. There are even specialized programs that only perform one single step of the simulation chain and must always be interfaced to other generators. The particular generators that were used in the scope of this thesis are listed in Section 5.1.2.

Event generation at NLO and parton shower matching. As mentioned, usually LO matrix elements are used in MC generators. The reason is that NLO matrix elements include the radiation of an additional gluon, which is also described by the parton shower. In order avoid double counting, i.e. to remove this overlap, the parton shower must be carefully matched to the NLO matrix element. As a result of this matching, negative weights may occur for some events, i.e. when counting the number of events in a sample these events must be counted as -1 (for more information about event weights see Section 5.1.2). Relevant examples in the scope of this thesis are the programs MC@NLO [14] and AcerMC.[15] MC@NLO was the first NLO order generator that became available. It produces up to 10%-15% negative-weight events. AcerMC is actually a leading order generator. It only performs parton shower matching for initial state b quarks that implicitly arise from a gluon splitting inside the proton. Therefore it only produces negative-weight events for some particular processes.

1.5 Experimental basics

This section introduces some experimental concepts that are commonly used in particle physics. First some typical kinematic variables that are used at hadron colliders are described. Then jets and b-tagging are introduced.

1.5.1 Kinematics at hadron colliders

The momentum fractions x_1 and x_2 of the partons that take part in a collision are generally not known. In particular, x_1 and x_2 are normally not equal. This implies that the center-of-mass system of the hard scattering generally is not at rest in the laboratory. Instead, it has some momentum along the beam axis (i.e. the proton–proton flight axis). Moreover, the actual momentum of the center-of-mass system of the hard collision in the laboratory, in the following called the *boost* of the hard collision, is generally not known. This has some impact on the kinematic variables that are used to describe hadron collisions.

Rapidity and pseudorapidity. One consequence of the unknown boost of the hard collision is that the same hard collision may look very different in the laboratory depending on its boost. Concretely, the polar angles of the final-state particles, i.e. their angles with respect to the beam axis, and thus also the angular distances between them, will be different. A useful concept to deal with this problem is the *rapidity* y of particle, which is defined as

$$y = \frac{1}{2} \ln\left(\frac{E + p_{\mathrm{L}}}{E - p_{\mathrm{L}}}\right),$$

where E is the energy of the particle and p_{L} its longitudinal momentum, i.e. the component of its momentum that points along the beam axis. The rapidity has the advantage that rapidity differences between particles do not depend on the boost of the center-of-mass system of the hard collision, i.e. they are invariant under boosts along the beam axis. When the energy of a particle is much greater than its mass ("high-energy" limit), such that $|\vec{p}| \approx E$, the rapidity can be approximated by the *pseudorapidity*,

η, which is defined as

$$\eta = \frac{1}{2} \ln\left(\frac{|\vec{p}| + p_{\mathrm{L}}}{|\vec{p}| - p_{\mathrm{L}}}\right) = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right),$$

where θ denotes the polar angle of the particle. The advantage of the pseudorapidity is that it is only a function the polar angle, while for the rapidity the particle's mass is needed. With help the pseudorapidity, a boost-independent distance measure between two particles, ΔR, can be defined, which is usually used at hadron colliders instead of the angular distance between particles:

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

where $\Delta\eta$ is the pseudorapidity difference between the particles and $\Delta\phi$ is the azimuthal difference, i.e the angular difference around the beam axis.

Missing transverse momentum. The second consequence of the unknown boost of the hard collision has to do with particles that escape the detector without being detected, such as neutrinos. If the boost was known, momentum conservation could be used to infer at least the sum of the momenta of all such particles that occurred, by demanding that the sum of the momenta of all final-state particles must add up to the initial boost. Any measured momentum imbalance would then be an indication that one or more particles escaped undetected, and the precise measured deviation would serve as the estimate for the sum of their momenta. At hadron colliders this concept must instead be modified. What can still be exploited here is that the center-of-mass system of the hard collision has a longitudinal momentum, but no transverse momentum (i.e. momentum component perpendicular to the beam axis). Any transverse momentum imbalance can therefore still be used in order to estimate at least the transverse momentum sum of escaping particles. This imbalance is called *missing transverse momentum* ($p_{\mathrm{T}}^{\mathrm{miss}}$).[6]

1.5.2 Jets

As mentioned in Section 1.3.4, colored particles, such as quarks and gluons, cannot exist freely but are confined in hadrons. When a colored object is produced in a particle collision it therefore hadronizes almost immediately, after about 10^{-23} s. This hadronization process normally gives rise to not only one single hadron that emerges from the quark or gluon, but to a collimated spray of several particles, which can then be measured in the detector. This is called a jet. In simple events, these can be easily identified by eye (aside from the fact that in practice this is not feasible due to the way too high event rates at colliders).

Jet algorithms

In more complicated events like they are usually produced at hadron colliders, jet identification by eye breaks down. Furthermore, in order to enable theoretical predictions for measurements that involve jets, a precise definition of what is called a jet must be constructed. Such definitions are provided by *jet algorithms*. Out of a set of four-momenta that represent final-state particles they pick subsets of particles

[6] In literature one also frequently finds the term "missing transverse *energy*", $E_{\mathrm{T}}^{\mathrm{miss}}$, as in practice mostly energy measurements of final-state particles are used to calculate it.

that are grouped together into jets. The four-momentum of a jet is then given by the sum of the four-momenta of its constituents. The input sets of four-momenta to a jet algorithm are given by measured particles (normally in the calorimeter) in case of real data; for simulated data also simulated particles at different stages of their decay chain may be used. It is desirable for a jet algorithm that the resulting jets do not depend on which stage of the decay chain is used. In particular certain types of gluon emissions by hadrons are supposed not to make a difference. This ensures that consistent theoretical predictions for jet final states can be made. Jet algorithms that fulfill these requirements are called *infrared-safe* and *collinear-safe*.

The most frequently used type of jet algorithms to date are the so-called sequential recombination algorithms. They define a distance measure for any pair of four-vectors, that contains both the geometrical distance between them and their energies. Furthermore for any arbitrary single four-vector, a cut-off distance is defined, that also depends on its energy. From the list of input four-vectors to the algorithm, all pairwise distances as well as all individual cut-off distances are calculated. In case the minimum of all these values is one of the distances, the corresponding pair of four-vectors is combined into one new four-vector that replaces its two constituents in the list. In case the minimum is rather one of the cut-off distances, the corresponding four-vector is called a jet and removed from the list. This is repeated until the list is empty. Sequential recombination algorithms contain a size parameter or resolution parameter, R, which is part of the distance measure and this way defines the typical sizes of the jets.

As an example for such an algorithm, the so-called Anti-k_T algorithm [16] is presented here. In this algorithm, the distance measure, d_{ij}, between two particles and the cut-off distance, d_i, for an individual particle are given by

$$\begin{aligned} d_{ij} &= \min(p_{\mathrm{T},i}^{-2}, p_{\mathrm{T},j}^{-2}) \frac{\Delta y_{ij}^2 + \Delta\phi_{ij}^2}{R^2} \\ d_i &= p_{\mathrm{T},i}^{-2}, \end{aligned}$$

where $p_{\mathrm{T},i}$ is the transverse momentum of particle i, Δy_{ij} and $\Delta\phi_{ij}$ are the rapidity and the azimuthal difference between two particles i and j and R is the resolution parameter. A unique feature of the Anti-k_T algorithm among the sequential recombination algorithms is that it produces very smooth, cone-shaped jets. Due to this property, which is quite attractive to experimentalists, the Anti-k_T algorithm is nowadays the default jet algorithm in many experimental collaborations in particle physics.

1.5.3 b-quark jets and b-tagging

In general, jets that originate from quarks of different flavors cannot be distinguished. The major exception is b-quark jets. They can be identified due to the comparably long lifetime of B hadrons (i.e. the hadrons that are formed by b quarks), which stand at the beginning of every b-quark jet. The lifetime of B hadrons is about 10^{-12} s, resulting in typical flight paths of up to a few millimeters before they decay. As a consequence, the B-hadron decay vertex, which the trajectories of the jet constituents point back to, is displaced with respect to the primary vertex of the event, at which the actual hard collision took place and to which the trajectories of other particles in the event point back. With modern tracking detectors it is possible to reconstruct such displaced secondary vertices and thus identify a jet as a b-quark jet. In practice various different techniques are employed to exploit different variables related to the secondary vertices. Identification of jets as b-quark jets is called *b-tagging*; and algorithms that serve that serve this purpose are called *b-taggers*. The typical output of a b-tagger is a continuous variable, sometimes called the b-tagger *weight*, that can be calculated for each jet and that is supposed to have high values for b-quark jets and low values for light-quark (i.e. non-b-quark) jets.

Chapter 2

ATLAS and the LHC

2.1 The Large Hadron Collider (LHC)

The LHC and its design parameters. The LHC [17, 18][1] is a circular proton–proton collider with a circumference of 26.659 km, running about 100 m underground. It is part of the facilities of the European Organization for Nuclear Research (CERN[2]), close to Geneva, Switzerland. The LHC accelerates protons in two counter-rotating beams and brings them to collision at four points. At design operation, collisions will happen at a proton–proton center-of-mass energy of 14 TeV. Each beam will consist of 2808 bunches of $1.15 \cdot 10^{11}$ protons each. The bunch spacing in each beam, and hence the bunch crossing rate, will be 25 ns, achieving an instantaneous luminosity of 10^{34} cm^{-2} s^{-1}. At this luminosity not only one but 23 collisions will happen per bunch crossing on average.[19] These additional interactions are called *pileup*. Besides proton–proton collisions, the LHC is also capable of colliding heavy ions, particularly lead nuclei.
Figure 2.1 [20] shows an illustration of the LHC and its hinterland as well as the location of the four major detectors. Also shown is the SPS, the last constituent of the LHC injector chain, which injects protons into the LHC at an energy of 450 GeV. Details of the injector chain are given e.g. in [17, 18].

Experiments at the LHC. The two LHC beams are brought to collision at four interaction points, at which the four large detectors/experiments[3] are located. Two of them, LHCb [21] and ALICE,[22] are special-purpose detectors: LHCb aims mostly at studying the physics of B-hadrons, with a particular focus on CP-violation. It is the only one of the four large detectors that that does not exhibit the typical collider detector layout with almost hermetic coverage of the collision point, but rather has a fixed-target detector geometry. The ALICE experiment is mostly dedicated to studying quark-gluon plasma, which is expected to be produced in heavy ion collisions. The other two detectors, ATLAS [19, 23] and CMS,[24] are general purpose detectors, aimed at measuring as many final states as possible. In addition to these four major experiments, there are three minor ones. The TOTEM experiment,[25] consisting of several components located up to 220 m away from the CMS interaction point, is designed to measure the total proton–proton cross section as well as elastic scattering and diffractive dissociation. The LHCf experiment,[26] located 140 m away from the ATLAS collision point, measures neutral particles produced at very low angles, which is hoped to provide input for the modeling of cosmic-ray-induced particle showers in the atmosphere. The MoEDAL experiment [27] is the youngest experiment at the LHC; it was approved only in 2010. Located around the LHCb interaction point, covering the hemisphere that is not occupied by LHCb itself, it will look for magnetic monopoles and other heavy, stable

[1] All numbers in this paragraph taken from there unless stated otherwise.
[2] In fact, nowadays CERN is rather a particle physics than a nuclear physics laboratory.
[3] The terms "experiment" and "detector" are frequently used synonymously in this context. Strictly speaking the names refer to the collaborations that operate the detectors of the same name and use them to conduct multiple experiments.

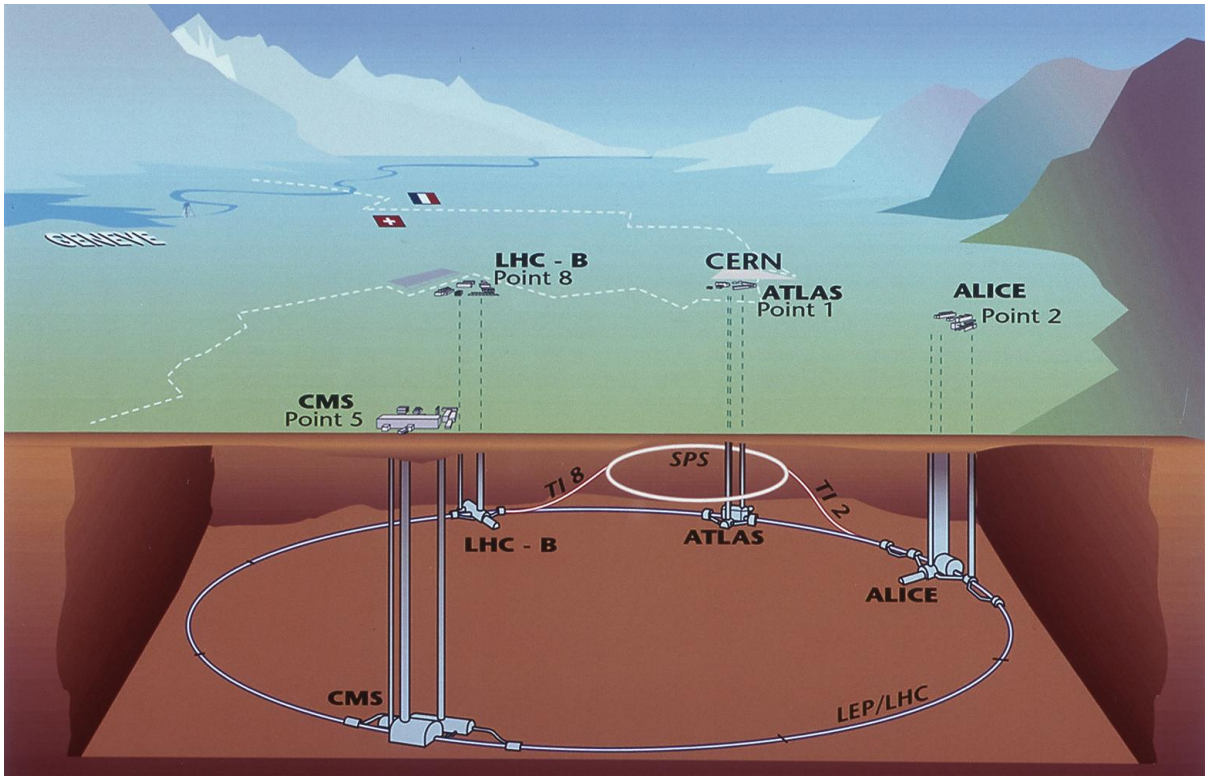


Figure 2.1: The LHC with the four major experiments and its situation in the hinterland.

or meta-stable, highly-ionizing particles. So far only a test setup of MoEDAL is installed, the full installation of the detector is foreseen for 2013.

Operation so far. LHC physics operation started in November 2009 with a short running period at a proton–proton center-of-mass energy ($\sqrt{s}$) of 900 GeV and very low instantaneous luminosity, as well as few runs at $\sqrt{s} = 2.36$ TeV. Most of 2010 was devoted to running at 7 TeV and to increasing the instantaneous luminosity by five orders of magnitude to $2 \cdot 10^{32}$ cm^{-2} s^{-1} over the course of the year. In 2011 the LHC reached an instantaneous luminosity of up to $3.65 \cdot 10^{33}$ cm^{-2} s^{-1} at a bunch spacing of down to 50 ns. The pileup amounted to up to 17 interactions per bunch crossing on average. The full 2011 data-sets amounts to about 5 fb^{-1}, still at a center-of-mass energy of 7 TeV. In 2012 the LHC is expected to reach its design luminosity. After that a long shutdown of more than one year is foreseen, during which some upgrades will be performed in order to also push it to its design energy. For this thesis the first inverse femtobarn of 2011 data was used.

2.2 The ATLAS detector – A Toroidal LHC ApparatuS

General-purpose detectors at particle colliders, such as the ATLAS detector, are designed to enable studies of preferably any final state that may occur. For this purpose they need to be capable of detecting as many of the particles that are produced as possible, measure their momentum and energy as precisely as possible and identify the particle types of as many of them as possible. These requirements drive the typical design characteristics of such detectors:

- almost hermetic coverage of the full solid angle in order to detect the full final state;
- components arranged in a shell structure, comprising (outward from the center):
  - a tracking system in the detector center, immersed in the magnetic field of a large solenoid coil, precisely measuring the flight path and the momentum of charged particles;
  - a calorimeter system measuring particles' energies (thereby stopping and absorbing them entirely), which is divided into one part that is optimized for electromagnetic showers and one part that is optimized for hadronic showers;
  - a muon system, exploiting the assumption that any detectable particles that are not stopped in the calorimeter are mostly muons.

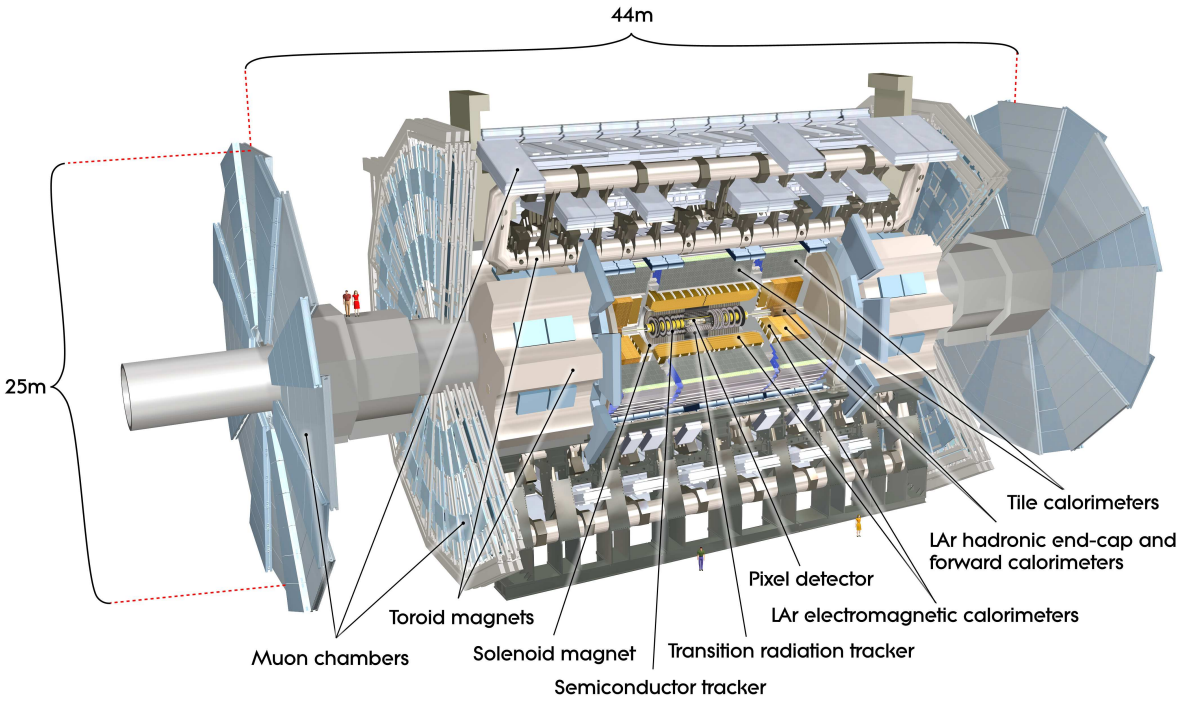


Figure 2.2: Overview of the ATLAS detector.

This general scheme also applies to the ATLAS detector. Figure 2.2 shows its overall design. The ATLAS central tracker is called the Inner Detector. It comprises a silicon pixel detector, a silicon strip detector and a straw tube detector. The solenoid coil is located between the Inner Detector and the calorimeter and produces a magnetic field strength of 2 T at the interaction point. The electromagnetic calorimeter uses liquid argon (LAr) technology with lead absorbers. The hadron calorimeter is divided up into the so-called tile Calorimeter in the central and semi-central region, which is a steel/scintillator sampling calorimeter, and the hadron end-cap, which uses liquid argon with copper absorbers. In addition there is the Forward Calorimeter, located very close to the beam-pipe, mostly inside the hadron end-cap. It uses LAr along with copper and tungsten absorbers in its electromagnetic and hadronic part, respectively. The most distinct feature of the ATLAS detector is the huge air-core toroid magnet system with a long barrel and two end-caps, which is arranged around the calorimeter. Its purpose is to enable the muon system to perform precise momentum measurements on its own, independent of the

Inner Detector. The muon system is located right before and behind the toroid, in the barrel also inside and between the racetrack-shaped coils. It consists of four different types of tracking chambers: two types with a very fast response that are suitable for triggering, and two types that are suitable to perform precision measurements, providing good momentum resolution. Driven by the size of the toroid that it has to contain, the muon system gives rise to the large overall dimensions of the ATLAS detector, measuring about 44 m in length in 25 m in diameter, while it weighs "only" 7000 t, compared to CMS, which weighs 12500 t having only half the size.

A detailed description of all components of the ATLAS detector is given in [19, 23]. In the following only a brief description of the basic components is given, mostly summarizing information from there. In particular all figures of Section 2.2 are taken from [23].

Coordinate system

The ATLAS coordinate system is a standard right-handed and -angled coordinate system with the x axis pointing toward the LHC center, the y axis pointing upward and the z axis pointing along the beam pipe. The origin is located at the nominal interaction point. The polar angle, θ, is the angle with respect to the z axis and the azimuthal angle, ϕ, is measured around the z axis as usual. The pseudorapidity, η, is defined as $\eta = -\ln(\tan(\theta/2))$. The *transverse* plane denotes the xy plane, the distance from the z axis in the transverse plane is usually denoted as r. Distances between two particles are normally expressed in $\eta\phi$ space via $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

2.2.1 The Inner Detector

The Inner Detector (ID) comprises three sub-components: the Pixel Detector (Pixel), the SemiConductor Tracker (SCT) and the Transition Radiation Tracker (TRT), each consisting of a barrel part and two end-caps. Pixel and SCT together are also referred to as the "silicon" trackers. Figures 2.3, 2.4, 2.5 and 2.6 provide an overview of the layout of the ID and its dimensions, amounting to about 6 m in length and 2 m in diameter. The ID covers a pseudorapidity region of $|\eta| < 2.5$.

The Pixel

The Pixel is a silicon pixel detector. It consists of three cylindrical barrel layers and three discs in each end-cap, such that each track typically crosses three Pixel layers. The sizes and positions of the individual layers and disks can be read off from figures 2.4, 2.5 and 2.6. The pixel employs only one type of sensor for both the barrel and the end-cap. Each sensor is 19×63 mm^2 in size and contains 47232 pixels, with a nominal pixel size of 50×400 μm^2. In the barrel, 13 modules – each carrying one sensor – form a stave. The three cylindrical barrel layers are composed of 22, 38 and 52 of these staves, respectively. Figure 2.7 (left) shows a bi-stave (a mechanical substructure of the barrel cylinders) loaded with modules. The six end-cap disks are all identical. Each disk is made up of eight azimuthal sectors with three modules arranged radially on each side, such that together they cover the full azimuthal range of the sector. A picture of a sector is shown in figure 2.7 (right). The Pixel achieves an intrinsic accuracy of 10 μm in $r\phi$ and an intrinsic z/r resolution of 115 μm in the barrel/end-cap, respectively. The Pixel covers the full ID pseudorapidity range of $|\eta| < 2.5$.

The SCT

The SCT is a silicon strip detector. It has four cylindrical layers in the barrel and nine discs in each end-cap. The positions and sizes of the end-cap discs are designed such that any particle typically crosses

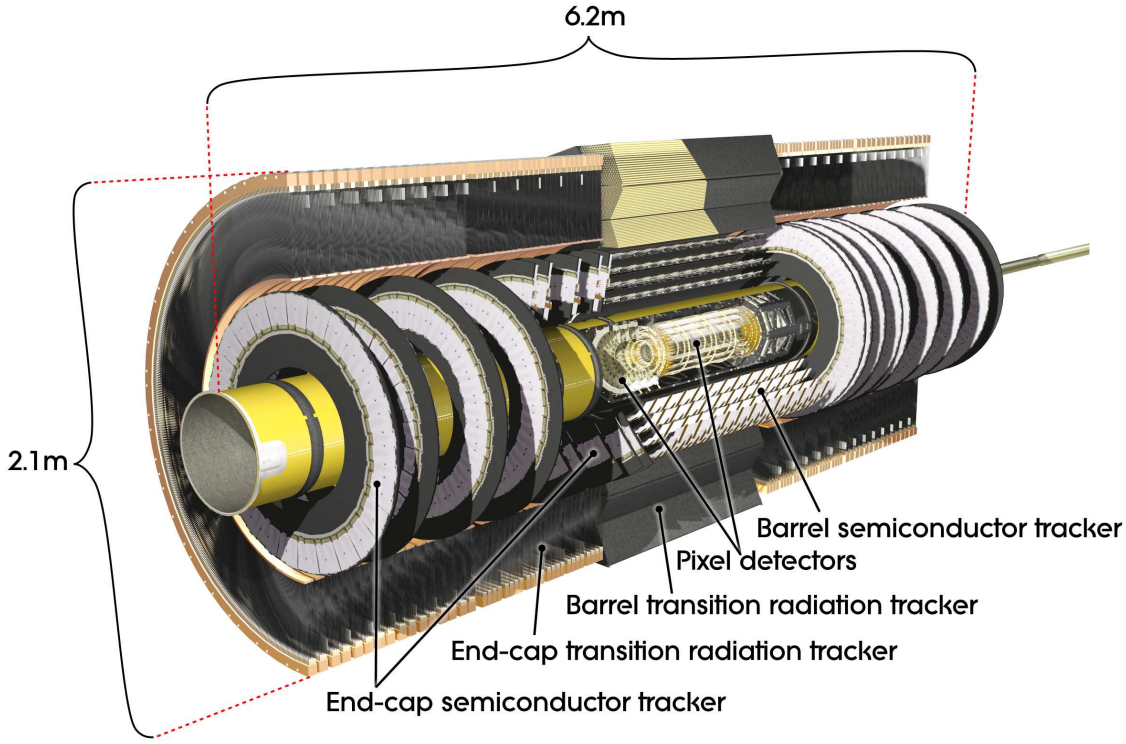


Figure 2.3: Cutaway view of the ATLAS Inner Detector.

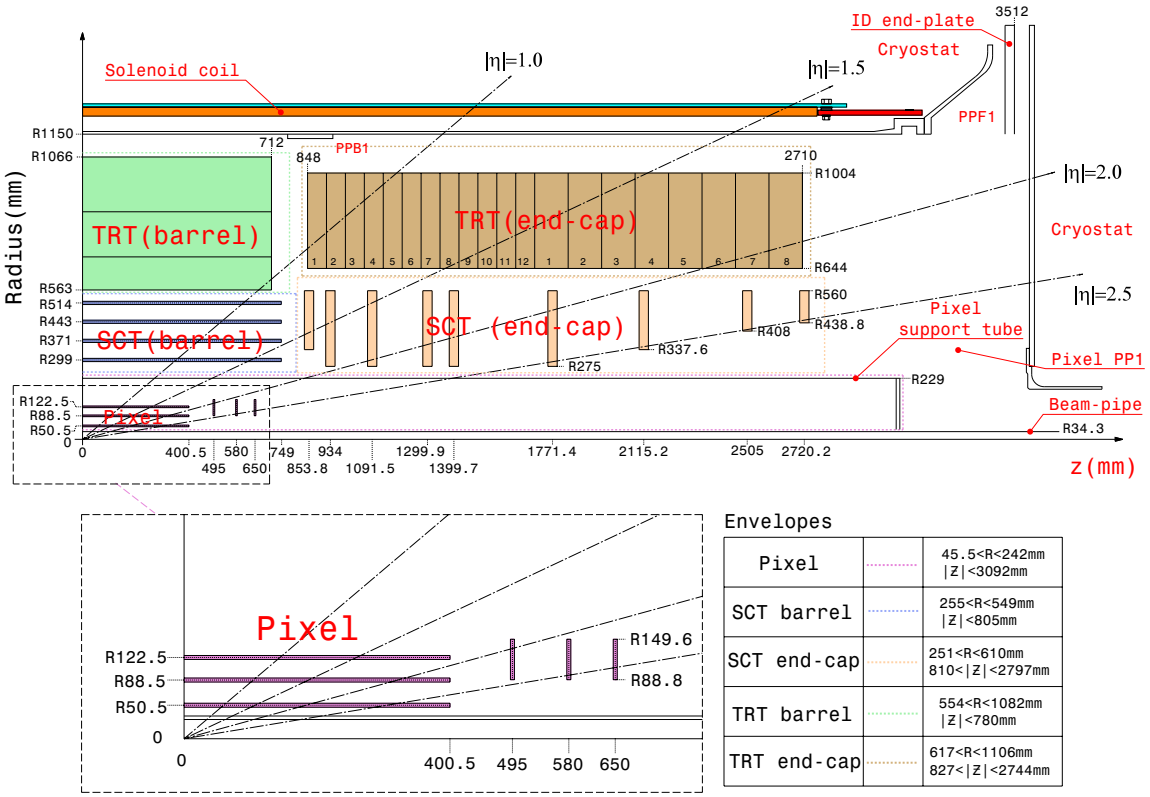


Figure 2.4: Plan view of a quarter-section of the Inner Detector.

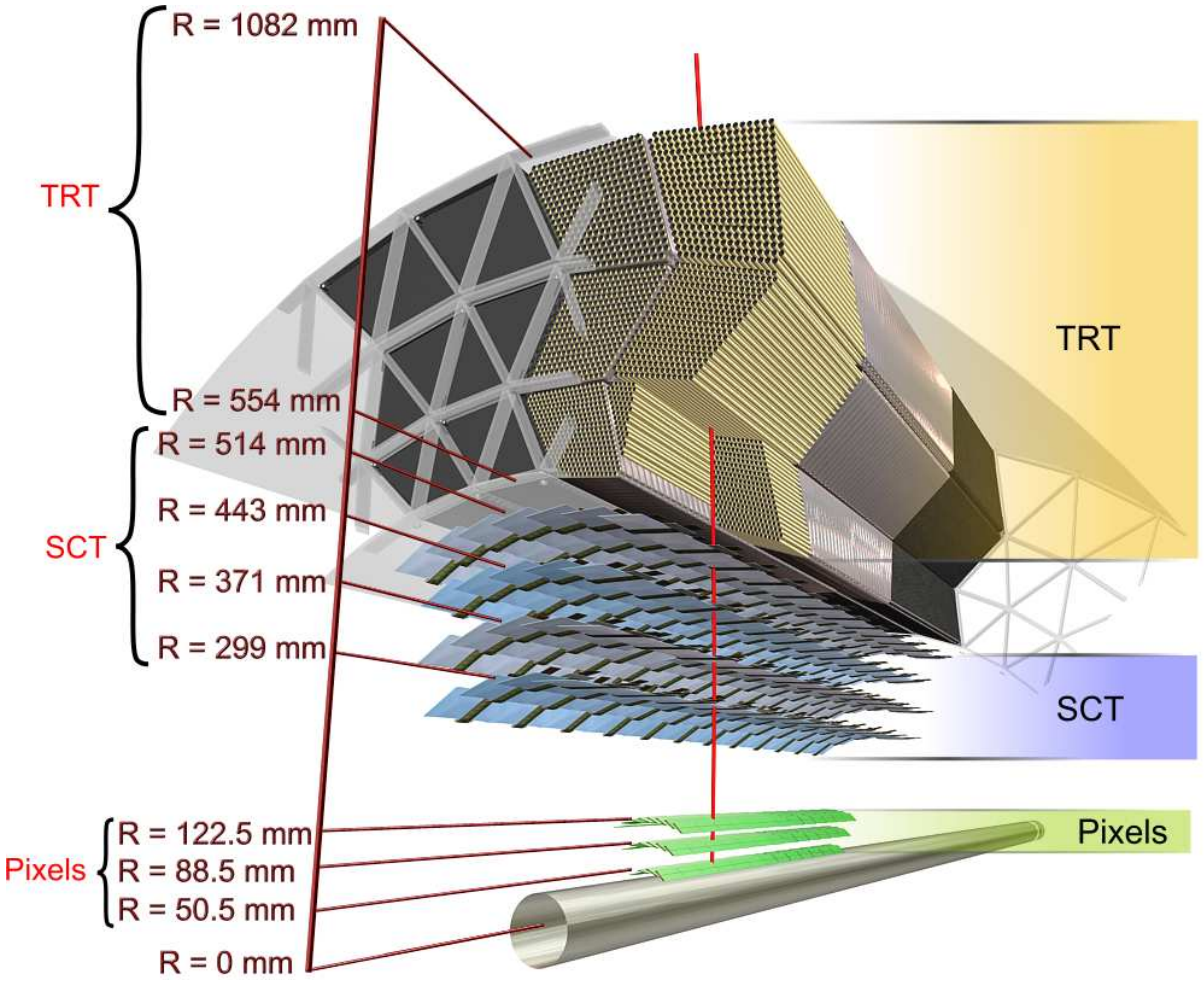


Figure 2.5: Three-dimensional overview of the Inner Detector barrel.

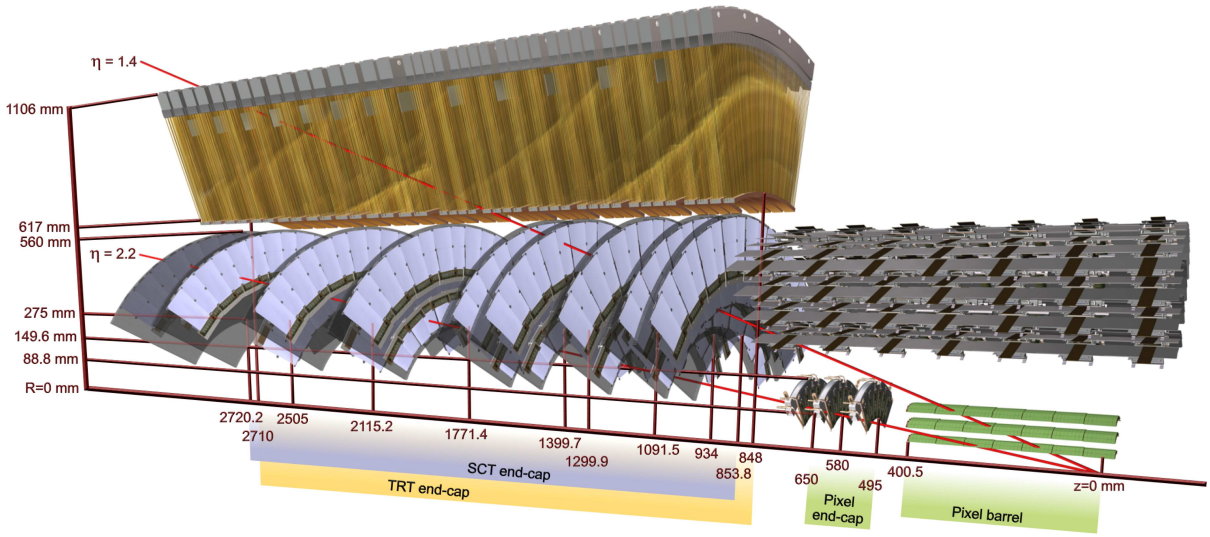


Figure 2.6: Three-dimensional overview of an Inner Detector end-cap (and part of the barrel).

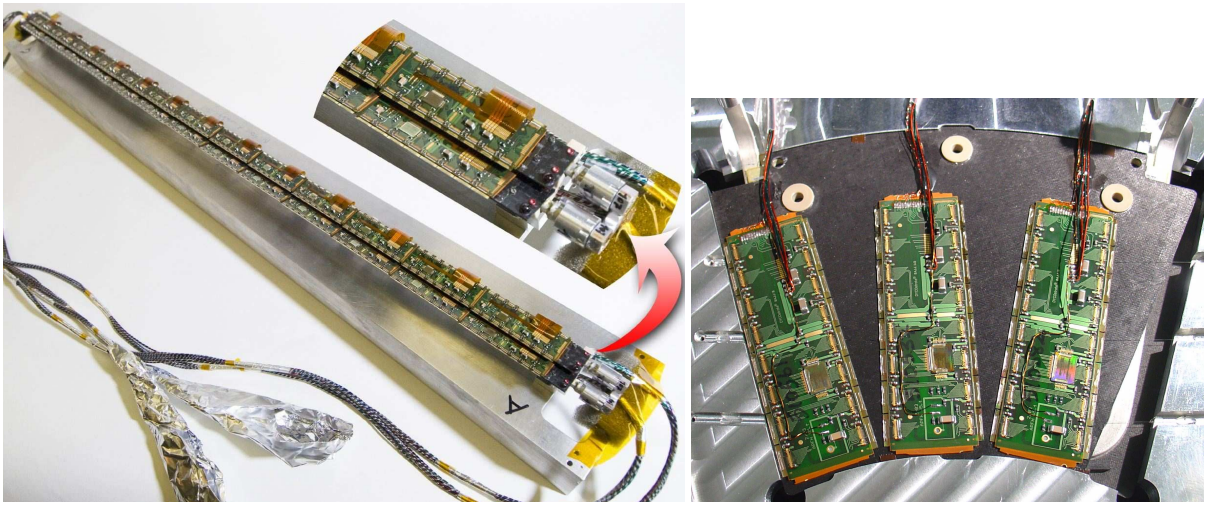


Figure 2.7: Photograph of a Pixel barrel bi-stave (left) and an end-cap sector (right). Three more modules are on the back of the sector, filling the azimuthal gaps.

four SCT layers, independent of its pseudorapidity. The exact sizes and positions of all layers and disks can be seen in figures 2.4, 2.5 and 2.6. In both the barrel and the end-cap the modules have sensors on both sides, hence each crossed SCT layer typically provides two SCT hits. The sensors on both sides of each module are rotated against each other by 40 mrad in order to provide the required resolution in z (barrel) and in r (end-caps). In the barrel, two rectangular sensors with a strip pitch of 80 µm are connected together on each side of each module, giving an active length of 12.6 cm. The sensors are individually mounted in rows of twelve to form the barrel cylinders, such that the strips on one side of each module are parallel to the z axis. In the end-caps, three different types (outer, middle and inner) of wedge-shaped modules are used. The strips on the sensors are oriented radially with a mean pitch of 80 µm. Like in the barrel, also the end-cap modules are individually mounted directly to the support structure. Photographs of all four sensor types are shown in figure 2.8. The SCT achieves an intrinsic accuracy of 17 µm in $r\phi$ and 580 µm in z/r (barrel/end-cap), respectively. Like the Pixel, the SCT covers the full ID acceptance range of $|\eta| < 2.5$.

The TRT

The TRT consists of 4 mm diameter straw-like polyimide tubes, each acting as an individual drift chamber with the signal wire spanned along the center of the straw. The barrel is divided into 3 rings of 32 modules each. The modules have triangular cross section in the transverse plane, the straws are arranged parallel to the z axis. In total the TRT barrel contains 73 layers of straws. The straws range from $z = -71.2$ cm to $z = 71.2$ cm, but are electrically split near their center (i.e. $z = 0$) and read out from both sides, reducing the occupancy and providing a minimal z measurement ($>$ or $<$ 0). Apart from that, the TRT barrel does not provide any z information but measures only in the transverse plane. In figure 2.5 the layout of the TRT barrel can be perceived well. Each TRT end-cap consists of 160 layers of radially oriented straws with a length of 37 cm, distributed uniformly in ϕ. These layers are grouped to 20 wheels of 8 layers each, where the first 12 and the last 8 wheels have a 8 mm and 15 mm spacing (in z direction), respectively between each two consecutive layers of straws. The layout of the TRT end-caps can be seen well in figure 2.6. Also the TRT end-cap does not provide measurements along the straw direction but only in z and in the azimuthal direction. The intrinsic resolution of the TRT is 130 µm (in

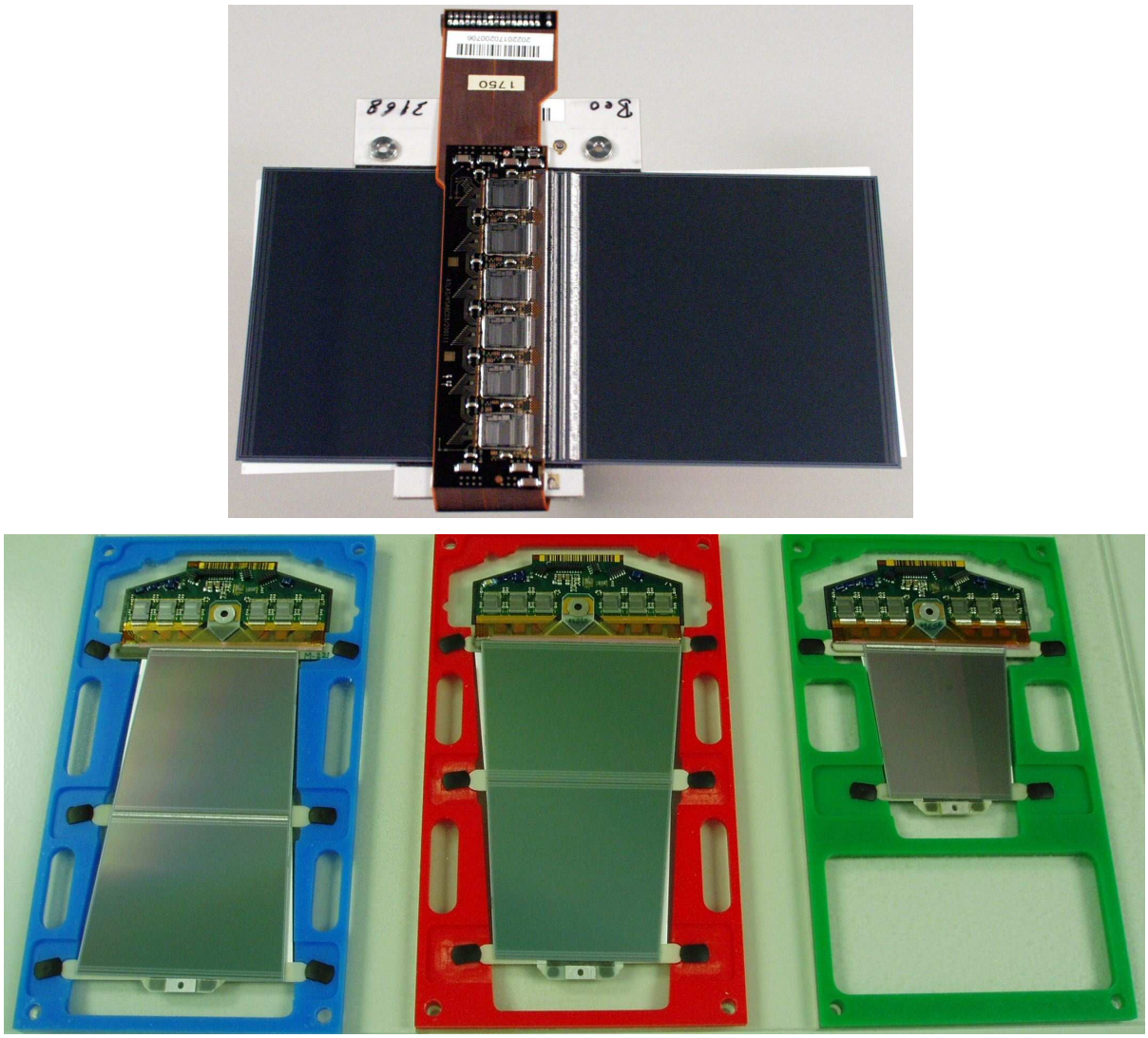


Figure 2.8: Photographs of SCT barrel (top) and end-cap (bottom) modules. Due to its rotation with respect to the top sensor by the 40 mrad stereo angle, also the bottom sensor is clearly visible for all module types.

both barrel and end-cap). Any particle with $p_T > 0.5$ GeV within the acceptance of the TRT traverses at least 36 straws, except for the transition region between barrel and end-cap (around $|\eta| \sim 0.9$), where that number can drop to a minimum of 22. Due to this large number of measurements and the longer track length inside the detector, the TRT can contribute significantly to the overall momentum measurement in spite of its lower intrinsic resolution compared to the silicon trackers. The TRT only covers a pseudorapidity range of $|\eta| < 2.0$, however.

The individual straw layers in the TRT are interleaved with fibers (barrel) and foils (end-cap), causing the crossing particles to produce transition radiation, which is used for particle (particularly electron) identification. This is done by operating the front-end electronics at a "high" and a "low" threshold simultaneously. Transition radiation photons give rise to much higher signal amplitudes than minimum-ionizing charged particles. Therefore particles that produce a lot of transition radiation, in particular electrons, yield a higher fraction of high-threshold hits than other particle types and can be identified that way. For electrons above 2 GeV, typically 7-10 high-threshold hits are expected.

2.2.2 The calorimeter

The ATLAS calorimeter system comprises several different components. The electromagnetic (EM) calorimeter consists of a barrel and two end-caps. The hadron calorimeter consists of the tile calorimeter, parted up in a barrel and two "extended barrels", and the hadron end-caps. Furthermore there is the forward calorimeter, which is usually listed as a hadron calorimeter, but also contains a part which is optimized for electromagnetic measurements. An overview of the calorimeter system is given in figure 2.9. Except for the tile calorimeter, which uses scintillators, all parts of the calorimeter system employ liquid argon as the active medium. Several different absorber materials are used, depending on the background and radiation conditions, resolution etc. The calorimeter covers a large pseudorapidity range of $|\eta| < 4.9$ in order to provide reliable missing transverse momentum measurements. Within the pseudorapidity range of the Inner Detector, the EM calorimeter has an increased granularity, enabling precision measurements of electrons and photons in this range. The coarser granularity outside that range is designed to still allow for good jet and missing transverse momentum measurements.

The electromagnetic calorimeter

The EM calorimeter consists of a barrel ($|\eta| < 1.475$) and two end-caps ($1.375 < |\eta| < 3.2$). Both the barrel and the end-caps use lead as absorber material. The absorber plates as well as the Kapton electrodes are accordion-shaped in order to provide full azimuthal coverage without any cracks. The barrel shares a vacuum vessel with the central solenoid coil in order to reduce the amount of dead material in front of the calorimeter. The end-caps are divided into two coaxial wheels at $|\eta| = 2.5$. Only the outer (i.e. more central) wheel has the finer granularity aiming at precision measurements. Furthermore, both the outer wheel and the barrel are longitudinally segmented into three sections, while the inner wheel is only divided into two sections. The thickness of the EM calorimeter is >22 radiation lengths (X_0) in the barrel and >24 X_0 in the end-caps. In the pseudorapidity region below 1.8 a presampler is used in addition.

The hadron calorimeters

The tile calorimeter. The tile calorimeter is divided into a barrel covering the pseudorapidity range of $|\eta| < 1.0$ and two extended barrels covering $0.8 < |\eta| < 1.7$. Employing steel absorbers along with scintillators, it is the only calorimeter in ATLAS that does not use LAr. It is segmented longitudinally

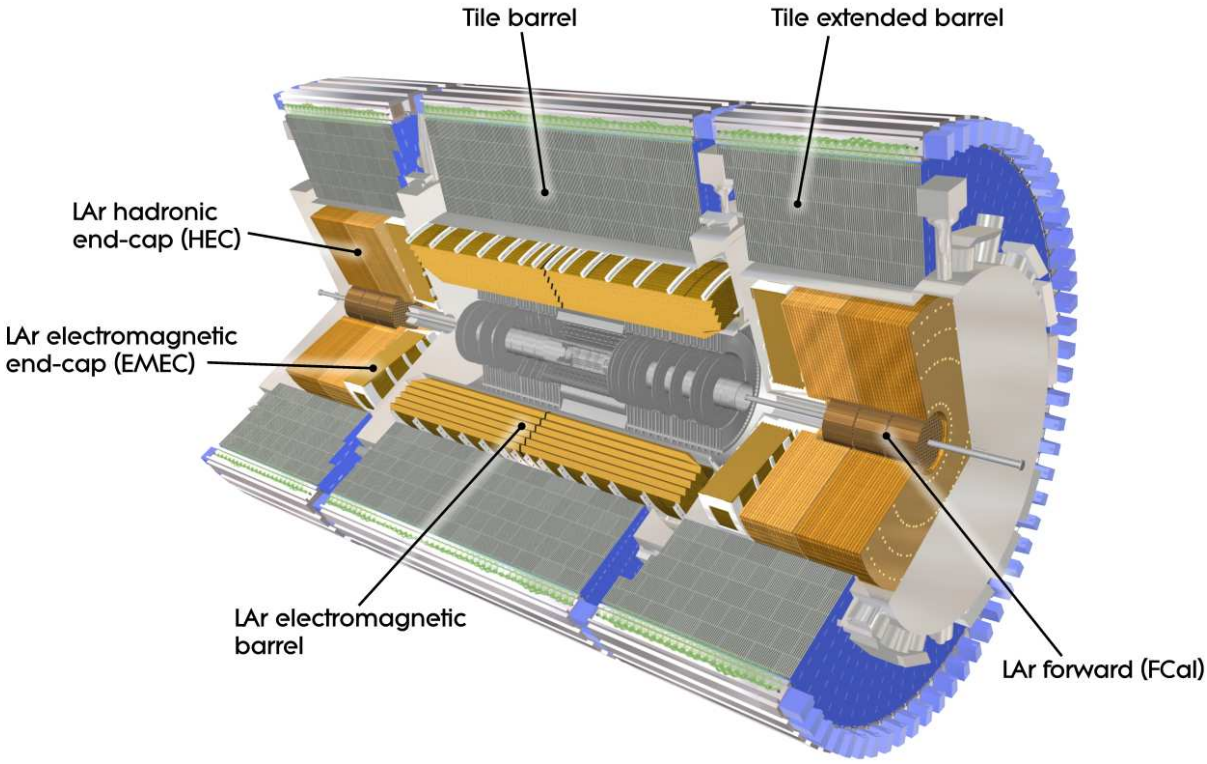


Figure 2.9: Cutaway view of the ATLAS calorimeter system.

into three layers of 1.5, 4.1 and 1.8 interaction lengths (λ), respectively, in the barrel and 1.5, 2.6 and 3.3 λ in the extended barrels. The total active calorimeter thickness (including the EM calorimeter) at $\eta = 0$ is 9.7 λ.

The hadron end-cap calorimeter. The hadron end-cap calorimeter is located directly behind (in the z direction) the end-caps of the EM calorimeter. It comprises two independent wheels per end-cap (in the z direction as well), which together cover a pseudorapidity region of $1.5 < |\eta| < 3.2$, providing some overlap with both the tile calorimeter and the forward calorimeter. The hadron end-cap calorimeter employs copper absorbers. Each end-cap is longitudinally segmented into four layers, two per wheel.

The forward calorimeter. The forward calorimeter is located inside (radially) the hadron end-cap calorimeter. It covers a pseudorapidity range of $3.1 < |\eta| < 4.9$. Though it is usually listed along with the hadron calorimeters, the first of its three modules, which uses copper absorbers, is in fact optimized for electromagnetic measurements. The other two modules use tungsten and are mostly focused on measuring hadronic showers. The thickness of the forward calorimeter is about 10 λ.

2.2.3 The muon spectrometer

The task of the muon system is to detect and measure charged particles that exit the calorimeter system, i.e. in particular to trigger on them and to precisely measure their momenta. For each of these two aspects the muon system comprises an independent set of planar tracking chambers. The trigger chambers cover a pseudorapidity range of $|\eta| < 2.4$, the high-precision chambers for cover a range of $|\eta| < 2.7$. Some

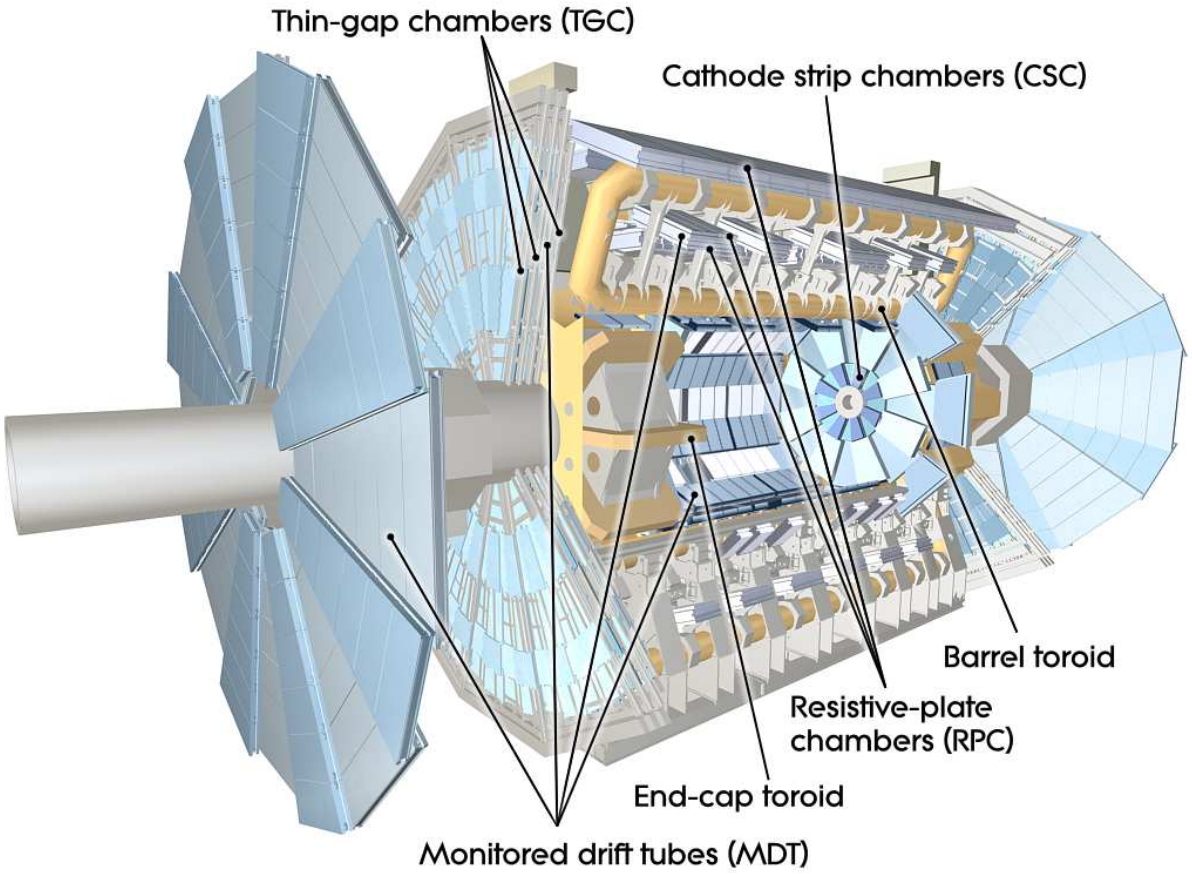


Figure 2.10: Cutaway view of the ATLAS muon spectrometer.

acceptance gaps exist at the positions of the main detector support feet (approximately at $\phi = 240°$ and 300° and $0.25 < |\eta| < 0.3$). The chambers are arranged in three coaxial cylinders with radii of about 5 m, 7.5 m and 10 m in the barrel and in 4 disks located at $|z| \approx 7.4$ m, 10.8 m, 14 m and 21.5 m in each end-cap. For each type of chambers, two different technologies are employed. For the high-precision chambers, mostly Monitored Drift Tubes (MDTs) are used, except for the innermost part of the innermost end-cap disk, where Cathode Strip Chambers (CSCs) are employed. For triggering, Resistive-Plate Chambers (RPCs) are used in the barrel and Thin-Gap Chambers (TGCs) in the end-caps. The momentum measurements are facilitated by the field of the huge air-core toroid magnet system, which also consists of a barrel and two end-cap parts. An overview of the muon system is given in figure 2.10. Figure 2.11 shows a schematic overview of the precise arrangement of the chambers.

The toroid magnet system

The toroid magnet system comprises eight magnet coils in the barrel and eight more in each end-cap. In the barrel, each of the huge coils is housed in its individual cryostat. In each end-cap all eight coils share a single cryostat. The end-caps are inserted into each end of the barrel toroid. The positions of the coils in the end-caps are rotated with respect to the barrel by 22.5° (i.e. half the azimuthal distance between two coils) in order to allow for some radial overlap between the barrel and end-cap coils and to optimize the magnetic field in the transition region between barrel and end-cap. The magnetic field created by the toroid coils points mostly in azimuthal direction. Thus the field is mostly orthogonal to the muon

trajectories, maximizing the bending power. Accordingly, the deflection of the trajectories takes place mostly in the polar direction. The achieved bending power can be defined as the integral $\int (|\vec{B}\times\vec{dl}|)$ of the normal component of the magnetic field along an infinite-momentum trajectory between the innermost and outermost crossed muon chamber plane. It amounts to between 1.5 Tm and 5.5 Tm in the barrel ($|\eta| < 1.4$) and between 1 Tm and 7.5 Tm in the end-cap ($1.6 < |\eta| < 2.7$), while in the transition region ($1.4 < |\eta| < 1.6$) the bending power is lower.

The precision chambers

The high-precision chambers have a very good spatial resolution in the main bending direction of the tracks in the magnetic field (i.e. z direction in the barrel, radial direction in the end-caps), enabling the desired good momentum measurements. In contrast, they provide no (MDTs) or only poor (CSCs) measurements perpendicular to this (i.e. in azimuthal) direction.

Monitored drift tubes. The MDTs make up most of the precision chambers in the muon spectrometer. The chambers consist of three to eight layers of independent drift tubes (slightly similar to the TRT). In the sensitive direction, they yield a resolution of about 80 μm per tube and about 35 μm for the whole chamber. Orthogonal to that direction, i.e. along the signal wires, they don't provide any measurement at all. The overall arrangement of the MDTs can be seen in the top part of figure 2.11 (they are labeled "BOL","BML", "EIL", "EEL" etc. there). The precise arrangement of the MDTs in the barrel is shown in the bottom part.

Cathode Strip Chambers. In the highest pseudorapidity region ($2 < |\eta| < 2.7$) of the innermost disk in each end-cap, where the highest occupancies are expected, CSCs are used as precision chambers instead of MDTs. The CSCs are multiwire proportional chambers whose cathode planes are segmented into strips. They provide a similar resolution as the MDTs in the sensitive direction (40 μm). In contrast to the MDTs they also provide at least a very coarse measurement perpendicular to this direction, with a resolution of about 5 mm. Furthermore, their high granularity gives them the higher rate capability with respect to the MDTs which is required in the position where they are mounted.

The trigger chambers

The trigger chambers are much less precise in their position measurement than the precision chambers, but have a very good timing resolution of 15-25 ns, including contributions from signal propagation and electronics. This enables them to perform precise bunch-crossing identification. There are three layers of trigger chambers in the barrel, two of them sandwiching the middle MDT layer and the third one mounted to the outer MDT layer. In the end-cap, three layers of trigger chambers sandwich the third MDT disk and a fourth layer is installed inside the innermost disk. The exact layout of the trigger chambers can be seen in figure 2.11 (top figure). The trigger chambers measure both space coordinates. The coordinate in bending direction is only used within the trigger in order to provide momentum thresholds. The orthogonal coordinate, which is not measured by the MDTs, is also used for the final muon reconstruction. This is the third main purpose of the trigger chambers, besides triggering and bunch crossing identification.

Resistive-Plate Chambers. The trigger chambers in the barrel are RPCs. RPCs are wireless gaseous detectors; a layer of them consists of two parallel resistive plates at 2 mm distance, aligned by insulating spacers. Between the electrodes there is an electric field of 4.9 kV/mm, which is strong enough that

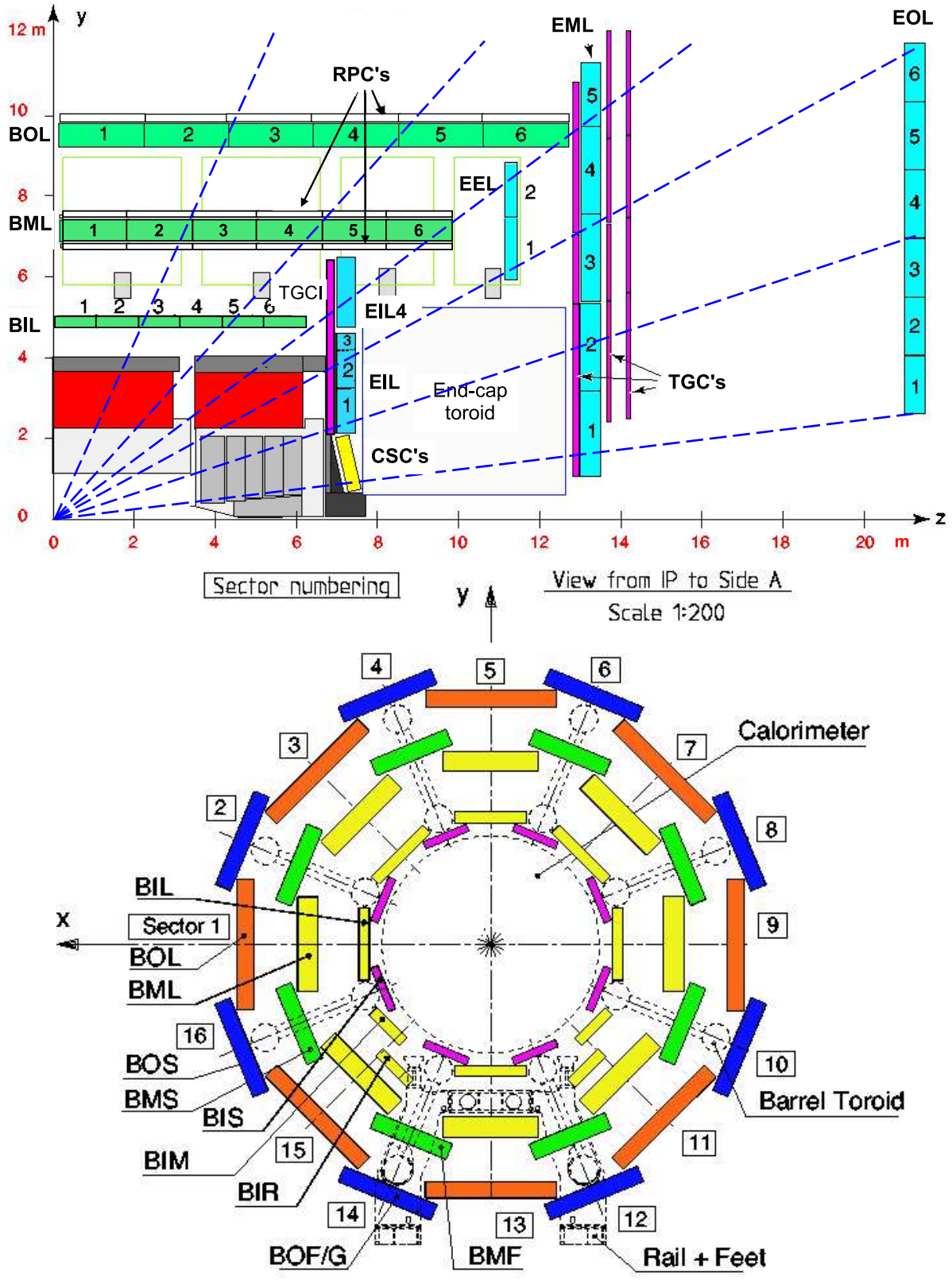


Figure 2.11: Schematic view of the ATLAS muon spectrometer in the *yz* plane (top) and in the *xy* plane (bottom).

primary electrons produced by traversing ionizing particles directly create avalanches. The signal is read out by metallic strips on the outer faces of the plates. Each RPC consists of two such layers, providing two measurements of each coordinate. The achieved spatial resolution is about 10 mm in both directions.

Thin-Gap Chambers. In the end-caps, TGCs are used as trigger chambers. The TGCs are basically "normal" multiwire proportional chambers with the distinctive feature that the distance of the wires to the cathode plate of 1.4 mm is smaller than the distance between two neighbored wires of 1.8 mm. By ganging the wires, the spatial resolution of the RPC can be adjusted according to the requirements, which depend on the pseudorapidity. The achieved resolution is 2-6 mm in r and 3-7 mm in the azimuthal direction.

2.2.4 Components used for luminosity measurements

In ATLAS, various methods for measuring the luminosity are employed, which make use of different detector components. Only those two components that were used to provide the main luminosity measurement in the 2011 data-taking period, which is used in this thesis, are briefly introduced in the following. The way how these detectors are used for luminosity measurement is described in Section 4.7.

LUminosity measurement using Cerenkov Integration (LUCID)

The two modules of LUCID are located at ±17 m distance from the interaction point, covering a pseudorapidity range of $5.6 < |\eta| < 6.0$. They are the primary online luminosity monitors of ATLAS, designed to measure inelastic proton–proton scattering at low angles. LUCID is a Cerenkov detector; each LUCID module consists of 16 C_4F_{10}-filled aluminum tubes of 1.5 m length, placed at a distance of 10 cm to the beam line. The Cerenkov light emitted by charged particles traversing the tubes is reflected by the tube walls until it is detected by photomultipliers at the back end of the tubes. A photomultiplier signal above a certain threshold is recorded as a hit for this event. Hits can uniquely be assigned to individual bunch crossings.

The Beam Condition Monitor (BCM) system

The primary purpose of the BCM is to monitor background levels and to trigger a beam-abort in case of risk of damage to ATLAS detectors due to beam losses. The BCM comprises two stations, each consisting of four modules, located at $z = \pm 184$ cm and arranged in a cross pattern around the beam pipe at a radius of $r = 5.5$ cm ($|\eta| = 4.2$) . Each module contains two radiation-hard diamond sensors, which are read out by very fast electronics with a rise time of 1 ns. Using the time-of-flight distance between the two stations, the BCM is able to distinguish between particles from normal collisions and stray protons.

2.2.5 Trigger

The ATLAS trigger system uses three levels of triggers, named *Level 1* (L1), *Level 2* (L2) and *Event Filter* (EF). The L2 trigger and the EF together are generally referred to as the High-Level Trigger (HLT). The trigger system is designed such that the amount of data transfer within the trigger system gets minimized in order to cope with the high collision rate of the LHC and to allow for maximum possible trigger rates.

Level 1 Trigger

The L1 trigger is completely hardware-based, employing purpose-designed electronics. The main L1 physics triggers use only information from the calorimeter with reduced granularity and from the muon trigger chambers. Trigger signatures are mostly given by multiplicities of high-p_{T} objects at different thresholds. In addition different thresholds of the total transverse momentum, the missing transverse momentum and the total transverse momentum of all jets can be triggered on. Further input to the L1 trigger is provided e.g. by the BCM and other minor detector components, which are used to trigger on things like filled bunch crossings or minimum-bias events. The L1 trigger has a maximum accept rate of 75 kHz. The decision must be taken at most 2.5 μs after the collision. In order to adjust the accept rates of the individual trigger items (i.e. the signatures that are triggered on) as desired, each trigger item can be prescaled by a factor of up to 2^{24}. As prescaled triggers are rather inconvenient for data analysis, this is only done for trigger items that would otherwise consume a too high fraction of the available overall trigger rate (or even exceed it).

High-Level Trigger

The HLT is software-based, implemented on commercially available hardware, concretely on 36 racks of server-class PCs, where multiple events can be processed in parallel. It is integrated within the ATLAS Data AcQuisition system (DAQ), which takes over the control of the data flow right after the L1 decision, starting with the readout of the data from the detectors and ending with writing out the events accepted by the EF to mass storage.

Level 2. The L2 trigger still does not analyze the full event yet. It does use full granularity and takes into account information from all detector components, but it only (with few exceptions) looks at the so-called "regions of interest" (ROI), which are the locations in η and ϕ of the object(s) found by the L1 trigger. This amounts to about $1-2\%$ of the whole event data, saving a substantial amount of data transfer. The L2 accept rate is up to 3.5 kHz with an average processing time of 40 ms per event.

Event Filter. The EF for the first time analyzes the full event. It employs mostly the same reconstruction algorithms that are also used offline, taking on average 4 s processing time per event. The accept rate of the EF is about 300 Hz.

Data streams. Besides accepting or rejecting events, the EF classifies the accepted events according to ATLAS-defined data streams. In principle, each event can be assigned to more than one stream, while in practice the overlap is small. Each accepted event is then written to one or more data files, according to the data streams they were assigned to by the EF. The most relevant physics streams are named "Egamma" (containing events with electron and/or photon signatures), "Muons" (muon signatures), JetTauEtmiss (jet, missing transverse momentum and τ lepton signatures) and "minBias" (events that pass a dedicated minimum bias trigger). Furthermore there is the express stream, which is used for monitoring the data quality before the bulk reconstruction is performed on the physics streams, and the calibration stream, that only records a minimum subset of the event data used for detector calibration.

Chapter 3

The lepton+jets decay mode of single top-quark production in the Wt-channel

This chapter discusses the signal channel of the analysis that is presented in this thesis, the lepton+jets decay mode of single top-quark production in the Wt-channel. First the top quark and its position in the Standard Model are introduced. Then its different production modes at hadron colliders are discussed briefly. Focus is on the LHC, while the situation at the Tevatron is usually given as well for comparison. The decay of top quarks and the typical resulting final states are discussed. Finally, some theoretical issues with the description of Wt-channel production at NLO are detailed a bit more, before a precise definition of the signal final state as it was used for analysis in the scope of this thesis is given and the relevant background processes are discussed briefly.

3.1 The top quark in the Standard Model

3.1.1 Properties of the top quark

The top quark is the up-type quark of the third generation. As such, it has an electric charge of +2/3 e. With a mass of about 173 GeV, it is the heaviest particle in the Standard Model, in particular it is by far heavier than all other fermions (the second-heaviest fermion, the b quark, has a mass of only 4.7 GeV). The width of the top quark as predicted by the Standard Model is $\Gamma_t = 1.34$ GeV for a top-quark mass of $m_t = 172.6$ GeV [28]. This corresponds to a very short lifetime of only $\tau_t \approx 0.5 \cdot 10^{-24}$ s, which has some important consequences.[1] Firstly, its lifetime is smaller than the typical time it takes for a quark to hadronize, $\tau_{\text{hadr}} \approx 3 \cdot 10^{-24}$ s. This means that top quarks decay before they are able to form bound states; they rather decay as bare quarks, enabling measurements of the top-quark mass that are much more precise than those of any other quark mass. Furthermore, top quarks also decay before their spin can depolarize due to the strong interaction, which makes it possible to study spin properties.

3.1.2 Discovery of the top quark

The top quark was predicted by theory long before its discovery. Already in 1973, when only three quarks had been found experimentally and a fourth one was anticipated in order to theoretically explain the experimentally observed absence of flavor-changing neutral currents (GIM mechanism [4]), Kobayashi and Maskawa proposed a third quark generation in order to explain CP violation in the Standard Model by a three-dimensional quark mixing matrix.[12] After the b quark was discovered in 1977 and identified as the down-type quark of the third generation, evidence was strong that indeed also its up-type partner should exist. One of the arguments was again the GIM mechanism. Furthermore the existence

[1] As the lifetime of the top quark is inversely proportional to the third power of its mass, one may in fact attribute these properties to the large top-quark mass.

of the top quark would ensure that the theory is free from anomalies, i.e. situations in which higher-order corrections break an apparent symmetry of the Lagrangian. The Standard Model contains such graphs which might break electroweak gauge symmetry, potentially destroying the renormalizability of the Standard Model. The symmetry breaking terms cancel out, however, if the partner of the b quark exists.[28–30]

Also the mass of the top quark, being one of its most outstanding characteristics, could be predicted by theory already before it was discovered. The higher-order corrections to some of the general relations between the fundamental electroweak parameters are dominated by top-quark loops, due to the high mass of the top quark. By calculating these corrections as a function of the top-quark mass, electroweak precision measurements performed at LEP made it possible to predict the top-quark mass to be in the order of 170 GeV shortly before its discovery.[31]

The discovery of the top quark finally took place in 1995 at the Tevatron, where CDF and D0 both measured the pair production of a top quark and an antitop quark.[32]

3.1.3 Top-quark related elements of the CKM matrix

The top-quark related elements of the CKM matrix can mostly be determined only indirectly, i.e. using processes that depend on these elements only due to higher-order graphs. The elements V_{td} and V_{ts} can be determined from measurements of certain processes involving B or K mesons which are dominantly mediated by box or loop diagrams containing top quarks, such as $B\bar{B}$ oscillations or certain rare decays. As such processes are also generally sensitive to beyond-the-standard-model contributions, any such measurement relies on the assumption that the Standard Model is valid, in particular that there are exactly three quark generations and that unitarity holds. The accuracy of such measurements is limited by the underlying theoretical uncertainties. The element V_{tb} can be determined from the ratios of the branching fractions of top-quark decays to a b quark and top-quark decays into any quark:

$$R = \mathcal{B}(t \to Wb)/\mathcal{B}(t \to Wq)\,.$$

Experimentally this ratio can be extracted from the fractions of measured top quark pair events with zero, one and two identified b-quark jets. This determination of V_{tb} again involves the assumption of unitarity.

The only direct (i.e. without assuming unitarity) determination of one of the third-row CKM matrix elements that is considered to date is the determination of V_{tb} via the measurement of single top-quark production cross sections at hadron colliders. As single top-quark production proceeds via the Wtb vertex, its cross section is directly proportional to the corresponding coupling strength, which contains V_{tb} as the only part that is not known from independent measurements. This direct measurement of V_{tb} was performed successfully for the first time in 2009 at the Tevatron, the value obtained from the combined D0 and CDF cross section is $V_{tb} = 0.88 \pm 0.07$.[33] In the future, such measurements will also be carried out the LHC. One of the ongoing single top-quark production analyses in ATLAS, which is intended to also provide a V_{tb} measurement at some point, is the subject of this thesis.

3.2 Top-quark production at the LHC

Top-quark production at hadron colliders can be classified in two general production mechanisms: the pair production of a top quark and an antitop quark via the strong interaction and the production of single top quarks via the weak interaction, more precisely the Wtb vertex.

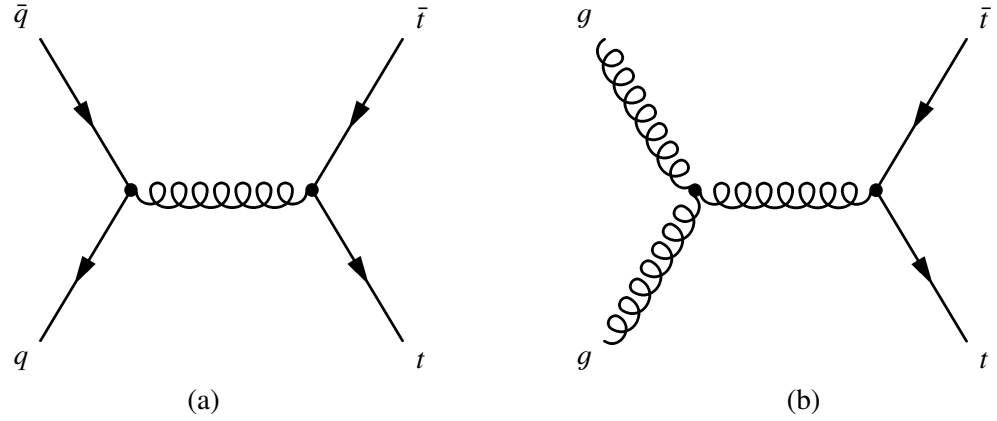


Figure 3.1: Example Feynman graphs for $t\bar{t}$ production at leading order via quark-antiquark annihilation (a) and gluon-gluon fusion (b).

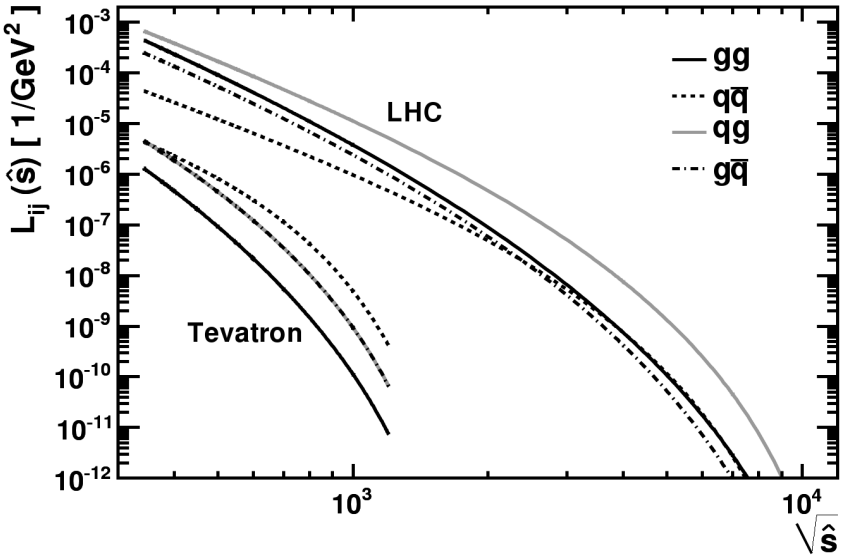


Figure 3.2: Parton luminosities for gg, gq, $g\bar{q}$ and $q\bar{q}$ interactions at the LHC and the Tevatron. The values on the x axis are in GeV. Taken from [28]

3.2.1 Top quark pair production

The dominant source of top quarks at both the LHC and the Tevatron is the pair production of a top quark and an antitop quark via the strong interaction, in the following simply denoted as top quark pair production or $t\bar{t}$ production. At leading order this process is induced either by quark-antiquark annihilation ($q\bar{q} \rightarrow t\bar{t}$) or by gluon-gluon fusion ($gg \rightarrow t\bar{t}$). Figure 3.1 shows an example Feynman graph for each of these two processes. While at the Tevatron about 85% of the top quark pair production is due to quark-antiquark annihilation, at the LHC gluon-gluon fusion is by far the dominant production mechanism. As the hard scattering processes are the same for both colliders, this difference must be due to the parton luminosities, according to the factorization theorem. The parton luminosities at the Tevatron and the LHC are shown in figure 3.2. One can see that all the parton luminosities increase toward lower partonic center-of-mass energies. Hence, most $t\bar{t}$ production must happen close to the kinematic threshold of $\sqrt{\hat{s}} \approx 350$ GeV, which is just at the left edge of the plot. One can clearly see that in this region $L_{q\bar{q}} \gg L_{gg}$ at the Tevatron, while at the LHC the situation just reversed. One reason for

this is that the Tevatron collides protons and antiprotons and therefore valence quarks can contribute to $q\bar{q}$-induced processes. The second reason is that the LHC has a much higher center-of-mass energy than the Tevatron, such that the same partonic center-of-mass energy corresponds to much lower momentum fractions, where gluons generally play a dominant role. In fact, at the LHC center-of-mass energies, gg fusion would dominate even for $p\bar{p}$ collisions.

The theoretical prediction for the total $t\bar{t}$ production cross section at the LHC at $\sqrt{s} = 7$ TeV is 165^{+11}_{-16} pb, assuming a top-quark mass of 172.5 GeV.[34, 35]

3.2.2 Single top-quark production

One distinguishes three different modes of single top-quark production, called the t-channel, the s-channel and the Wt-channel. Example Feynman graphs at leading order for all of them are shown in figure 3.3. Theoretical predictions for the total cross sections at LHC and Tevatron, mostly at NNLL, are shown in table 3.1. The total single top-quark production cross section at the LHC at $\sqrt{s} = 7$ TeV is predicted to be 84.4 pb, i.e. about 50% of the $t\bar{t}$ cross section. Single top-quark production was first experimentally observed in 2009 at the Tevatron by combining the t-channel and the s-channel to a single analysis channel.[33] In ATLAS, t-channel production could be measured for the first time in 2011.[36]

Common to all three production modes is that they proceed via the electroweak Wtb vertex. Any single top-quark production cross section measurement therefore can be used to extract the V_{tb} CKM matrix element.

	$\sqrt{s}$	t-channel	s-channel	Wt-channel
$\sigma(p\bar{p} \to t/\bar{t})$ [pb]	1.96 TeV	$2.08^{+0.00+0.12}_{-0.04-0.12}$	$1.05^{+0.00+0.06}_{-0.01-0.06}$	0.25 ± 0.03
$\sigma(pp \to t)$ [pb]	7 TeV	$41.7^{+1.6\ +0.8}_{-0.2\ -0.8}$	$3.17^{+0.06+0.13}_{-0.06-0.10}$	$7.8^{+0.2+0.5}_{-0.2-0.6}$
$\sigma(pp \to \bar{t})$ [pb]	7 TeV	$22.5^{+0.5\ +0.7}_{-0.5\ -0.9}$	$1.42^{+0.01+0.06}_{-0.01-0.07}$	$7.8^{+0.2+0.5}_{-0.2-0.6}$

Table 3.1: Predicted total cross sections in picobarn for all three single top-quark production modes for LHC at $\sqrt{s} = 7$ TeV and for Tevatron at $\sqrt{s} = 1.96$ TeV including theoretical uncertainties. The Wt-channel cross section for Tevatron is taken from [28]; it was evaluated at a top-quark mass of 175 GeV and the quoted error includes both scale and PDF uncertainties, where the PDF set is MRST2004 NNLO.[37] All other numbers are from [38], calculated at NNLL (more precise: approximate NNLO at NNLL accuracy) and evaluated at $m_t = 173$ GeV. The first of the quoted uncertainties refers to the scale dependence, the second one is the PDF uncertainty at 90% confidence level, the PDF set is MSTW2008 NNLO.[39] The values for Tevatron are the sums of t and $\bar{t}$ production.

t-channel

In t-channel production, a sea b-quark exchanges a W boson in the t-channel with another – typically light – quark, thereby turning into a top quark. The t-channel production is expected to be the dominant production mode for single top quarks at both the LHC and the Tevatron. At the Tevatron its fraction of the total single top-quark production cross section is about 60%, at the LHC it is even about 75%.

In fact, the t-channel process is of the gq type, as the b quark actually comes from a gluon splitting. By adding this gluon splitting to the Feynman graph shown in figure 3.3(a), the final state would obtain an additional b quark (see also the Wt-channel paragraph). For analysis, however, this additional b quark is normally not taken into account, as it usually lies outside the acceptance range of the typically used

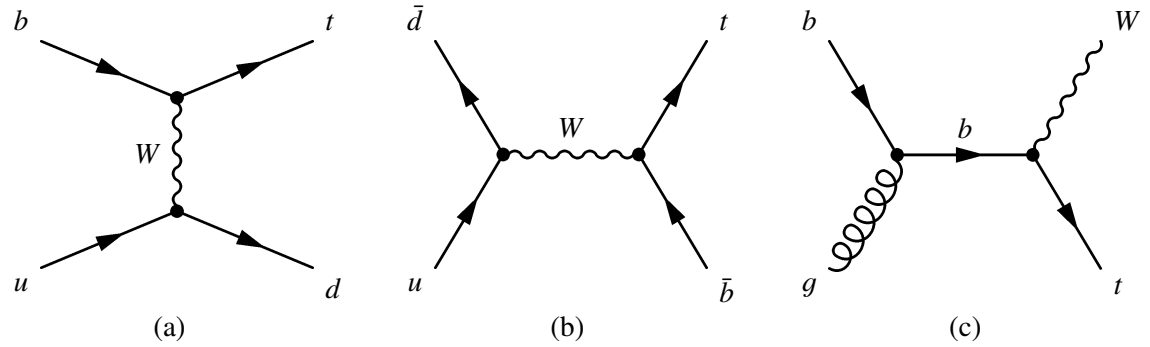


Figure 3.3: Example Feynman graphs at leading order for the three single top-quark production modes: the *t*-channel (a), the *s*-channel (b) and the *Wt*-channel (c).

jet definitions. Instead, the most distinct signature of the *t*-channel which is exploited for analysis in ATLAS is its final-state light quark, which typically has very high pseudorapidity. Requiring a light-quark jet with typically $|\eta| > 2.5$ is therefore one of the most important cuts in *t*-channel analyses.

s-channel

In *s*-channel production, an up-type quark and a down-type antiquark or vice versa – normally indeed up and down flavored – annihilate to a *W* boson in the *s*-channel, which then "decays" into a top quark and an antibottom quark (or vice versa). At the Tevatron the *s*-channel is the second-most important single top-quark production channel, with a cross section that amounts to about 50% of that of the *t*-channel. At the LHC in contrast the *s*-channel has the smallest cross section of all three production modes, making up only about 5% of the total single top-quark production. The reason is again that at the LHC $q\bar{q}$ annihilation processes play a much smaller role than at the Tevatron.

Experimentally, it has been attempted at the Tevatron to separate *s*-channel from *t*-channel production, while at the LHC the *s*-channel will be the last single top-quark production mode that will be measurable.

Wt-channel

In *Wt*-channel production, the top quark is produced in association with an on-shell *W* boson, denoted the *prompt W* in the following. At leading order, the prompt *W* boson is emitted by a (sea) *b* quark, which at the same time turns into a top quark, either before or after the quark gets struck by a gluon. The diagram for the former case, shown in figure 3.4(a), has a top-quark line in the *t*-channel, while the (much more frequently shown) diagram for the latter case, shown in figure 3.4(b), has a *b*-quark line in the *s*-channel. While at the Tevatron the *Wt*-channel makes up only about 7% of the total single top-quark production and will probably not be measurable, it is the second-most important production mode at the LHC, contributing about 18% to the total single top-quark production cross section.

As in *t*-channel production, the initial-state *b* quark in the graphs shown in figures 3.4(a) and (b) originates from a gluon splitting. The "full" version of figure 3.4(b) including the gluon splitting is shown in 3.4(c). Whether this gluon splitting is considered to be part of the hard scattering or part of the proton (or a bit of both) depends on the used factorization scheme and scale. In a pure four-flavor scheme, which does not involve a *b*-quark PDF, 3.4(c) is the actual full LO graph rather than 3.4(b). When calculating the cross section in this scheme, terms proportional to $\ln((Q^2 + m_t^2)/(m_B^2))$ appear, which give rise to a singularity in case the *b* quark is taken to be massless. When the *b*-quark mass is

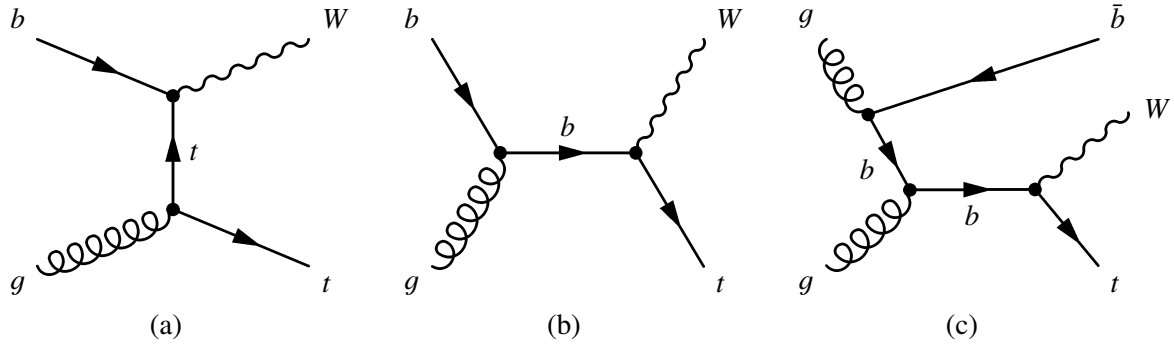


Figure 3.4: Leading-order Feynman graphs of *Wt* production: *t*-channel-like (a), *s*-channel-like (b) and including the gluon splitting which produces the initial-state *b* quark of the two former graphs (c).

taken into account, they still cause the perturbation series to converge only slowly.[28] This issue, which the *Wt*-channel shares with the *t*-channel, gets solved by working in a five-flavor scheme, employing a *b*-quark PDF, which implicitly describes the gluon splitting inside the protons. In this scheme, figure 3.4(b) indeed becomes the leading order graph and 3.4(c) is a NLO correction, that is already partially included in the PDF.[28]

When calculating the *Wt*-channel cross section at this level, another problem occurs, however, which is specific to the *Wt*-channel and which turns out to be much more severe. This problem is discussed in Section 3.4.1.

3.3 Top-quark decay

In the Standard Model the top quark decays into a down-type quark and a W boson. Due to the structure of the CKM matrix, the decay into a b quark has a branching fraction of almost 100%, while decays to s or d quarks are highly suppressed and are therefore neglected throughout this thesis. In the following, the typical final states of top-quark production are characterized.

Final states of top-quark production

The final state of a top-quark decay is usually categorized by the decay mode of the W boson, which decays into a pair of (mostly light) quarks ("hadronic W/top decay", $\mathcal{B} = 68\%$) or into a charged lepton and its corresponding neutrino ("leptonic W decay"/"semileptonic top decay", $\mathcal{B} = 32\%$). Such leptons from leptonic W decays are typically hard (i.e. they have high p_{T}) and isolated (the same actually holds for leptons from Z decays); this provides a clear signature that is well-suited for selecting events. The same holds for the large amount of missing transverse momentum due to the neutrino. For production modes that have two W bosons in their decay chain, like the *Wt*-channel (or $t\bar{t}$ production), there are three different overall decay modes/channels:

- both W bosons decay hadronically ("all-hadronic" channel, $\mathcal{B} = 46.2\%$);
- one W decays hadronically, one decays leptonically ("lepton+jets" channel, $\mathcal{B} = 43.5\%$);
- both W bosons decay leptonically ("dilepton" channel, $\mathcal{B} = 10.3\%$).

The all-hadronic channel has the highest branching fraction of all three decay channels. But as it contains solely jets in the final state it is very difficult to separate it from background (particularly

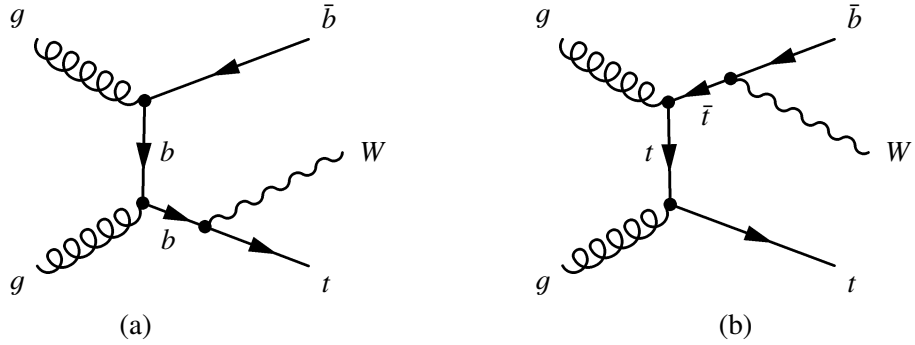


Figure 3.5: Next-to-leading-order Feynman graphs of *Wt* production: singly resonant (a) and doubly resonant (b), involving an additional internal top-quark line

multi-jets, see Section 3.4.3), and therefore this channel is usually considered last for analyses. The lepton+jets channel has a much more pronounced signature: one – typically hard and isolated – charged lepton, along with large missing transverse momentum, as well as some jets, among them the *b*-quark jet(s) from the top-quark decay(s). This makes the lepton+jets channel much easier to isolate than the all-hadronic channel, while its branching fraction is only slightly lower. Due to this rather good compromise between easy separation from background and high branching fraction, it is sometimes referred to as the "golden channel" (mostly in the context of $t\bar{t}$ analyses). The dilepton channel features even two charged leptons – both being hard and isolated – and it also has large missing transverse momentum. The only jets in the final state are the *b*-quark jet(s) from the top-quark decay(s) as well as jets that may have been produced in association with the top quark(s) (e.g. by gluon radiation). This signature allows an even cleaner selection, which just about compensates its low branching fraction. This thesis studies the lepton+jets decay mode of the *Wt*-channel. A precise signal definition as well as an overview of the relevant background processes are given in the next section.

3.4 The lepton+jets decay mode of *Wt* production

3.4.1 Theoretical issues with the definition of the *Wt*-channel at next-to-leading order

The *Wt*-channel at NLO. As mentioned in Section 3.2.2, a severe problem arises when calculating the *Wt*-channel cross section at NLO, i.e. at the level of figure 3.4(c). At this level diagrams contribute that contain an additional, internal top-quark line which can become on-shell. These diagrams are referred to as doubly resonant diagrams in the following, in contrast to the diagrams where only the final-state top quark can become on-shell, which are called singly resonant. Figure 3.5 gives an example for each type of diagram. Figure 3.5(a) represents the singly resonant type. It is the same diagram as 3.4(c), only slightly rearranged in order to illustrate the similarity to the doubly resonant diagram shown in figure 3.5(b). They turn into each other by simply moving the *Wtb* vertex along the quark line. Moving the vertex until it is between the two quark-gluon vertices would result in the (singly resonant) diagram that corresponds to 3.4(a).

The problem. The doubly resonant diagrams can be interpreted as LO $t\bar{t}$ production graphs with the subsequent decay of one of the top quarks. Apparently, any LO $t\bar{t}$ graph contributes to the NLO *Wt* cross section this way. Given that the cross section for $t\bar{t}$ is more than ten times the *Wt* cross section, it is therefore not surprising that, owing to the kinematic regions where the second top-quark line is on-shell,

the doubly resonant diagrams give rise to NLO corrections that are much larger than the actual LO cross section, which basically makes the perturbation series collapse.

In fact, the precise nature of the problem is even more complicated. A detailed explanation is given in [40]. It is related to the fact that NLO here only refers to QCD, while the order of the electroweak coupling is kept constant at leading order. Now the doubly resonant diagrams give rise to a singularity, when the internal top-quark line becomes resonant, which can only by avoided by taking into account the non-zero width of the top quark. This finite width, however, is an all-order result in the weak coupling, contradicting the constant-order assumption. In other words, the only way to avoid a singularity in the NLO QCD corrections given by the diagrams of figure 3.5 (and others) implies that the diagrams in 3.4(a)/(b) are not the corresponding LO contribution (due to the different order in the weak coupling). To phrase it even more drastically: there is no perturbative QCD expansion whose LO contribution is given by the graphs shown in figure 3.4(a)/(b)! The consequence is that, in a strict sense, at NLO order the *Wt*-channel is not a priori well-defined.

The workaround. It is however possible to recover a workable definition of the *Wt*-channel at NLO in a sense of giving it an operative meaning. Operative meaning here denotes that it can be used to calculate meaningful theoretical predictions as well as in an experimental context (which includes the need for MC simulation), even if it may be not fully consistent theoretically. Several such solutions have been proposed (see [40] for an overview and references therein for details). The general idea of all these approaches is to separate out $t\bar{t}$ production and *Wt* production from each other, i.e. to remove the $t\bar{t}$ contribution (as this is given by the doubly resonant diagrams, which were shown to be the actual source of the problem) from the *Wt* cross section in some way. What makes things difficult is that a clear separation is only possible if there is no interference between both processes, i.e. the singly resonant and doubly resonant diagrams, which is not necessarily the case in general.

The solution proposed in [40] nicely illustrates this aspect, as will become clear below. It is driven by the demand that it ought to be workable both for theoretical calculations and in an experimental environment, including the application in the context of MC generators. In particular no kinematic cuts need to be imposed at generator level, such that the impact of any such cuts can be studied with full flexibility at analysis level. In fact, two distinct definitions for the *Wt* channel are given. The first one, called *Diagram Removal* (DR), simply excludes ("removes") all the doubly resonant diagrams from the calculation at amplitude level.[2] The second method, called *Diagram Subtraction* (DS), tries to subtract the contribution from the doubly resonant diagrams at cross-section level. The difference between the cross sections is then supposed to provide a measure for the strength of the interference between *Wt* and $t\bar{t}$ production. This gets a bit clearer when one writes down the total NLO amplitude, $\mathcal{A}$, as the sum of the contribution from all singly resonant diagrams, $\mathcal{A}_{Wt}$, and the contribution from all doubly resonant diagrams, $\mathcal{A}_{t\bar{t}}$.

$$\mathcal{A} \equiv \mathcal{A}_{Wt} + \mathcal{A}_{t\bar{t}}\,.$$

The calculation of the cross section involves the square of this amplitude

$$\begin{aligned} |\mathcal{A}|^2 &= |\mathcal{A}_{Wt}|^2 + 2\mathcal{R}\{\mathcal{A}_{Wt}\mathcal{A}^*_{t\bar{t}}\} + |\mathcal{A}_{t\bar{t}}|^2 \\ &\equiv \mathcal{S} + \mathcal{I} + \mathcal{D}\,. \end{aligned}$$

In the DR definition, $\mathcal{A}_{t\bar{t}}$ is entirely removed from the calculation such that both the $t\bar{t}$ contribution $\mathcal{D}$

[2] This violates gauge invariance, which is shown to be unproblematic in practice, however.[40]

and the interference term $\mathcal{I}$ disappear, leaving only the Wt contribution $\mathcal{S}$:

$$|\mathcal{A}_{\mathrm{DR}}|^2 = \mathcal{S}\,.$$

In the DS definition, a subtraction term $\tilde{\mathcal{D}}$ is introduced, which is constructed such that it locally, i.e. when the additional top-quark line is on-shell, cancels the $t\bar{t}$ contribution and is furthermore gauge invariant. In a simplified but illustrative manner, one can write

$$\begin{aligned} |\mathcal{A}_{\mathrm{DS}}|^2 &= \mathcal{S} + \mathcal{I} + \mathcal{D} - \tilde{\mathcal{D}} \\ &\approx \mathcal{S} + \mathcal{I}\,. \end{aligned}$$

One can see that, apart from the only local (and thus incomplete) cancellation of the $t\bar{t}$ contribution in the DS definition, the difference between DS and DR is exactly the interference term.[3] Turning this around one can conclude that when both definitions give approximately the same result, the interference between Wt and $t\bar{t}$ is small and an operative definition of the Wt-channel in the sense of above is achieved. One can then say that the Wt-channel is well-defined.

It turns out that the interference term amounts to about 10% of the total cross section, which is not substantial, but still higher than desirable. It is therefore nessecary to define analysis cuts to further reduce the interference and hence ensure that the given definition of the Wt-channel indeed has the desired operative meaning. As an example, a veto on the p_{T} of the second hardest (if present) b/B quark/hadron/jet in the event (depending on the analysis level) is proposed in [40] and [41]. This is based on the idea that, due to the high top-quark mass, a b quark from the decay of a second top quark is normally harder than the additional b quark in the singly-resonant graphs (like figure 3.5(a)), which are taken to come from a gluon splitting inside one proton and therefore supposed to be rather soft. It is shown in [41] that even in a somewhat realistic analysis scenario (i.e. using typical analysis cuts, (several different) finite b-tagging efficiencies etc.), the interference can be brought down to a sufficiently low level that way. The conclusion therefore is that the Wt-channel can indeed be regarded as a well-defined channel, given that adequate analysis cuts are applied.

3.4.2 Definition of the signal final state

As mentioned, this thesis studies the lepton+jets decay mode of the Wt-channel. For analysis only electrons and muons are considered as charged leptons, which are both long-lived enough to be detected directly. These decay modes are referred to as the electron and muon channel, respectively. Tau leptons in contrast can only be detected via their decay products, which leads to a much lower efficiency, purity and accuracy compared to electrons or muons. Therefore, the tau channel is not considered for analysis. Hadronically decaying taus are in fact rather considered as background. Taus that decay leptonically, i.e. into an electron or muon plus two neutrinos, can however be assigned to the electron/muon channel and therefore be recovered as signal. Combining the overall three occurring neutrinos to one – which happens in the missing transverse momentum measurement anyway – the final state looks exactly as if the W had decayed directly into an electron or muon. The difference is that the tau decays will typically have different kinematics, in particular the electron or muon will usually be softer. Subtracting the hadronic tau decays from the lepton+jets channel, the signal branching fraction decreases to 34.1% of the overall Wt production.

With the given signal definition, the final state of Wt production in the lepton+jets decay mode con-

[3] As mentioned, this view is slightly simplified; the precise estimation of the size of the interference is a bit more involved, but still possible and performed in [40].

tains one hard and isolated charged electron or muon, large missing transverse momentum, one b-quark jet and two light-quark jets. For selection primarily the charged lepton, the missing transverse momentum and the single b-quark jet are exploited. Furthermore one or more appropriate jet bins are chosen for analysis. The individual jet bin that is most populated by the signal final state is the three-jet bin, but also the two- and the four-jet bins contain significant fractions of the signal.

3.4.3 Background processes

The background processes for top-quark analyses are usually classified by their contents of either (on-shell) W and Z bosons or top quarks (the only actual relevant process that contains both is the Wt-channel). This classification gives rise to the following background categories, where in each category additional jets may be produced along with the heavy particles:

- "multi-jets" (also denoted "QCD"): no W, Z or t is produced;
- "W+jets": production of a single W boson;
- "Z+jets": production of a single Z boson;
- "diboson": production of WW, WZ or ZZ;
- single top-quark t-channel production;
- single top-quark s-channel production;
- $t\bar{t}$ production.

Figure 3.6 shows the total cross sections for some of these processes as a function of the center-of-mass energy. It nicely illustrates the hugely different orders of magnitude of the cross sections of the different processes. The higher the cross section of a given background is, the stronger it needs to be rejected by the event selection, which at the same time imposes some requirements on the reconstruction. In the following the individual background processes are discussed briefly.

Multi-jets background

Multi-jets background denotes any processes that do not involve the production of any (on-shell) weak bosons or top quarks. It has by far the largest cross section of all backgrounds (approximately represented by σ_{tot} in figure 3.6), superseding the cross sections of any other process by orders of magnitude. Very efficient rejection of multi-jets background is therefore crucial (not only) for any top-quark analysis. Due to the absence of W and Z bosons in the final state, neither hard isolated leptons nor considerable missing transverse momentum is expected. Accordingly, such events are only able to pass the event selection due to mismeasurements of those objects ("fake leptons"/"fake $p_{\text{T}}^{\text{miss}}$"), also the fake leptons are usually less isolated than signal leptons. Therefore the required good multi-jets rejection depends to a large extent on a good fake rejection by the electron and muon reconstruction including appropriate isolation criteria. Also the kinematics of the fake $p_{\text{T}}^{\text{miss}}$ with respect to the rest of the event typically differs from signal events, which can also be exploited in the event selection. Finally, b-tagging requirements further reduce multi-jets background – as one can see in figure 3.6, the cross section for $b\bar{b}$ production (which is approximately the subset of multi-jets background that contains true b quarks) is almost three orders of magnitude below the total multi-jets cross section. However due to its huge cross

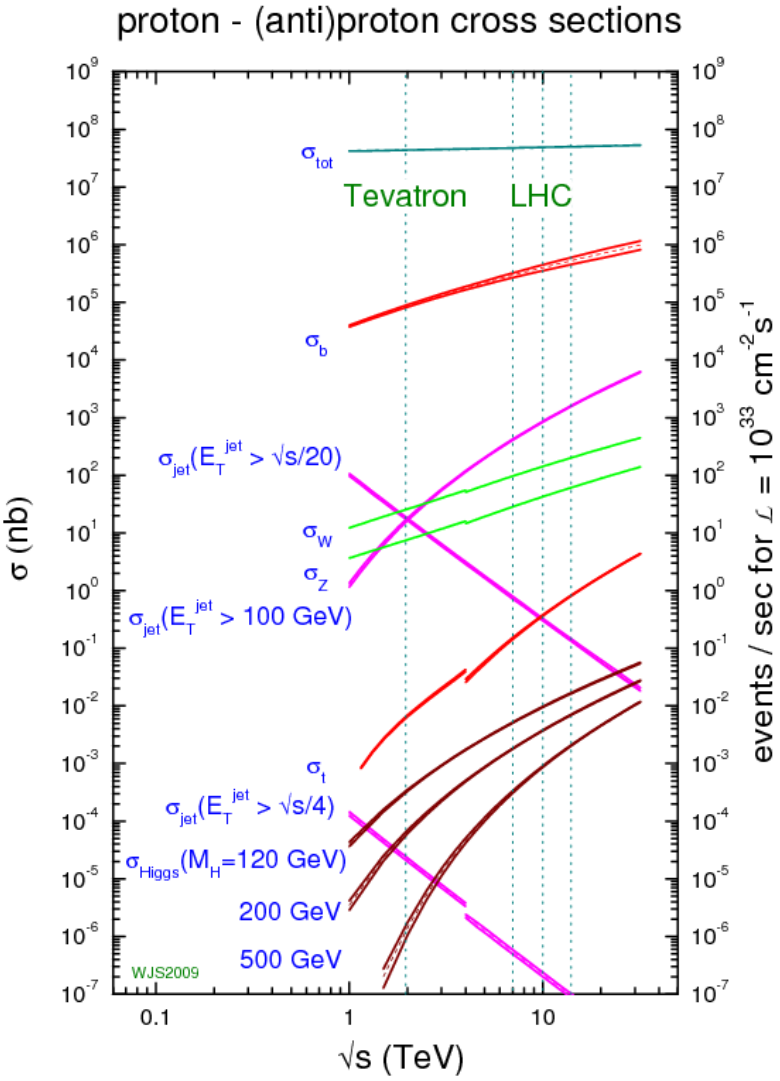


Figure 3.6: Total cross sections and production rates for an instantaneous luminosity of $\mathcal{L} = 10^{33}$ cm^{-2} s^{-1} for different processes at the LHC and the Tevatron as a function of the center-of-mass energy. Below the discontinuity at $\sqrt{s} = 4$ TeV, cross sections are for $p\bar{p}$ collisions, above they are for pp collisions. The dotted lines indicate the Tevatron energy of 2 TeV, the current LHC energy of 7 TeV, the intermittently foreseen LHC start-up energy of 10 TeV and the LHC design energy of 14 TeV. Figure taken from [42].

section, multi-jets background is generally still non-negligible, though it is normally among the minor backgrounds in top-quark analyses.

Another consequence of the cancellation between huge cross section and huge rejection by the selection, which is much more severe than the actual magnitude of multi-jets background after selection, is that it is virtually impossible to estimate the multi-jets background from simulation. Even given that the simulation described the data well, the number of events that one would need to generate in order to end up with sufficient statistics after event selection, would by far exceed available computing resources. Therefore multi-jets background generally needs to be estimated by data-driven methods for analysis (see Section 5.1.3), which always gives rise to large systematic uncertainties.

W+jets

W+jets background has the second-highest cross section among the background processes. In case the W boson decays leptonically, it contains two of the main signatures of the signal: the charged lepton and missing transverse momentum. Furthermore, the W may be produced in association with one or two b quarks. In this case the W+jets background contains all the main signal signatures, which makes the event selection based on these signatures not too efficient against this background.

Instead, most of the W+jets background gets suppressed by the choice of jet bins: the production of each additional jet is suppressed by a constant factor ("Behrends scaling"), therefore the W+jets production cross section drops exponentially with the number of additional jets. As a result, most W+jets production in fact happens in the zero- and one-jet bins, where only little signal is located. The large overall cross section and the signal-like signature still make W+jets production one of the two main background processes in the end.

Z+jets

Z+jets production has a cross section smaller but similar to W+jets production. As the Z boson can only decay into either zero or two leptons and furthermore typically does not produce considerable missing transverse momentum, it can be suppressed much better by the selection than W+jets. Additionally it also obeys the "Behrends scaling", i.e. it mostly populates the lowest jet bins. Therefore Z+jets production only poses a minor background.

Dibosons

Diboson production summarizes three different processes – WW, WZ and ZZ production – which have quite different signatures. ZZ production has similar characteristics to single Z production, as leptons can be produced only in pairs and $p_{\mathrm{T}}^{\mathrm{miss}}$ is typically small. WZ production has a somewhat similar final state to W+jets in case the W decays leptonically and the Z decays hadronically. WW production with one leptonically and one hadronically decaying W boson contains all signatures of the signal, except for the b-quark jet. In summary, two of the three diboson production modes can be expected to have a selection efficiency at least of the order of W+jets. But as diboson production has a much smaller cross section than single boson production, it only poses a minor background.

t-channel production

The cross section for t-channel production is about four times as large as that of the Wt-channel. In case of a leptonically decaying top quark it also has all the main signal signatures: the lepton, $p_{\mathrm{T}}^{\mathrm{miss}}$ and b-quark jet. It has, however, only one light-quark jet, such that the main jet bin of the t-channel is the

two-jet bin, followed by the one- and three-jet bins. In the end, t-channel production is among the minor backgrounds.

s-channel production

Like t-channel production, also s-channel production has all the main signal signatures in case of a leptonically decaying top quark. Its cross section is only less than a third of the Wt-channel cross section, however. The s-channel has no light-quark jets, but a second b-quark jet instead. Vetoing a second b-quark jet – which is mostly aimed at reducing $t\bar{t}$ (in particular at reducing potential interference, see Section 3.4.1) – therefore provides some additional suppression. Overall, the s-channel is the smallest of all backgrounds.

Top quark pair production

The final state of top quark pair production contains the final state of Wt-channel production plus an extra b quark. In addition, $t\bar{t}$ production might even interfere with Wt production (see Section 3.4.1). Therefore it is clear that this is the background channel which is most difficult to separate from signal. The most important handle against $t\bar{t}$ production is the veto on an additional b-quark jet. With a cross section of about ten times the signal cross section, it is not only the hardest background to get rid of, but also among the two largest backgrounds after selection.

Chapter 4

Reconstruction

As stated in Chapter 3, the final state of interest in the scope of this thesis contains one – presumably hard and isolated – electron or muon, a neutrino giving rise to missing transverse momentum, two light-quark jets and one b-quark jet. In order to make use of this rich signature in analysis, one needs to properly reconstruct all these objects from the raw detector data.
In this chapter, the relevant reconstruction algorithms from the ATLAS software that were used by the reconstruction in the scope of this thesis are described briefly; for the objects used for analysis also some performance benchmarks are given. References to more detailed descriptions and performance studies of the algorithms are given in the individual sections. The information presented here summarizes the relevant parts of these sources.

4.1 Tracking

Tracking in the scope of this thesis basically enters in the electron and muon reconstruction as well in b-tagging. The track reconstruction in ATLAS is explained in great detail e.g. in [43] and [44]. The latter also contains studies of the expected performance. The tracking performance is studied with 900 GeV data in [45]. A brief summary of the standard inside-out track reconstruction sequence as described in the given references is presented in the following.

Before the actual track reconstruction, the raw data in the detectors is transformed to hits that can be used for tracking. Those are called "clusters" for the silicon trackers and "calibrated drift circles" for the TRT, but for the sake of simplicity both are just referred to as *hits* in the following. The default primary tracking sequence, which is the mentioned inside-out sequence, then starts with a track finding in the silicon detectors. First, measured space points are combined to track seeds, where a measured space-point denotes either a single hit in the Pixel or a pair of SCT hits on both sides of the same SCT module. Further silicon hits are then picked up by a Kalman Filter.[46] The resulting track candidates are ranked by a sophisticated scoring algorithm in order to resolve overlapping tracks and reject fake, incomplete and badly reconstructed tracks. Only the remaining tracks are extrapolated to the TRT, where additional hits are picked up. These extended tracks are refitted using the hits from all three ID components and scored again. If the refitted extended track has a lower score than the original silicon-only track, the original track is kept instead and the TRT hits are only associated as outliers, i.e. they are still assigned to the track but not used for the fit. Subsequent to this inside-out sequence an outside-in sequence is applied using the remaining hits, followed by another inside-out sequence again using the hits left then, which is mainly aimed at very low-p_{T} tracks that may not reach the TRT. Both are not discussed here in any more detail.

The track reconstruction software is written in such a modular and flexible way that it is not only capable of fitting Inner Detector tracks, but can also be used to fit tracks in the Muon Spectrometer or even combined Inner Detector and Muon Spectrometer tracks. This ability is used in the reconstruction of combined muons (see Section 4.4).

4.2 Calorimetry

The reconstruction of jets, missing transverse momentum and electrons is based on calorimeter measurements. In this section, two general aspects of calorimeter-based reconstruction are discussed, that are relevant for all of these: clustering and energy scale calibration.

4.2.1 Clustering

In any reconstruction of objects from calorimeter measurements, the calorimeter cells first need to be grouped to clusters, which are then typically assumed to correspond to individual particles that hit the calorimeter. These clusters are then used e.g. as inputs to jet algorithms or for missing transverse momentum reconstruction. Several different clustering algorithms are used in ATLAS. Detailed descriptions of all of them can be found in [44]. Those that were used by the reconstruction in the scope of this thesis are briefly introduced in the following.

Topological clusters

Topological clusters are the default input objects for jet algorithms and for missing transverse momentum reconstruction in the scope of this thesis. Topological clusters are three-dimensional objects built from the calorimeter cells. Clusters are seeded by cells with a signal-to-noise ratio of more than four. Then all directly neighboring cells (in all three dimensions) with a signal to-noise-ratio of more than two are iteratively added to the cluster, until none are left. Finally, only once all directly neighboring cells are added regardless of their signal-to-noise ratio. Clusters that have local maxima are split by a splitting algorithm.

Sliding window

The sliding window is a rather simple algorithm, which is basically used for electron/photon reconstruction in ATLAS. A rectangular window of fixed size is placed such that the energy inside the window gets maximized. The cells within this window then form the cluster.

4.2.2 Energy scale calibration

Energy scale calibration is a fundamental aspect of calorimetry. The main issue here is that electromagnetic showers and hadronic showers, which are the two ways how particles deposit their energy in the calorimeter, have different detector responses. That is, an electromagnetic shower and a hadronic shower initiated by two particles of identical energy lead to different measured signal strengths in the calorimeter. Therefore both types of showers need to be calibrated separately. This implies that for precise measurements the shower type must be known. In case of electrons and photons, which mostly give rise to purely electromagnetic showers, this is comparably easy. In case of hadrons this is much more difficult, as hadronic showers in general also contain an substantial electromagnetic part due to e.g. neutral pions produced in the shower which immediately decay into two photons. This electromagnetic fraction is also subject to strong statistical fluctuations, i.e. it strongly varies from shower to shower.

The baseline energy scale for calorimeter measurements in ATLAS is the electromagnetic (EM) scale, i.e. the calibration for electromagnetic showers. Any calorimeter measurement is first calibrated to this scale. Calibration to the hadronic scale is mostly needed for jets, but also for the measurement of transverse missing energy. In fact the hadronic calibration schemes in ATLAS mostly rather aim at

calibrating to the jet energy scale (JES), which is not exactly the same as the straight hadronic scale, as the jet energy scale needs to take into account things like some constituents of the jet not reaching the calorimeter, while the hadronic scale just calibrates to the energies of the individual hadrons that hit the calorimeter. Basically three different approaches for jet energy scale calibration are employed in ATLAS: the *EM+JES* scheme, the *Global Cell Weighting* (GCW, formerly also referred to as "H1-style" calibration) and the *Local hadronic calibration* (or Local Cell Weighting, LCW). Only the latter is actually able to also produce a straight hadronic scale. For jets, additional corrections may be applied on top of any of the schemes (see Section 4.5). Detailed descriptions of all schemes (including an extension of the EM+JES scheme called *Global Sequential Calibration* (GSC), which is not mentioned any further here) can be found in [47], a brief description is given in the following.

EM+JES calibration. The EM+JES scheme is the simplest of all three approaches, originally designed for the commissioning phase. It directly calibrates from the EM scale to the jet energy scale, applying a JES correction factor to each each object (i.e. jet or, in the case or missing transverse momentum, topological cluster), which depends only on the object's pseudorapidity and transverse momentum. In the scope of this thesis the EM+JES was used for both jets and missing transverse momentum.

Global cell weighting. The GCW first applies to each jet a correction that only depends on the energy density in its constituent cells, basing upon the fact that electromagnetic showers are much denser than hadronic showers. Afterward a JES correction factor depending on η and p_{T} (similar to the EM+JES calibration) is applied in order to correct for detector effects.

Local hadronic calibration. The LCW uses shower shape variables in order to calculate a probability for each cluster to be of electromagnetic or hadronic nature. Clusters are then calibrated to the hadronic scale by applying weights based on these probabilities. Further corrections are applied one by one on top of this ("out-of-cell" correction, dead material correction) to reach a straight hadronic calibration as defined above. Only then is the final jet level calibration applied in order to reach the jet energy scale.

4.3 Electrons

Electron finding in ATLAS is two-staged. First the *electron reconstruction* provides a set of electron candidates, which is mostly optimized for efficiency. Besides from "real" electrons from different sources, these electron candidates also contain a substantial amount of "fake" electrons, i.e. electron candidates that arise from other physical objects – mostly hadrons – which are misidentified as electrons. The second step, denoted *electron identification*, then tries to identify the real electrons out of these candidates and reject as many of the fake electrons as possible. In many analyses – like the one presented in this thesis – one is basically interested in isolated electrons that are produced at the primary vertex. In this sense other sources of true electrons are considered as background, too, and hence the electron identification tries to reject them as well. This holds in particular for electrons from photon conversions and Dalitz decays, but also for those from semileptonic decays of charmed or bottom hadrons. Such electrons are denoted *background electrons* in the following.

Both the electron reconstruction and the electron identification, including all respective algorithms that are available in the ATLAS software, are described in detail in [48], [49] and, even more exhaustive but based on an older version of the ATLAS software, in [44].

4.3.1 Electron reconstruction

Three algorithms are available for electron reconstruction in ATLAS: the *standard electron* algorithm, the *soft electron* algorithm and the *forward electron* algorithm. The standard algorithm is a calorimeter-based algorithm, that aims mostly at isolated electrons with high transverse momentum. The *soft electron* algorithm is based on Inner Detector tracks and aims mostly at low-p_{T} electrons. The *forward electron* algorithm reconstructs electrons in the very forward region, outside the acceptance of the Inner Detector.

Only standard electrons are used in this thesis. Standard electrons are seeded by a cluster in the electromagnetic calorimeter, reconstructed by the sliding window algorithm (see Section 4.2.1). These clusters are then matched to Inner Detector tracks. The energy of the electron candidates is given by a weighted mean of the cluster energy and the track momentum. The full four-momentum is then constructed using track η and ϕ (in case of TRT-only tracks η is rather provided by the cluster instead) and applying zero mass.

4.3.2 Electron identification

Different electron identification techniques are available in order to reject both fake electrons and background electrons. The baseline electron identification in ATLAS uses cuts on several different discriminating variables. Among those are pure calorimeter quantities such as shower shapes and hadronic leakage, pure Inner Detector quantities such as track quality requirements or TRT high-threshold hits as well as quantities from the track-cluster matching, such as the η-ϕ distance between the track and the cluster and the ratio between cluster energy and track momentum. Three reference sets of cuts are defined for use in analyses, denoted "loose", "medium" and "tight" in the order of increasing background rejection and decreasing efficiency. Cuts on quantities that employ the Inner Detector are used starting from the "medium" cuts. In particular the presence of a matched track is required for these categories.

Other electron identification techniques available in the ATLAS software employ different multivariate techniques instead of cuts. These were not use in the scope of this thesis, however, nor are they commonly used for (top-quark) analyses in general.

Electron isolation

Isolation criteria are not used in the electron identification as described so far. Nevertheless, isolation criteria are an important tool in order to discriminate against electrons from semileptonic decays, which are least suppressed by the electron identification with respect to the isolated high-p_{T} electrons we are looking for. This holds not only for top-quark analyses in general, but also for any other analyses that have isolated electrons as part of their signal (in particular any electrons from W or Z decays). However, as the required isolation criteria are considered to be analysis dependent, these are rather applied on top of the electron identification cuts. Two main types of variables are available for building isolation requirements: $E_{\mathrm{T,\,cone}}(\mathrm{XY})$ and $p_{\mathrm{T,\,cone}}(\mathrm{XY})$. The $E_{\mathrm{T,\,cone}}(\mathrm{XY})$ variables denote the reconstructed transverse energy in a cone of half opening angle $\Delta R = 0.\mathrm{XY}$ around the electron, excluding the electron itself. The $p_{\mathrm{T,\,cone}}(\mathrm{XY})$ variables denote the scalar p_{T} sum of all tracks in a cone of half opening angle $\Delta R = 0.\mathrm{XY}$ around the electron that obey some quality cuts.

The barrel/end-cap transition region

In the transition region between barrel and end-cap of the electromagnetic calorimeter, in the small pseudorapidity region of $1.37 < |\eta| < 1.52$, there is a huge spike in the distribution of dead material

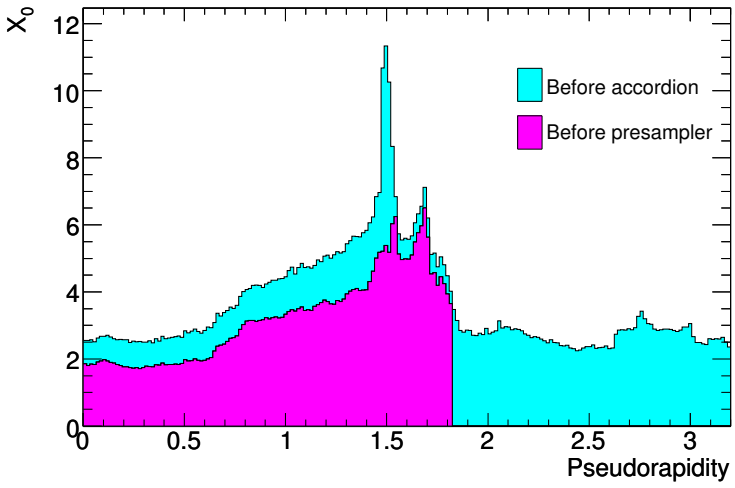


Figure 4.1: Distribution of the amount of material in front of the electromagnetic calorimeter (denoted "accordion" in the legend") in units of radiation length X_0, as a function of $|\eta|$.

$E_T > 20$ GeV, $\|\eta\| < 2.5$							
	Efficiency (%)		Jet rejection (total)	Surviving candidates (%)			
	$Z \to ee$ ±0.03	$b, c \to e$ ±0.5		iso	non-iso	bkg	had
Reconstruction	97.58	-	91.5 ± 0.1	0.1	0.8	23.3	75.8
Loose	94.32	36.7	1065 ± 5	0.9[1]	1.9	56.7	40.4
Medium	90.00	31.5	$(6.84 \pm 0.07) \cdot 10^3$	6.0	9.9	50.5	33.4
+ isol. 99%	88.87	-	$(10.3 \pm 0.13) \cdot 10^3$	9.2	7.6	56.2	27.0
+ isol. 98%	87.99	-	$(13.6 \pm 0.20) \cdot 10^3$	11.7	5.8	58.7	23.8
+ isol. 95%	85.06	-	$(20.0 \pm 0.35) \cdot 10^3$	16.0	3.5	60.5	20.0
+ isol. 90%	80.67	-	$(27.1 \pm 0.55) \cdot 10^3$	16.9	2.6	60.7	16.8
Tight	71.59	25.2	$(1.39 \pm 0.06) \cdot 10^5$	29.9	44.9	11.4	13.8
+ isol. 99%	70.78	-	$(1.98 \pm 0.11) \cdot 10^5$	42.3	32.7	12.5	12.5
+ isol. 98%	70.16	-	$(2.50 \pm 0.15) \cdot 10^5$	51.6	24.1	12.3	12.0
+ isol. 95%	68.05	-	$(3.79 \pm 0.28) \cdot 10^5$	65.5	13.3	11.7	9.5
+ isol. 90%	64.83	-	$(5.15 \pm 0.45) \cdot 10^5$	73.5	8.3	11.0	7.2

Table 4.1: Courtesy of [49] (incl. this caption, except for minor adjustments): Expected isolated-electron efficiencies, non-isolated-electron efficiencies and jet rejections for the standard sets of identification cuts and and E_T-thresholds of 20 GeV and $|\eta| < 2.5$ (including the barrel/end-cap transition region). The total jet rejection includes hadron fakes and background electrons from photon conversions and Dalitz decays. The last four columns give the fraction of surviving electron candidates in a filtered dijet sample after each selection level. The categories considered for the electron candidates are described in the text. The efficiencies are computed on a $Z \to ee$ sample and rejections are computed on the dijet sample. The quoted errors are statistical.

that a particle has to traverse before the calorimeter (see figure 4.1 [23]). This leads to a significantly reduced performance (regarding both efficiency and energy resolution) in this region. Performance studies therefore often exclude this region from the reported results, and in analyses it is also left out accordingly.

4.3.3 Performance

Reconstruction efficiency and background rejection

The expected performance of the electron reconstruction and identification is summarized in table 4.1.[1] [49] The efficiencies in the two left-most columns have been determined using an inclusive $Z \rightarrow ee$ MC sample. The rejections in the third column refer to not only fake electrons but also include background electrons from Dalitz decays and photon conversions. They were computed on a filtered inclusive dijet sample generated with PYTHIA, that contains all hard QCD processes, heavy flavor production, prompt photon production, and W and Z production. The fractions of surviving candidates in the four right-most columns are also based on the dijet sample. The used classification of electrons is based on truth classification (using a common ATLAS tool), the classes are defined as follows:

- isolated electrons ("iso"):
  electrons that match a true electron from a W or Z decay;
- non-isolated electrons ("non-iso"):
  electrons that match a true electron from a charmed or bottom meson;
- background electrons ("bkg"):
  electrons that match a true electron from a Dalitz decay or a photon conversion;
- fake electrons ("had"):
  electrons that do not match a true electron, muon or tau.

Energy resolution

In [48] the electron energy resolution and energy scale uncertainty were determined from real data using W, Z and J/ψ decays. In the central region (i.e. $|\eta| < 2.47$) the systematic uncertainty on the energy scale amounts to $0.3 - 1.6\%$, depending on the pseudorapidity. For the two pseudorapidity regions where the resolution is highest and lowest, the energy scale uncertainty as a function of the electron E_{T} is shown in figure 4.2. The relative electron energy resolution is parametrized as $\sigma(E)/E = a/\sqrt{E} \oplus b/E \oplus c$. Only the constant term c was determined, amounting to about 1.2% and 1.8% in the barrel ($|\eta| < 1.37$) and end-cap ($1.52 < |\eta| < 2.47$) regions, respectively. In [50] the electron performance was studied on 900 GeV data and the performance at higher energies was estimated. From the results shown there one can deduce that the expected electron energy resolution is in general between 1% and 3% for most pseudorapidity and energy regions and at most 8%, at an electron energy of 25 GeV and close to the transition region. These numbers however refer to the energy resolution from the calorimeter only and do not take into account tracking information, which is expected to improve the resolution at lower energies.

[1] In the original table in [49] the fraction of surviving isolated electrons in Loose selection is actually 0.3. I believe that that number is a typo. On the one hand the sum of fractions of surviving candidates would be only 99.3% then, furthermore it would lead to considerable inconsistencies between numbers for loose, medium and tight cuts. Both issues get resolved by the assumption that the correct value is 0.9.

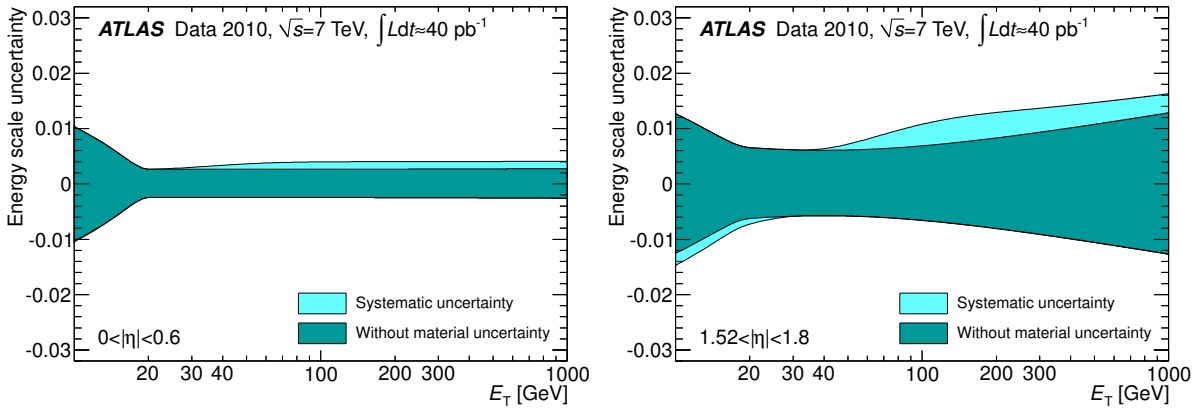


Figure 4.2: Total uncertainty on the electron energy scale depending on E_{T} for two different η regions. The left plot shows the region $|\eta| < 0.6$, which has the smallest uncertainty, the right plot shows the region $1.52 < |\eta| < 1.8$, which has the largest uncertainty.

4.4 Muons

Most information given in this section is a small extract from [44], where a detailed (though in some details outdated) description of all muon reconstruction algorithms can be found.

In ATLAS there are three types of reconstructed muons: *standalone* muons, *combined* muons and *tagged* muons. Standalone muons consist of a reconstructed track in the muon spectrometer, which gets extrapolated to the beam line. Combined muons are created by combining a standalone muon with a matching (i.e. nearby) Inner Detector track. For tagged muons, Inner Detector tracks are extrapolated to the muon system and get "tagged" either with matching segments in the muon system ("segment tagging", note that a segment is less than a full reconstructed track in the muon system) or with matching signals from minimum ionizing particles in the calorimeter ("calorimeter tagging").

Calorimeter-tagged muons are treated separately from all other muon types by the ATLAS software (one dedicated reconstruction algorithm, candidates are stored in a separate collection) and are not considered in the following. For all other types of muons, two separate chains of algorithms exist in the ATLAS software, each containing one algorithm for each type of muon. Both chains are named after their respective algorithm for reconstructing combined muons: MuID [51] and StaCo. By now MuID has become standard for analyses (at least this is the case within the top working group). More specifically, for top-quark analyses in general, only combined muons from MuID are used. MuID employs a combined refit of the Inner Detector and the muon system track. The muon reconstruction provides three quality flags named "loose", "medium" and "tight", similar to the reference sets of cuts in the electron reconstruction; MuID combined muons by definition all fulfill the "tight" quality. For muons the same isolation variables (i.e. $E_{\mathrm{T,\,cone}}(XY)$ and $p_{\mathrm{T,\,cone}}(XY)$) as for electrons are available.

Performance

The transverse momentum resolution for muons was studied in [52], using cosmic muons. For combined muons it is parametrized as

$$\frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}} = P_1 \oplus \frac{P_0 \times p_{\mathrm{T}}}{\sqrt{1 + (P_3 \times p_{\mathrm{T}})^2}} \oplus P_2 \times p_{\mathrm{T}} \,.$$

In this parametrization P_1 refers to multiple scattering, which dominates at lowest momenta. In this region, the resolution is mostly achieved by the Inner Detector. The parameter P_2 refers to the intrinsic resolution, dominating at very high momenta. In this region, the resolution is driven mainly by the Muon System. The P_3 term describes the intermediate region from about 50 to 150 GeV, where both systems together achieve the best resolution. Only the parameters P_1 and P_2 are determined in [52], the values are 3.8% and 2×10^{-4} GeV^{-1}, respectively. The actual relative momentum resolution amounts to about 2%, 3.5% and 9% at muon transverse momenta of 25 GeV, 100 GeV and 400 GeV.

The muon reconstruction efficiency is studied in [53], based on Z decays in real data, and is "found to be well above 92% for combined muons".

4.5 Jets

A wide variety of jet collections is available from the ATLAS reconstruction, which differ in

- the used jet algorithm (both algorithm type and jet size parameter);
- the type of input objects to the jet algorithm;
- the energy scale calibration.

The standard jet algorithm used for analyses is the Anti-k_T algorithm with a size parameter of 0.4 or 0.6. The default type of input objects used in top-quark analyses are topological clusters. The current baseline calibration scheme used within the top working group is EM+JES (see Section 4.2.2). On top of that calibration, corrections are applied to take into account contributions due to pileup and to adjust the position/direction of the jet such that it points back to the primary vertex (rather than to the origin of the coordinate system[2]).

Performance

Jet energy resolution. The jet energy resolution has been studied in [54]. The relative jet p_T resolution is parametrized as

$$\frac{\sigma_{p_T}}{p_T} = \frac{N}{p_T} \oplus \frac{S}{\sqrt{p_T}} \oplus C$$

there, where N represents detector and electronics noise as well as energy offset due to multiple interactions, S represents the statistical fluctuations of the energy deposit of particles in the calorimeters and C represents any fluctuations that are a constant fraction of the jet energy. Table 4.2 shows the values of these parameters that were determined for Anti-k_T jets with $R = 0.4$ and the EM+JES calibration in different rapidity bins. For the lowest and the highest rapidity bin – representing the worst and the best resolution, respectively – this amounts to resolutions of 25%, 13%, 7.5% and 20%, 11%, 6.8% for a jet p_T of 25 GeV, 100 GeV, 400 GeV, respectively.

Jet energy scale uncertainty. The uncertainty on the jet energy scale for Anti-k_T jets has been studied in [47] using both real data and MC. Estimations on the uncertainty are given there in bins of $|\eta|$ for different example values of jet p_T for both $R = 0.4$ and $R = 0.6$, and as a function of jet p_T for a few example $|\eta|$ bins only for $R = 0.6$. Table 4.3 shows the determined uncertainties for $R = 0.4$ in those η bins that are relevant for this thesis, at $p_T = 20$ GeV and $p_T = 200$ GeV. Figure 4.3 shows the uncertainties as function of p_T for $0.3 < |\eta| < 0.8$ and $2.1 < |\eta| < 2.8$.

[2] The actual beam spot is slightly displaced with respect to the nominal one.

Rapidity range	S	N	C
$0.0 < \|y\| < 0.8$	1.14±0.02	2.1±0.2	0.048±0.005
$0.8 < \|y\| < 1.2$	1.12±0.03	1.6±0.2	0.045±0.008
$1.2 < \|y\| < 2.1$	1.07±0.03	1.9±0.2	0.05 ±0.01
$2.1 < \|y\| < 2.8$	0.93±0.03	1.1±0.1	0.05 ±0.05

Table 4.2: Resolution parameters for Anti-k_T jets with $R = 0.4$ and EM+JES calibration as determined in [54].

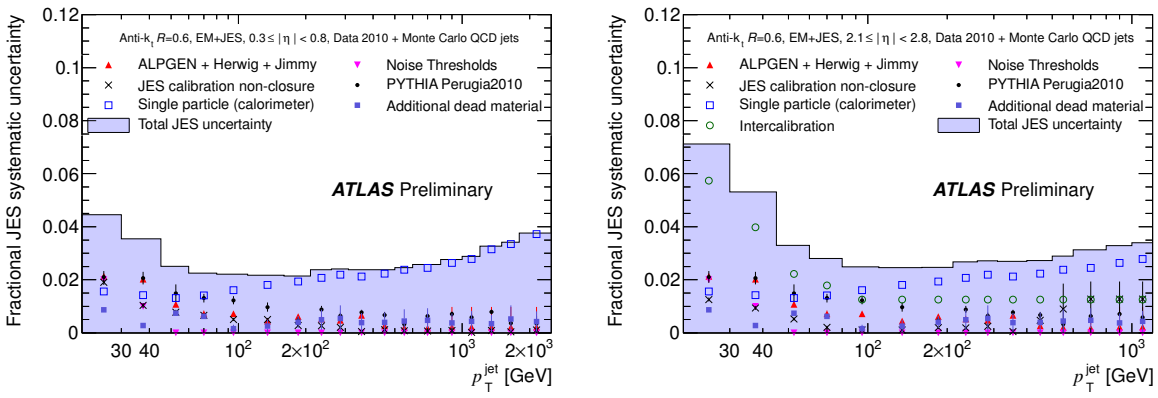


Figure 4.3: Uncertainty on the jet energy scale for Anti-k_T jets with $R = 0.6$, depending on the jet p_T for the two η regions $0.3 < |\eta| < 0.8$ (left) and $2.1 < |\eta| < 2.8$ (right), serving as examples for barrel and end-cap.

η range	Max. rel. JES uncertainty	
	$p_T = 20$ GeV	$p_T = 200$ GeV
$0.0 < \|\eta\| < 0.3$	4.1%	2.3%
$0.3 < \|\eta\| < 0.8$	4.3%	2.4%
$0.8 < \|\eta\| < 1.2$	4.4%	2.5%
$1.2 < \|\eta\| < 2.1$	5.3%	2.6%
$2.1 < \|\eta\| < 2.8$	7.4%	2.7%

Table 4.3: Summary of the maximum EM+JES jet energy scale systematic uncertainties for different p_T and η regions for Anti-k_T jets with $R = 0.4$.[47]

4.5.1 *b*-tagging

A variety of *b*-tagging algorithms are available in the ATLAS reconstruction. A recent overview focusing on what is called the "high-performance" or "advanced" taggers in the ATLAS jargon, including performance studies, is given in [55], a more comprehensive but somewhat outdated overview is given in [44].

The current default for top-quark analyses, which is also the one that was used in the scope of this thesis, is the so-called "JetFitterCombNN" tagger, which is a neural-network-based combination of the "JetFitterTagNN" algorithm and the "IP3D" algorithm. The IP3D algorithm uses a likelihood ratio formalism that is common to many *b*-tagging algorithms in ATLAS, using the two-dimensional distribution

of the transverse versus the longitudinal impact parameter significance as input.[44] The JetFitterTagNN algorithm employs a neural network, (which is not related to the one that combines JetFitterTagNN and IP3D to JetFitterCombNN) using input variables provided by the JetFitter algorithm. The JetFitter is a sophisticated Kalman-Filter style vertexing algorithm that fits at the same time a primary vertex, the intended B-hadron flight axis (starting from that primary vertex) and multiple vertices, all lying on this flight axis, which are supposed to represent the B-hadron decay chain. A detailed description of the original JetFitter algorithm is given in [56], while the construction of the b-tagging weight from the output of the JetFitter algorithm given there is deprecated by now.

Working points. Output from b-taggers is normally used for analyses in ATLAS in the form of particular *working points* that are provided by the ATLAS flavor tagging working group. A working point basically consists of a cut value on the b-tagger output, above which a jet is considered b-tagged, and nominal values for the efficiency ϵ and for the light-quark rejection factor R of the tagger at this cut value. Efficiency here denotes the probability for a true b-quark jet to be tagged at this working point, and the rejection is the inverse of the probability for a true light-quark jet to be wrongly tagged (i.e. the number of light-quark jets that are rejected per light-quark jet that is tagged). The nominal efficiency and the rejection factor are average values, determined from MC, that are to be taken as benchmark values; in reality both are functions of jet η and p_{T}.

4.6 Missing transverse momentum

A recent detailed description of how missing transverse momentum is reconstructed in ATLAS as well as performance studies can be found in [57], older studies can be found in e.g. [58] and [44]. The missing transverse momentum reconstruction can be summarized from these sources as follows.

First the vector sum of all relevant calorimeter measurements is calculated. By default these are given by topological clusters (see Section 4.2.1). The clusters get associated to reconstructed physics objects as far as possible and calibrated according to these objects. The association is done in a defined order: electrons, photons, hadronically decaying τ-leptons, jets, muons. Double counting of cells is avoided by using each calorimeter cell for only one type of object; in case of several objects of the same type sharing a cell the energy in the cell is parted up. The association of electrons is driven by a choice of electron identification cuts: the default in the top working group is to use the same as for signal electrons (i.e. "tight" cuts), in contrast to the general default, which is to use "medium" electrons. Cells that can not be associated to any object are still taken into account, this is called the "cell-out" term. The calibration of the individual object types can be done in various ways. In the scope of this thesis, the following calibrations were used, following the top working group default choice:

- electrons: default electron calibration;
- photon, tau leptons, soft (i.e. $p_{\mathrm{T}} < 20$ GeV) jets, cell-out term: standard EM scale;
- jets ($p_{\mathrm{T}} > 20$ GeV): EM+JES calibration.

This is in contrast to recent performance studies, which usually rather use LCW calibration for any hadronic objects.

What has been described so far is normally referred to as the "calorimeter term". This does not yet take into account the contribution from muons, which deposit only a very small fraction of their energy in the calorimeter. The muon contribution is therefore added on top of the calorimeter term. This can happen in two ways. For isolated muons, the combined ID and muon spectrometer measurement is used,

and the energy deposited in the calorimeter by the muon is removed from the calorimeter term in order to avoid double counting of that energy. For non-isolated muons the energy deposited in the calorimeter can not be resolved from the surrounding deposits and therefore this procedure cannot be applied. For these muons the p_{T} measurement from the muon spectrometer alone is used, as this automatically takes the energy loss of the muon in the calorimeter into account.

The main variables that are produced by the missing transverse momentum reconstruction are the two-dimensional $p_{\mathrm{T}}^{\mathrm{miss}}$ vector and the total transverse momentum $\sum p_{\mathrm{T}}$, which is just the scalar sum of all contributions to $p_{\mathrm{T}}^{\mathrm{miss}}$.

Performance

The performance of the missing transverse momentum reconstruction was studied e.g. in [57]. It was found there that the resolution, σ, of the x and y components of the missing trasverse momentum can be parametrized as $\sigma = k \cdot \sqrt{\sum E_T}$, with $k \approx 0.5$ GeV$^{1/2}$. The average overall uncertainty of the $p_{\mathrm{T}}^{\mathrm{miss}}$ scale was estimated to be 2.6 GeV in $W \to e\nu$ and $W \to \mu\nu$ events.

4.7 Luminosity

A variety of different techniques for measuring the luminosity has been considered in ATLAS. A detailed report on all originally considered techniques can be found in [59]. An excellent and detailed description of the general idea behind the techniques that have been applied so far, as well as the concrete luminosity determination for the 2010 data, is given in [60]. The luminosity determination for the 2011 data-set is described in detail in [61]. All the following explanations are taken from these sources.

All luminosity determination techniques that have been employed in ATLAS so far are based on measuring visible event rates, i.e. the *fraction of bunch crossings that contain at least one interaction which satisfies some given selection criteria.* In the following, the term "(measured) event rate" is generally used in precisely this sense. Seveal different algorithms, i.e. selection criteria, each connected to one or more detector components, have been employed. For the 2011 running period, mainly two detector components have been used in this scope: the LUCID and the BCM (see Section 2.2.4), which have in common that they consist of two stations at equal distance to the interaction point, one on each side. Two algorithms were used for both the LUCID and the BCM: an "Event_OR", where at least one hit in one of the stations is required, and an "Event_AND", where at least one hit in both of the stations is required (i.e. coincidences are counted). Each algorithm provides an independent event (rate) count. For the BCM, the horizontal and the vertical sensor pair are read out separately, and they were used as separate counters, denoted BCMH and BCMV. This leads to a total of 6 different algorithms (i.e. measured event rates). Using different detectors and algorithms at the same time and comparing them to each other is an integral part of how ATLAS wants to assess and control systematic uncertainties of luminosity measurements.

In order to relate the visible event rates to the luminosity, one can write the luminosity as

$$\mathcal{L} = \frac{R_{\mathrm{vis}}}{\sigma_{\mathrm{vis}}} = \frac{\mu^{\mathrm{vis}} n_b f_r}{\sigma_{\mathrm{vis}}}, \tag{4.1}$$

where σ_{vis} is the visible cross section for interactions that satisfy the given selection criteria and R_{vis} is the rate of such visible interactions,[3] which can be expressed by the average number of visible in-

[3] Note that here rate denotes interactions per unit of time, in contrast to the measured event rate, that was defined above as a fraction of bunch crossings

teractions per bunch crossing μ^{vis}, the number of bunch crossings per revolution n_b and the machine revolution frequency f_r. It is important to note that μ^{vis} is not exactly the same as the measured event rate: while the latter is bounded above by one, the former is in principle only bounded by the total number of collisions per bunch crossing. If, and only if, $\mu^{\text{vis}} \ll 1$, it can, however, be directly approximated by the measured event rate. Otherwise one needs to take into account that in a single bunch crossing more than one visible interaction may occur. In this case, which is considered to be the normal case at the LHC, Poisson statistics needs to be employed in order to calculate μ_{vis} from the measured event rates. The exact procedure depends on the algorithm type (Event_OR or Event_AND) and is described elsewhere.[60] This procedure is only consistent if the selection criteria are defined such that they only depend on whether at least one visible collision takes place during a bunch crossing or not, but not on the multiplicity of (visible) interactions within the bunch crossing.[4] Furthermore the signal definition should be chosen such that the visible cross section gives rise to a "moderate" visible bunch crossing rate over a broad luminosity range. "Moderate" means that for too high visible cross sections the measured event rate is approximately one, independent of the luminosity, and the technique becomes insensitive. When the visible cross section and therefore the rate is too low instead, statistical uncertainties become an issue.

In order to calculate the luminosity from equation 4.1 once μ^{vis} has been determined, one needs to know the visible cross section. This is generally not the case. The approach that is used to obtain it, is by determining the luminosity from machine parameters during dedicated runs, so-called *van-der-Meer* (vdM) scans (also called *beam-separation* or *luminosity* scans), where at the same time the event rates are measured as discussed above. Then the visible cross section is the only unknown in equation 4.1 and can be calculated. This calibrates the algorithms for use in normal physics runs, for which then just the visible cross section that was measured in the vdM scans needs to be inserted into equation 4.1 in order to obtain an absolute luminosity measurement. The uncertainty on the luminosity determined by this procedure was estimated to be about 3.7% for 2011 data, dominated by uncertainty on the beam current (3%), which is one central ingredient to the vdM scan, and followed by the estimated long-term stability and the dependence on the actual average number of visible interactions per bunch crossing (1% each). The algorithm that was chosen as the reference for physics results in 2011 is the BCMH Event_OR algorithm.

Luminosity blocks. The smallest piece of data, for which an integrated luminosity is determined in ATLAS is called a luminosity block. It corresponds to in the order of one to two minutes of data taking. The total integrated luminosity for any amount of data is then simply the sum of the luminosity blocks.

[4] This means it should e.g. be ruled out that two collisions only pass the selection criteria when they occur together in the same bunch crossing, while none of them would pass the selection on its own (such things may happen when e.g. too high thresholds are involved).

Chapter 5

Analysis setup: datasets, selections and general strategy

After all the necessary ingredients have been discussed in the previous chapters, this chapter is now devoted to the setup of the actual analysis. The first section presents all the datasets that were used, the two subsequent sections define the event selections. Finally, based on the outcome of these selections, the general analysis strategy is outlined.

5.1 Used datasets

In this section, the datasets that were used in the scope of this thesis are discussed. The signal and most background processes were estimated by MC simulated datasets. For the multi-jets background a data-driven estimation was used.

5.1.1 Real data

Data stream and trigger selection. Two distinct sets of real data were used for the electron channel and the muon channel. For the electron channel, data from the "Egamma" stream, selected with an electron trigger, was used; for the muon channel, data from the "Muons" stream, selected with a muon trigger, was used. In each case the unprescaled trigger with the lowest energy/momentum threshold was chosen, in order to maximize the rate and the acceptance. For the data that was used in the scope of this thesis, this was the "EF_e20_medium" and "EF_mu18" trigger for the electron and the muon channel, respectively. The numbers in the trigger names denote the required p_{T} threshold in GeV.

Good run lists. Before some data that was taken is signed off for analysis, Data Quality (DQ) flags are assigned that indicate whether detector, reconstruction etc. were performing well during data taking and hence the data is suitable for analysis. Such flags exist for all the individual detector components, for the triggers, and for reconstructed objects (electrons, muons etc., see Chapter 4), where e.g. the flags for reconstructed objects can depend on the relevant detector flags. The DQ flags are then combined in *good run lists* (GRLs), which specify all the luminosity blocks in a defined set of runs for which the corresponding data is supposed to be suitable for analysis. In addition, the GRL also contains the corresponding total luminosity of all the specified good luminosity blocks. For top-quark physics analyses in ATLAS, GRLs are defined and provided by the top reconstruction group. The list of required DQ flags comprises some general flags that are needed for any analysis, as well as all the detector components and all reconstruction that is normally used for top-quark physics analyses. Concretely, flags for global detector status, luminosity determination, the magnet system, trigger and reconstruction of electrons, muons and jets, reconstruction of missing transverse momentum as well as some flags related to tracking, vertexing and b-tagging are required.

Used data-taking period. The data that was used in the scope of this thesis was taken between end of March and end of June 2011. After selection of good runs (i.e. after applying the relevant GRL), this dataset corresponds to an integrated luminosity of 1035 pb^{-1}.

5.1.2 MC datasets

In this section first some general aspects of the generation of the MC datasets and their usage are discussed. After that, a brief description of all the individual datasets that were used is given.

General aspects

Generators, PDFs and particle masses. All the MC samples that were used in the scope of this thesis were produced by the ATLAS production group. For HERWIG [62] and AcerMC [15] samples the LO* PDF set MRST2007 [63] was used, the PDF set for MC@NLO [14] samples was CTEQ6.6 [64] and for ALPGEN [65] the CTEQ6L1 [66] PDF set was used. For all HERWIG samples the simulation of multi-parton interactions (MPI) was done with JIMMY.[67] While MC@NLO and ALPGEN were interfaced to HERWIG/JIMMY for parton showering/MPI, AcerMC was interfaced to PYTHIA.

For all samples, the pole mass and width of the W were set to $M_W = 80.4$ GeV and $\Gamma_W = 2.1$ GeV, respectively. For the pole mass and width of the top quark, values of $M_t = 172.5$ GeV and $\Gamma_t = 1.5$ GeV were used, respectively.

Event weights. In real data, each event obviously always has a weight of one, i.e. it is always counted as one event. In MC, it is sometimes useful to apply different weights to some or all events. A single event may obtain several event weights that serve different purposes. In this case all these weights are multiplied in order to obtain the overall event weight. In the following some situations are described, in which event weights are used.

In the simplest application of event weights, every event in a given dataset obtains the same weight. This corresponds to simply normalizing the whole sample as desired and this is just what it is used for.

Sometimes it is important that the distribution of a particular variable in data is welldescribed by MC, but in reality this is not the case. Then the actual distributions of that variable in data and MC can be used to extract event weights for MC, which depend only on the value of that variable in a MC event. These are calculated such that the overall normalization of the MC is conserved while the distribution of the variable in MC properly describes that in data, when the weights are applied. This technique is referred to as *reweighting* in the following.

Often the efficiency to reconstruct a certain object is not the same in MC as in data. This leads to a bad description of the multiplicity distribution of that variable, and, as soon as a selection cut is applied on this multiplicity, to a wrong prediction of the yield of events that pass the selection criterion. This can be corrected for by applying appropriate weights to the MC. This is done by first assigning a weight to each object of the relevant type that was reconstructed in the event, and then multiplying the weights of all these objects in order to obtain the actual event weight. As reconstruction efficiencies in general depend on the kinematics of the object, the individual object weights are normally functions of $|\eta|$ and p_{T} or other similar variables.

Normalization of the datasets. For all MC samples, reference production cross sections were provided. In most cases these are the cross section given by the generator, normalized to some higher-order theory calculations where applicable. For the three single top-quark production channels, the values from the theoretical calculations (see Section 3.2.2) are used. From this reference cross section

and the number of generated events, the equivalent integrated luminosity of each sample is calculated, i.e. the amount real data that is supposed to contain the same number of such events. The ratio of this equivalent luminosity and the actual luminosity of the set of real data which was used then provides the appropriate event weight to correctly normalize the MC to data.

Pileup reweighting. Pileup can have a considerable impact on physics analyses and therefore needs to be included in the simulation. This is done by adding different amounts of minimum bias events to the events of the actually simulated process. This needs to be done such that the multiplicity distribution of pileup interactions in MC matches that in data. But, as in general the MC is produced before or during the data taking, this multiplicity distribution can only be guessed in advance, which is generally not perfect. The resulting differences are therefore corrected using the reweighting procedure described above in the section on event weights. The distribution of the pileup interaction multiplicity in real data, which is needed for this purpose, is provided along with the GRL by the top reconstruction group.

Individual datasets

Signal. For the *Wt*-channel signal two different MC samples are available. One was simulated with MC@NLO using the diagram removal scheme (see Section 3.4.1), the other one was done with AcerMC. The MC@NLO sample contains about 900000 generated events in total; taking into account negative weights, the effective number of events is 800000. With the reference cross section of 15.6 pb, this corresponds to an equivalent luminosity of about 51 fb^{-1}. The AcerMC sample contains about 300000 events, corresponding to a luminosity of 19 fb^{-1}.

The MC@NLO sample was used as the baseline sample, as the NLO prediction is considered superior to the LO description provided by AcerMC. Furthermore it has almost three times the statistics. The AcerMC sample was rather used to cross-check the results obtained with the MC@NLO sample.

***W*+jets.** For the production of the *W*+jets sample ALPGEN was used. Only the three leptonic *W* decay modes were simulated. Production was split up into a couple of subsamples, depending on the number and type of partons generated in association with the *W* boson:

- production of zero to five light quarks or gluons ("*W* plus light flavor", "*W*LF+jets");
- production of a single *c* quark and zero to four light quarks or gluons ("*Wc*+jets");
- production of a $c\bar{c}$ pair and zero to three light quarks or gluons ("*Wcc*+jets");
- production of a $b\bar{b}$ pair and zero to three light quarks or gluons ("*Wbb*+jets").

The *W*+LF samples were furthermore split up according to the decay mode of the *W* boson, i.e. electron, muon or tau channel. The *Wc*, *Wcc* and *Wbb* samples together are later referred to as "*W* plus heavy flavor" or "*W*HF+jets". The equivalent luminosity of most of the samples amounts to about 7-8 fb^{-1}. The only major exceptions are the *W*+LF samples with zero or one parton, each of which corresponds to only about 0.4 fb^{-1}. The number of generated events ranges from about 30000 for the *Wc* four-parton sample and 70000 for the *W*+LF five-parton samples and the *Wbb* three-parton sample up to about $3.5 \cdot 10^6$ for the *W*LF zero- and two-parton samples and $6.5 \cdot 10^6$ for the *Wc* zero-parton sample.

The different *W*+jets samples are not strictly mutually disjoint regarding their flavor content; rather there is some overlap in phase space between the samples. For example, an event in the *W*LF+jet sample may still contain a $c\bar{c}$ pair from a gluon splitting, while this event may also be contained in (and belongs

to) the *Wcc*+jets sample. In order to avoid double counting, this overlap needs to be removed. For this purpose, a tool is available in ATLAS which flags those events that need to be removed. This procedure is denoted *heavy flavor overlap removal*, and it is applied before the actual event selection. In addition, the normalization and flavor composition of the full *W*+jets sample is not taken from theory predictions, but determined by data-driven techniques.

Z+jets. Like the W+jets, also the Z+jets channel was simulated with ALPGEN, considering only the three leptonic decay modes. Production was split up depending on the number of zero to five partons and the decay mode of the Z boson, extra samples for Z plus b- or c-quarks were not used. All Z+jets samples correspond to equivalent luminosities of about 8-10 fb^{-1}. The number of generated events ranges from about 10000 for the five-parton samples to about $6.5 \cdot 10^6$ for the no-parton samples.

Dibosons. Three diboson samples – WW, WZ and ZZ production – were all produced with HERWIG. In contrast to the single-boson samples, the decay modes of the W and Z were not restricted to leptonic decays. Instead, a generator-level cut was applied, requiring at least one electron or muon with $p_{\mathrm{T}} > 10$ GeV and $|\eta| < 2.8$. Each of the samples contains about 250000 events. This corresponds to luminosities of about 15, 45 and 200 fb^{-1} for the WW, WZ and ZZ sample, respectively.

Background from single top-quark production. The simulation of single top-quark t-channel and s-channel production was done with AcerMC. Only semileptonic top-quark decays were simulated, separate samples were produced for the electron, muon and tau channel. Each sample contains about 200000 events. In the t-channel sample, some events have negative weights,[1] such that the effective event number per sample goes down to about 185000. The corresponding luminosity of the samples is about 27 fb^{-1} for the t-channel and about 400 fb^{-1} for the s-channel.

Top quark pair production. The simulation of $t\bar{t}$ production was done with MC@NLO. Separate samples were produced for the all-hadronic decay mode and for all the rest ("non-all-hadronic"). For the all-hadronic sample about 10^6 events were generated, resulting in 930000 events when taking generator weights into account. The corresponding luminosity amounts to about 12 fb^{-1}. For the non-all-hadronic sample about $15 \cdot 10^6$ events were generated, resulting in $11.5 \cdot 10^6$ events when generator weights are taken into account, which corresponds to a luminosity of about 130 fb^{-1}. Due to the long processing time for so many events, only 10% of the non-all-hadronic sample were used in the scope of this thesis, however, which still provided by far sufficient statistics.

5.1.3 Multi-jets

As mentioned in Section 3.4.3, multi-jets background needs to be estimated by data-driven techniques. A number of different such methods are used in ATLAS. The common principle of these methods is to use a modified event selection to define a dataset which is assumed to contain events that are suited to model the multi-jets content by applying appropriate event weights to them. These weights can be either global weights that simply normalize the selected dataset or individual weights for each event which do the actual modeling of the multi-jets estimation (or a combination of both).

[1] The negative weights are related to the initial-state b quark of the t-channel, which implicitly arises from a gluon splitting inside the proton (see Section 3.2.2). Therefore they do not occur in the s-channel sample.

For this thesis the results from the so-called "jet-electron" method [68] were used for multi-jets estimation. In this method some criteria are defined to select jets that are considered likely to be misreconstructed as an electron. These are called jet-electrons. The dataset that is used to estimate the multi-jets contribution is obtained from the "JetTauEtmiss" stream using a jet trigger. Only events that contain no signal lepton, but exactly one jet-electron are used, where here for the electrons a slightly looser definition than normal is used (see Section 5.2.1). In these events the jet-electron is removed from the list of jets and instead inserted as the signal electron; apart from that, the normal event selection is applied. It turns out that this method can even be used in the muon channel, when the jet-electron is used as the signal muon instead. Appropriate normalization is obtained by fitting the jet-electron sample to data (more precisely: to the difference between data and the sum of the MC samples) in a sideband region. Usually the $p_{\mathrm{T}}^{\mathrm{miss}}$ range below the analysis threshold is used for this purpose.

5.2 Object definitions

The precise definitions of physics objects to be used for top-quark analyses, based on the different available reconstruction and calibration schemes (see Chapter 4), are provided by the top reconstruction group. These are updated regularly as needed. The object definitions that were used in the scope of this thesis are summarized in the following. They basically follow the recommendations from the top reconstruction group as of summer 2011, with a few modifications that are specific to the single-top group, which are indicated in the text.

5.2.1 Electrons

Two types of electron definition were used in the scope of this thesis: a "tight" one, which is the standard definition used for analysis and a "loose" one which is used for multi-jets estimation. These are not to be confused with the sets of electron identification cuts of the same names.

For both definitions, electrons that were reconstructed by the "standard electron" algorithm (see Section 4.3.1) are used. For each electron, an "object quality" flag is checked, which indicates whether any of the calorimeter cells from which the electron's cluster was built suffered from important hardware problems. In this case the electron is rejected. A slightly modified electron four-momentum definition with respect to the electron reconstruction is used: the electron energy as provided by the electron reconstruction is replaced by the cluster energy only, while the electron direction is kept as it is and the zero mass assumption is preserved as well.
For the tight definition, the electrons then need to satisfy the "tight" set of electron identification cuts (see Section 4.3.2) as well as an isolation criterion, consisting of two parts: the $E_{\mathrm{T,\,cone}}(30)$ variable of the electron must be larger than 0.15 times its transverse momentum and the $p_{\mathrm{T,\,cone}}(30)$ variable must be larger than 0.1 times the transverse momentum. This particular isolation requirement is specific to the single-top group. For the loose definition, electrons need to fulfill at least the "medium" set of cuts and they need to have a hit in the innermost Pixel layer on their track.
The same kinematic cuts are then applied for both definitions. The required minimum transverse momentum is 25 GeV, which is driven by the requirement that the trigger efficiency is on its plateau at that value. The pseudorapidity of the electrons is required to be within the acceptance range for high-precision electromagnetic measurements, which is $|\eta| < 2.47$, excluding the transition region of $1.37 < |\eta| < 1.52$.

Electron energy scale/resolution and efficiency corrections. The overall reconstruction efficiency for signal electrons is composed of the trigger efficiency, the efficiency of the electron reconstruction algorithm and the reconstruction and matching efficiency of the corresponding ID track. Any discrepancies between data and MC regarding these efficiencies are corrected using the method described in Section 5.1.2.

Furthermore, discrepancies in the electron energy scale and resolution are corrected for. The scale is corrected by applying a η-dependent factor to the measured electron energy in data, such that the Z mass peak is centered around its nominal value. The resolution is corrected by smearing the measured electron energy in MC such that the resolution is the same as in data. The modified electron energies are furthermore propagated to the reconstructed missing transverse momentum, i.e. the reconstructed $p_{\mathrm{T}}^{\mathrm{miss}}$ is adjusted electron-by-electron according to the changes.

5.2.2 Muons

The muon definition uses combined muons that were reconstructed by MuID. These muons automatically fulfill the "tight" quality flag, which technically is required explicitly in addition. The same isolation criterion that is used for the electron definition is also applied to the muons: $E_{\mathrm{T,\,cone}}(30)$ must be larger than 0.15 times and $p_{\mathrm{T,\,cone}}(30)$ must be larger than 0.1 times the muon's transverse momentum. Then some Inner Detector hit requirements are applied. Firstly, if the muon crossed an active part of the innermost Pixel layer, a measured hit in this layer is required. Secondly, the sum of the number of hits and the number of crossed dead sensors in the Pixel and SCT must be larger than one and five, respectively. Thirdly, the total number of holes on the track (i.e. crossed detector layers for which a hit is expected but none is measured) in the Pixel and SCT together must be less than three. Finally, there is a complicated requirement on the hits in the TRT. Be N the sum of the number of TRT hits and TRT outliers. Then if $|\eta| \geq 1.9$ and $N > 5$, the fraction of TRT outliers with respect to N must be less than 90%. If $|\eta| < 1.9$ then N *must* be > 5 *and* the fraction of TRT outliers must be less than 90%. The transverse momentum of the muons is required to be larger than 25 GeV, where the precise value of this threshold is again driven by the trigger, and less than 150 GeV.[2] Finally, the absolute value of the muon's pseudorapidity is required to be less than 2.5.

Overlap with jets. In addition to the criteria defined above, muons that have a distance ΔR of less than 0.4 to any selected jet are discarded. "Selected" is to be understood in the sense of Section 5.2.3, with the exception that the p_{T} threshold is lowered to 20 GeV here. The reason for this additional cut is that such muons are produced mostly by the decay of a hadron within that jet.

Muon momentum scale/resolution and efficiency corrections. The overall reconstruction efficiency for signal muons is composed of the trigger efficiency, the efficiency of the muon reconstruction in the muon spectrometer and the reconstruction and matching efficiency of the corresponding ID track. As for electrons, any discrepancies between data and MC regarding these efficiencies are corrected using the method described in Section 5.1.2.

Furthermore, discrepancies in the muon momentum scale and resolution are corrected for. The scale is corrected by applying an η-dependent factor to the measured muon transverse momentum in MC, such that the Z mass peak has the same mean value as in data. The resolution is corrected by smearing the measured muon transverse momentum in MC such that the resolution is the same as in data. As for the electrons, these corrections are also propagated to the reconstructed missing transverse momentum.

[2] This upper limit for the muon p_{T} is necessary due to some technical problems with the trigger simulation in MC.

5.2.3 Jets

The jets used for analysis are Anti-k_T-jets with $R = 0.4$ using topological clusters as input and calibrated in the EM+JES scheme as described in Section 4.5. A transverse momentum of at least 25 GeV is required, the pseudorapidity range of $|\eta| < 2.5$ is used.

Overlap with electrons. Jets and electrons may be reconstructed from the same calorimeter cells. Therefore, the jet that is closest (in ΔR) to a selected electron (in the sense of Section 5.2.1) is discarded, if the distance is less than 0.2. This is done after the muon-jet overlap removal.

5.2.4 Tagged jets

For b-tagging the JetFitterCombNN tagger (see Section 4.5.1) was used. Three different working points were considered in the scope of this thesis, employing cuts on the tagger output at 2.0, 0.35 and -1.25. These working points refer to nominal b-tagging efficiencies of 60%, 70% and 80% and light-quark rejection factors of 345, 96 and 21, respectively. In the following, the individual working points are usually referred to by their nominal efficiencies.

5.3 Event selections

5.3.1 General cleaning cuts

Some general, rather analysis-independent, cuts are applied in order to ensure good-quality collision data.

The first one is later referred to as "LAr cleaning". During about the last two thirds of the data-taking period that was used in the scope of this thesis, corresponding to 84% of the luminosity, there was a a problem with six dead front-end boards in the LAr calorimeter, resulting in an acceptance gap which is customarily referred to as "the LAr hole" in ATLAS. For electrons this is taken into account by the object quality flag. For jets, a somewhat more drastic procedure is applied. Any event that contains at least one jet which lies in the LAr hole, is rejected completely. Given the high fraction of affected data, this procedure was simply applied to the full dataset on both real data and MC.

Then there is the "vertex cleaning". The first primary vertex of the event is required to have at least four tracks in order to reject non-collision background, such as beam halo events or cosmic ray events.

Finally, events are removed that contain at least one jet which is classified as a so-called "bad" jet. The presence of such a jet suggests that either there was a problem in calorimeter in the event, such as a spike or coherent noise, or the event is not a collision event. This cut is later referred to as "jet cleaning".

5.3.2 Analysis selection

The analysis selection is applied on top of the general cleaning cuts. It comes in two variants, the *pretag* selection and the *tag* selection.

Pretag selection

Pretag selection basically denotes the full selection except for b-tagging. It mostly focuses on the presence of a leptonically decaying W boson, i.e. a hard and isolated lepton and a hard neutrino in the final state.
Therefore, in the electron channel the presence of exactly one electron and no muon is required, and in

the muon channel exactly one muon and no electron. Furthermore this electron/muon has to match the object that fired the trigger within a cone of $\Delta R = 0.15$. A missing transverse momentum of more than 25 GeV is required in order to account for the neutrino.
Furthermore the so-called *triangular cut* is applied: the sum of the missing transverse momentum and the transverse W mass, calculated from the missing transverse momentum and the signal lepton, has to be more than 60 GeV. This cut mostly serves to reject multi-jets background.

Tag selection

For the tag selection, exactly one b-tagged jet is required on top of the pretag selection, aiming at the b quark from the top-quark decay. The presence of a second (or more) b-tagged jet is vetoed in order to suppress $t\bar{t}$ background and reduce the interference between the Wt-channel and $t\bar{t}$ (see Section 3.4.1).

Obviously, the tag selection comes in different variants depending on the b-tagging working point that is used. Like the working points themselves, also the corresponding tag selections are usually referred to by the nominal b-tagging efficiency of the working point that is used (e.g. "the (ϵ =)70% selection").

b-tagging calibration. The distribution of most b-tagger weights in data is not perfectly described by MC, which leads to different b-tagging efficiencies and rejection factors in data and MC. This is corrected using the efficiency correction method described in Section 5.1.2, where the individual jet weights not only depend on the jet kinematics but also on the true jet flavor and on whether the jet was tagged or not. This is referred to as *b-tagging calibration*. The calibrations are performed by the ATLAS flavor-tagging working group. A tool is provided to apply these calibrations for analysis. Besides the actual calibrations, this tool also provides individual per-jet efficiencies and rejections including uncertainty estimations.

5.4 Results of event selection and analysis strategy

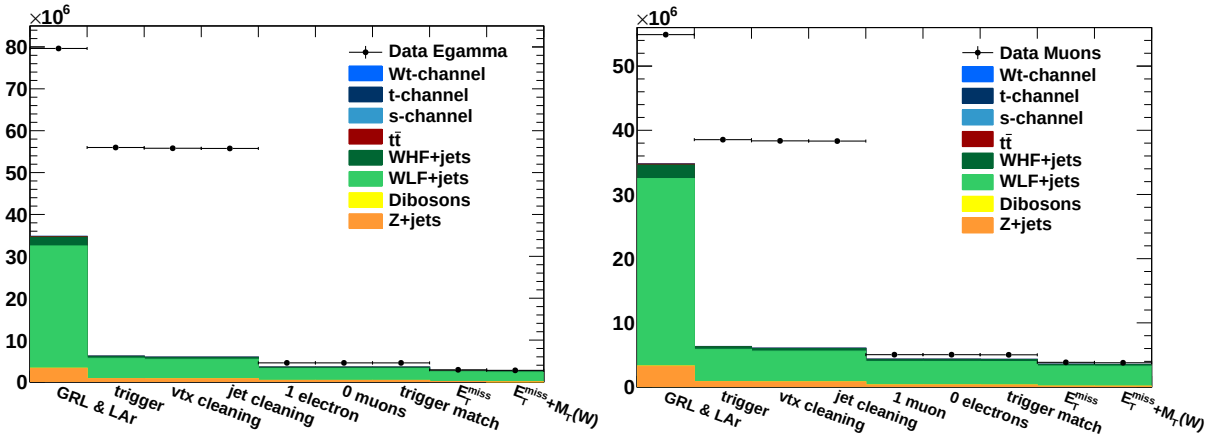
In this section the results that one obtains with the presented event selections are discussed, first for the pretag selection and then for the tag selection. Based on this, the general strategy of the analysis is motivated and the choice of a jet bin and a working point for analysis are discussed.

5.4.1 Pretag selection

Cut flow

The results of the general cleaning cuts and the pretag selection presented in the previous section are displayed in figure 5.1. It shows the number of events in data and MC after each consecutive cut in the electron channel (a) and in the muon channel (b). The histograms for MC are stacked, i.e. the full visible bin content corresponds to the sum of all MC samples, while the visible area of each color represents the respective contribution from the according MC sample. The W+jets background is split up into WLF+jets and WHF+jets. A multi-jets estimation is not included in the plot, it is rather assumed that any major discrepancy between data and the MC sum approximately represents the multi-jets contribution.

The first bin of each plot shows all the events that are in the GRL after heavy flavor overlap removal (W+jets MC only) and the LAr cleaning cut. A multi-jets fraction of $O(50\%)$ can be deduced, which is some orders of magnitude less than the relations among total production cross sections of the different processes (see figure 3.6) suggest. This is due to the fact that in order to be written to the Egamma or the Muons stream, which were used here, an event must pass at least one arbitrary electron/photon or muon



(a) (b)

Figure 5.1: Stack plot of the pretag selection cut flow on data and MC in the electron channel (a) and in the muon channel (b).

trigger, respectively. The observed moderate multi-jets fraction can therefore be seen as an indication that these triggers indeed filter out multi-jets events very efficiently.

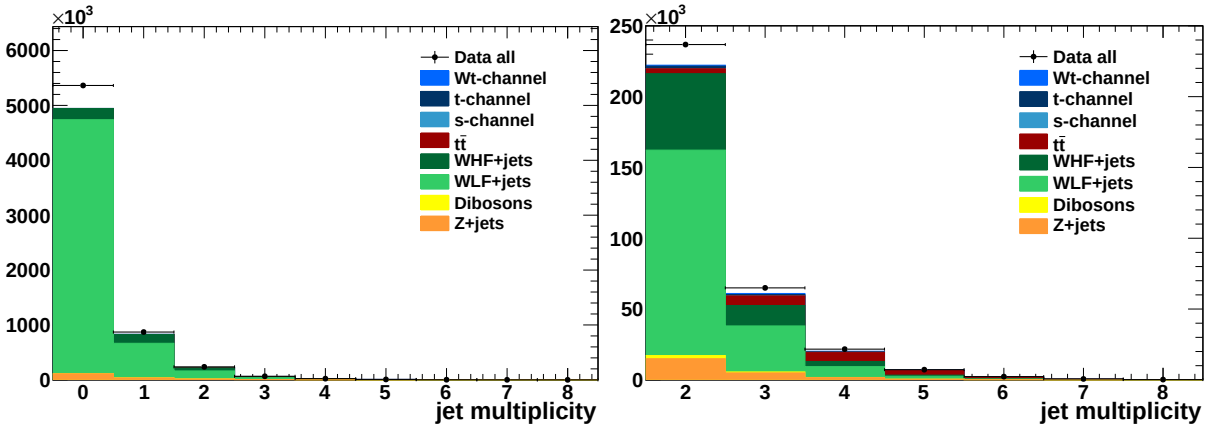
The next major reduction of multi-jets happens by requiring the signal lepton, while the dominating (within MC) W+jets contribution is reduced only rather little. This indicates that the used lepton definitions are indeed appropriate for efficiently selecting hard and isolated leptons such as typically arise from W decays, while the fake lepton background arising from multi-jets is mostly rejected. The last visible multi-jets rejection then happens due to the missing transverse momentum requirement. After that, the multi-jets contribution is hardly visible on the scale of the plots shown in figure 5.1.

Distribution of jet multiplicities

As soon as the signal lepton is required, the by far dominating contribution in figure 5.1 is given by W+jets background, while none of the top-quark production channels is even visible. However, as indicated in Section 3.4.3, W+jets production predominantly populates the lowest jet bins. A look at the jet multiplicity distribution is therefore advisable. Figure 5.2 shows this distribution; for the sake of simplicity only the sum of the electron and muon channel is shown. In figure 5.2(a) one can see that indeed about 95% of the W+jets events have zero or one jet. Figure 5.2(b) therefore shows a closeup of the jet multiplicity range with at least two jets. Both plots show clearly that in fact there still seems to be a small but non-negligible multi-jets contribution at least until the three-jet bin, which was hardly visible in the cut flow plot. Furthermore, in 5.2(b) the contribution from $t\bar{t}$ production is clearly visible, peaking at the three- and four-jet bins. One can even perceive some tiny contribution from single top-quark production, though on this scale it is virtually impossible to distinguish the individual production modes. In particular, the signal fraction is so tiny in any jet bin, that a Wt-channel analysis on the pretag selection seems quite hopeless. Therefore in the next section the tag selection is evaluated.

5.4.2 Tag selection

Figure 5.3 shows the jet multiplicity distribution in the tag selection for each of the considered working points. Naturally, this distribution only starts with the one-jet bin, as at least the b-tagged jet must be present. Although still very small, the signal contribution is now visible in the two-, three- and four-jet



(a) (b)

Figure 5.2: Jet multiplicity distribution in the pretag selection (a) and a closeup of only the range with at least two jets (b). Both plots show the sum of electron and muon channels.

bins. One has to note that the y axis scale strongly differs between the distributions for the individual working points, approximately by a factor of two from one working point to the next. Therefore, while it appears as if the number of signal events significantly decreases with increasing working point efficiency, in fact it increases slightly (see table 5.1).

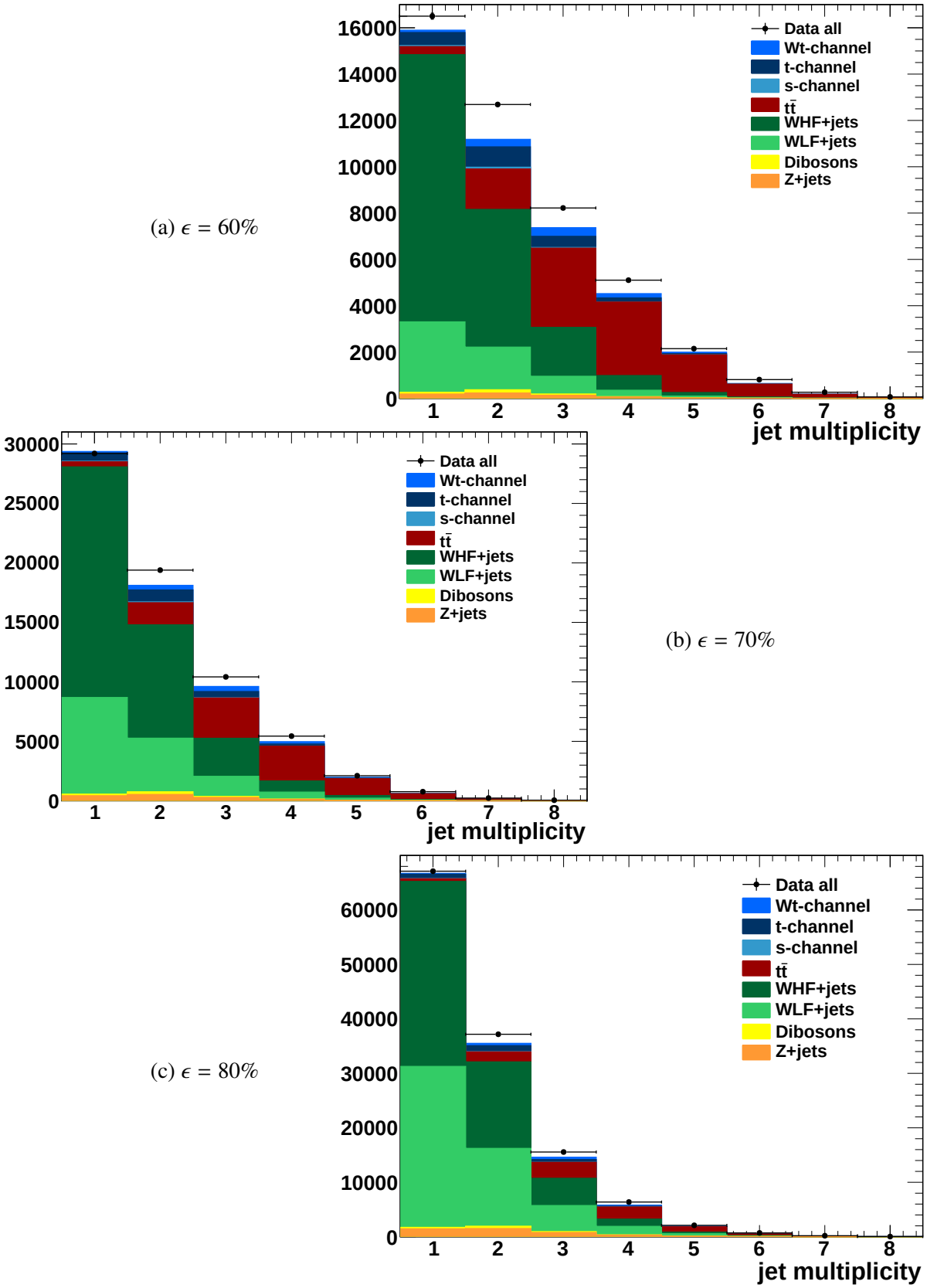
Background composition

Among the background channels, both W+jets and $t\bar{t}$ production are now the two dominant backgrounds. The relative contributions from each, as well as the light and the heavy flavor components of W+jets, vary strongly with the jet bin and the working point. W+jets production, as its production cross section falls exponentially with increasing number of jets, dominates at low jet multiplicities, while $t\bar{t}$ production, peaking in the three- and the four-jet bins, dominates at high multiplicities. The cross-over, where both have similar contributions, is in the three- or four-jet bin, depending on the working point. Furthermore, the amount of W+jets increases strongly with the working point efficiency, whereas at the same time $t\bar{t}$ decreases slightly. In fact, the different scales of the three histograms are mostly due to this increase of W+jets. Within W+jets production the heavy flavor component is dominant, while the light flavor fraction increases slightly with the number of jets and substantially with the working point efficiency. Among the minor backgrounds, t-channel production and Z+jets have the largest contribution, while dibosons only have a noteworthy contribution in the two-jet bin and s-channel production is almost completely negligible. Some explicit numbers on the background composition are shown exemplarily for the three-jet bin in table 5.2

5.4.3 Analysis strategy

Choice of analysis tools and general strategy

Given the very low signal fraction even in the tag selection, further background suppression is desirable. Looking at the distributions of typical variables for analysis, such as the four-momenta of the measured final-state particles and any combinations thereof, or standard topological variables, one finds that there are some variables that discriminate moderately between signal and W+jets, while there are fewer variables that discriminate against $t\bar{t}$, mostly with only very weak discrimination. Consequently, a simple



(a) $\epsilon = 60\%$

(b) $\epsilon = 70\%$

(c) $\epsilon = 80\%$

Figure 5.3: Jet multiplicity distribution in the tag selection.

	$\epsilon = 60\%$	$\epsilon = 70\%$	$\epsilon = 80\%$
	Events / S/B / $S/\sqrt{B}$	Events / S/B / $S/\sqrt{B}$	Events / S/B / $S/\sqrt{B}$
2 jets	324 / 3.0% / 3.1	385 / 2.2% / 2.9	431 / 1.2% / 2.3
3 jets	373 / 5.3% / 4.5	426 / 4.6% / 4.4	431 / 3.0% / 3.6
4 jets	185 / 4.3% / 2.8	197 / 4.1% / 2.9	184 / 3.3% / 2.4

Table 5.1: Number of signal events, signal-to-background ratio (S/B) and statistical significance (estimated by $S/\sqrt{B}$) in the tag selection, depending on the jet bin and the b-tagging working point.

cut-based analysis does not appear very promising.

The basic strategy that is followed in this thesis is to employ a kinematic fit to the signal topology in order to provide some additional variables that are directly sensitive to how much an event looks like signal. These variables, together with others, are then combined using a multivariate analysis tool, concentrating the full discrimination power in one single variable. Finally, the output variable from this multivariate tool is used to extract the signal.

For the kinematic fit, the KLFitter package was used. The implementation and validation of the kinematic fit are two of the main topics of this thesis and are detailed in Chapters 6 and 7, respectively. NeuroBayes was chosen as the multivariate analysis tool; it comprises an artificial neural network along with a powerful preprocessor that can be used to assess and choose appropriate input variables. A description of NeuroBayes is given in Section 8.1. Within this thesis it is used with only variables from the kinematic fit in order to demonstrate that these are indeed useful for analysis. This is done in Chapter 8. Also an expected statistical significance depending on a cut on the NeuroBayes output is extracted there. The full analysis using also non-kinematic-fit variables as well as the actual signal extraction are not part this thesis, however, as this was not done by the author and is furthermore still ongoing.

Choice of the jet bin

It still needs to be clarified which jet bin(s) and which b-tagging working point(s) are suited best for analysis. Table 5.1 shows the number of signal events, the signal-to-background ratio ("S/B") and the statistical signal significance, estimated by the signal-over-square-root-of-background ratio ("$S/\sqrt{B}$"), for the two-, three- and four-jet bins and for all three b-tagging working points, respectively. It appears quite clear that the three-jet bin is favorable, as for any working point is has not only the highest number of signal events but also both the highest signal-to-background ratio and estimated statistical significance among the jet bins. The only actual exception is the total number of events at the 80% working point, which is about the same in the two-jet bin as in the three-jet bin, but this does not affect the conclusion. For analysis, i.e. in Chapter 8 of this thesis, therefore the three-jet bin is used. For the validation of the kinematic fit in Chapter 7, additionally the four- and the five-jet bins are studied for comparison. The two-jet bin cannot be used there, as at least three jets are needed for the kinematic fit.

Choice of the working point

The choice of the optimal b-tagging working point is a bit more involved. The point here is that, as indicated above, it turns out that with further analysis W+jets is much easier to separate from signal than $t\bar{t}$. A working point that has a slightly lower overall statistical significance may therefore still be

	$\epsilon = 60\%$	$\epsilon = 70\%$	$\epsilon = 80\%$
	Evts. / $B_i/\Sigma B$ / $S/\sqrt{B_i}$	Evts. / $B_i/\Sigma B$ / $S/\sqrt{B_i}$	Evts. / $B_i/\Sigma B$ / $S/\sqrt{B_i}$
$t\bar{t}$	3420 / 48.9% / 6.4	3390 / 36.9% / 7.3	2930 / 20.7% / 8.0
W+jets all	2880 / 41.1% / 7.0	4880 / 53.2% / 6.1	9790 / 69.0% / 4.4
Other bkg.	699 / 10.0% / 14.1	912 / 9.9% / 14.1	1460 / 10.3% / 11.3
All bkg.	7000 / 100% / 4.5	9190 / 100% / 4.4	14200 / 100% / 3.6

Table 5.2: Number of events (B_i), fraction of all backgrounds ("$B_i/\Sigma B$") and individual statistical signal significance in the tag selection for selected background channels and for all three working points in the three-jet bin. Individual statistical signal significance denotes the signal significance with respect to each single background ("$S/\sqrt{B_i}$"). "Other bkg." represents the sum of s-channel and t-channel production, dibosons and Z+jets. All event numbers are rounded independently to three significant digits, therefore the sum of the single backgrounds doesn't precisely match the "All bkg." number.

preferable if it has a lower $t\bar{t}$ fraction.[3] For this reason it is instructive to have a closer look at the different compositions of backgrounds depending on the working point. Table 5.2 gives an overview of this composition from the main background channels in the three-jet bin. For $t\bar{t}$, W+jets and the sum of the minor backgrounds, it shows the number of selected events from the channel, the channel's fraction of all backgrounds ("$B_i/\Sigma B$") and the statistical significance of the signal with respect to each single background channel ($S/\sqrt{B_i}$) at each working point. In addition, the sum of all backgrounds is shown for comparison and for completeness.

Examining the working-point dependence of the main backgrounds, one finds that the $t\bar{t}$ contribution decreases moderately toward higher efficiencies. This is quite expected; assuming that $t\bar{t}$ events mostly contain two true b-quark jets, the probability that exactly one of them is tagged at an efficiency ϵ is $2 \cdot \epsilon \cdot (1 - \epsilon)$, which has its maximum value of 0.5 at $\epsilon = 50\%$ and then drops toward both higher and lower efficiencies. Moreover, the individual statistical significance against only $t\bar{t}$ increases toward higher working points. In case of a perfect discrimination against any other background, the highest-efficiency working point would therefore be preferable. In contrast, the W+jets contribution strongly increases toward higher efficiencies. This is mostly due to fact that higher efficiency also means lower light-quark rejection and therefore more and more events without a true b-quark jet survive the event selection due to a mistagged light-quark jet. Consequently, the individual statistical significance against W+jets drops toward higher b-tagging efficiencies.

Comparing the individual working points, one finds that the 60% and the 70% working points have quite similar overall significances, while $t\bar{t}$ and W+jets approximately change their roles: at 60% efficiency, $t\bar{t}$ is the largest background with a 49% background fraction, while at 70% W+jets becomes dominant with a background fraction of 53%. Given that W+jets can be separated somewhat better from signal by further analysis, the 70% working point should therefore be preferable. At the 80% b-tagging efficiency, the overall signal significance drops from 4.4 (at 70%) to 3.6, mostly driven by the W+jets contribution, which almost doubles compared to the 70% working point. It can be assumed that in order to profit from the individual statistical significance against $t\bar{t}$ being higher than at 70%, an extremely good discrimination against W+jets is needed, which is at least questionable whether it can be achieved. For this reason, the working point at 70% efficiency is considered likely to provide the best measurement in the end.

[3] In fact, this argument may also apply to the two-jet bin, which compares to the three-jet bin in precisely this way. However, as the kinematic fit needs at least three jets the two-jet bin cannot be used.

In this whole discussion, systematic uncertainties have not been taken into account. These may have a large effect on the final significance of the measurement, however, which may also depend on the working point. Hence, a definitive answer to the question which working point is best can only be given after the full analysis. The general analysis strategy therefore foresees to perform the full analysis for all three working points and only then decide on the best one. For the validation of the kinematic fit in Chapter 7 all three working points are compared in order to study the behavior of the fit. Chapter 8 instead exemplarily restricts the full study to the 70% working point.

Chapter 6

Implementation of a kinematic fit using the KLFitter package

The principle of kinematic fitting is to assume a given event topology to be the true topology of a measured event and then use known properties of this topology in order to provide improved estimates for some observables, which (better) match those properties than the actually measured values. In other words, the measurements are constrained to fulfill certain properties of the event that are assumed to be true. In the following, the assumed event topology is referred to as the (fit) *hypothesis* or the *model*, the properties used to constrain the measurements are simply referred to as the *constraints*. When performing kinematic fits, it is important to correctly take into account the measurement resolutions of the observables that enter the fit. Otherwise the effect of the constraints is too strong or too weak. The main aims of kinematic fitting are to provide a test for the fit hypothesis, to correctly assign the measured objects to the final-state particles that are part of the hypothesis (which may be regarded as a special case of the hypothesis test) as well as to provide improved estimates for those observables that are subject to the constraints in case the fit hypothesis is indeed true.

6.1 The KLFitter

The Kinematic Likelihood Fitter (KLFitter) [69] is a package for kinematic fitting using a likelihood approach. It is based on the Bayesian Analysis Toolkit (BAT).[70] The KLFitter was originally written for the $t\bar{t}$ lepton+jets channel, but it is designed such that it can be adopted to any other process.

6.1.1 Basic principle

The centerpiece of the KLFitter is the likelihood function. It defines the actual fit hypothesis. The likelihood function is supposed to return the likelihood to obtain a certain measurement, given the model, depending on some model parameters:

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

The fit is then performed in order to maximize the likelihood function, using (a subset of) the model parameters as the fit parameters, and yielding that set of fit parameters that has the highest likelihood of leading to the observed measurement. The final values of fit parameters then present the improved estimates in the above sense.

6.1.2 Typical application

In all current applications of the KLFitter the fit hypothesis is a specific decay chain of a particular production process. The model parameters are essentially the assumed true four-momenta of the model's

final-state particles. In fact, by fixing the masses of the final-state particles, the actual number of parameters per particle is usually reduced to three.

The final-state particles get associated to measured objects that are considered to correspond to them: final-state quarks get associated to measured jets, charged leptons get associated to the measured leptons, in case the final state contains a neutrino it gets associated to the measured missing transverse momentum. All these measured objects together then denote the "measurement" for which the likelihood is calculated.

By associating the measured particles to the model particles, also the model parameters canonically obtain their associated observables. The measured values of these observables (in the following simply denoted the "measured values of the parameters") serve as the initial values of the parameters for the fit. Some of the model parameters might be not used as fit parameters. These parameters keep their initial values throughout the fit. In case the final state contains a neutrino, only the x and y component of its momentum can be associated to the respective components of the measured missing transverse momentum, while the z component does not have a measured value. Its initial value must be calculated in a different way. Also it is clear that, without additional tricks, final states with more than one neutrino can not be treated like this.

In general, the association of the measured particles to the model particles is not unique. The KLFitter therefore is capable of calculating all possible meaningful associations of the measured particles to the model particles (referred to as "permutations" in the following). All permutations are fitted individually and then ranked (see Section 6.1.6). The best-ranked permutation then is normally considered as the true one.

The likelihood function for the described scenario is composed of two parts: the *transfer functions* (TFs) and the actual kinematic constraints:

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

The TFs are the part of the likelihood that takes the measurement resolutions into account. They represent the likelihood to measure a certain value of an observable, given the true value of the associated model parameter. The TFs are detailed in Section 6.1.3. The kinematic constraints are given by the invariant masses of the decay vertices within the decay chain. They are described in Section 6.1.4. Furthermore, b-tagging information can be used in order to refine the ranking of the permutations. The way this is realized within the KLFitter is explained in Section 6.1.5.

6.1.3 Transfer functions

As mentioned above, the TFs are defined as the likelihood to measure a certain value of an observable, given the true value of the associated model parameter:

$$\mathcal{L}_{\text{TF}} = W(x_{\text{meas}}|x_{\text{true}}) \,.$$

The KLFitter provides several such TFs for typical observables:

- for the energy of light-quark jets, b-quark jets, gluon jets, electrons and photons;
- for the transverse momentum of muons;
- for missing transverse momentum (x and y component);
- for the direction (i.e. for η and for ϕ) of light-quark jets and b-quark jets.

Observable	η-bins	Parametrization	Parameter dependencies
Light-quark jet energy	5	Double-Gaussian	Parameters are functions of E_{true}
Light-/b-quark jet η and ϕ	4	Gaussian	
b-quark jet energy	5	Double-Gaussian	Parameters are functions of E_{true}
Gluon jet energy	4	Double-Gaussian	Parameters are functions of E_{true}
Electron energy	4	Double-Gaussian	Parameters are functions of E_{true}
Photon energy	4	Gaussian	
Muon p_{T}	3	Gaussian	
Missing transverse momentum	1	Gaussian	Width is a function of $\sum p_{\text{T}}$

Table 6.1: Overview of the transfer functions that are part of the KLFitter with the number of η bins, parametrization and dependencies of the parameters.

Regarding the TFs for jet properties, one has to keep in mind that the model particles to which measured *jets* get associated, are final-state *quarks*. Hence, all those TFs represent the likelihood to measure a certain value of a *jet* property, given the true value of the corresponding *quark* property. The reason why there is a separate TF for b-quark jet energies is that b-quark jets frequently contain a neutrino and/or a muon from the semileptonic decay of a B or D hadron within the jet. These can carry away a significant amount of energy, which is then missing in the jet, such that the measured jet energy underestimates the true quark energy. As a result, the TF for b-quark jet energies has a significantly longer tail than the one for light-quark jet energies in order to accommodate this, which justifies the separate treatment.

All TFs except for the one for missing transverse momentum come in different bins of pseudorapidity. The TFs for the jet angles and for the photon energy are parametrized by simple Gaussians. The TF for missing transverse momentum is also parametrized by a Gaussian, but here the width of the Gaussian is parametrized as a function of the $\sum p_{\text{T}}$ variable, motivated by the observed behavior of the $p_{\text{T}}^{\text{miss}}$ resolution (see Section 4.6). All the other TFs (in particular most of the ones for particle energies) are parametrized by double-Gaussian functions, in order to be able to take asymmetric tails into account:

$$p(E_{\text{meas}}|E_{\text{true}}) = \frac{1}{2\pi(p_2 + p_3 p_5)} \cdot \left(e^{-\frac{(x-p_1)^2}{2p_2^2}} + p_3 \cdot e^{-\frac{(x-p_4)^2}{2p_5^2}} \right),$$

where $x = (E_{\text{true}} - E_{\text{meas}})/E_{\text{true}}$ and the parameters p_1-p_5 are themselves functions of the true energy. An overview of the available TFs with their basic properties is given in table 6.1. As an example, the TF for the b-quark energy is shown in figure 6.1.

6.1.4 Mass constraints

The mass constraints are realized by adding factors to the likelihood function that denote the probability for a particle to have a mass m, given its true pole mass M and width Γ. For the mass m one inserts the invariant mass of the two model particles that are to be constrained to their mother particle's mass, calculated from the values of their model parameters. The mass constraints are parametrized by

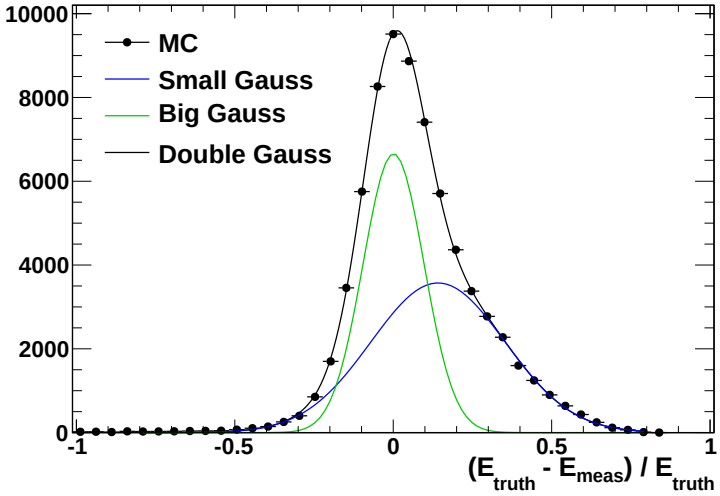


Figure 6.1: Transfer function for the b-quark energy for $0.0 < |\eta_{\text{meas}}| < 0.8$. The green and the blue line show the composition of the full double-Gaussian from two individual Gaussians. Courtesy of the KLFitter developers.

(relativistic) Breit-Wigner functions:

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

6.1.5 Using b-tagging information in the KLFitter

The KLFitter is capable of taking into account information from b-tagging. The idea is to compare the true flavors of the final-state quarks in the model to the measured flavors of the associated jets (where "flavor" in both cases denotes either "light" or "b"). As this does not depend on the model parameters, b-tagging information is not used in the actual fit. Instead, each permutation obtains a weight which can be interpreted as a probability for the b-tagging configuration of the permutation, i.e. for the pattern of how many (non-)b-tagged jets are associated to b- (light) quarks or vice versa. This weight is then multiplied with the likelihood from the fit, and the resulting quantity is called the *event probability*:

$$\mathcal{L}_{\text{event}} = w_{b\text{-tag}} \cdot \mathcal{L} \,.$$

The KLFitter provides two different modes for applying such weights to the permutations: the *working point mode* and the *veto mode*. Running without using b-tagging information is later on referred to as the *no-b-tagging mode*.

Working point mode

In the working point mode, the total b-tagging weight for a permutation is calculated as the product of the individual b-tagging probabilities for each parton in the final state. Individual b-tagging probability here denotes the probability to obtain the observed jet flavor, given the true parton flavor:

$$w_{b\text{-tag, WP}} = \prod_{\text{final-state partons}} p_{b\text{-tag}}(\text{ meas. jet flavor} \,|\, \text{true parton flavor}) \,.$$

These probabilities are given by the efficiency and rejection of the b-tagging working point that is used to determine the measured jet flavors. Explicitly, for an efficiency ϵ and a rejection R this reads:

$$\begin{aligned} p_{b\text{-tag}}(\text{ tag} \quad |\text{b} \quad) &= \epsilon\,, \\ p_{b\text{-tag}}(\text{ no tag}\,|\text{b} \quad) &= 1-\epsilon\,, \\ p_{b\text{-tag}}(\text{ tag} \quad |\text{light}\,) &= \frac{1}{R}\,, \\ p_{b\text{-tag}}(\text{ no tag}\,|\text{light}\,) &= 1-\frac{1}{R}\,. \end{aligned}$$

For the efficiency and rejection there are two options. The simpler one is to use the benchmark efficiency and rejection of the working point. The second method is to use the per-jet efficiency and rejection that one retrieves from the b-tagging calibration. This way, both quantities become functions of the jet p_{T} and η as they are in reality. The other advantage of this option is that the b-tagging calibration also provides uncertainties on the efficiency and rejection, which makes it easy to evaluate b-tagging related systematic effects on the analysis. However, this method was implemented too late in the KLFitter to be considered for this thesis.

Veto mode

In the veto mode, all permutations in which at least one b-tagged jet is associated to a light quark obtain a zero weight, while all other permutations obtain weight one. Up to a constant factor of 0.5 per final-state b quark, this is equivalent to using the working point mode with $\epsilon = 50\%$ and $R = \infty$:

$$\begin{aligned} p_{b\text{-tag}}(\text{ tag} \quad |\text{b} \quad) &= 0.5\,, \\ p_{b\text{-tag}}(\text{ no tag}\,|\text{b} \quad) &= 0.5\,, \\ p_{b\text{-tag}}(\text{ tag} \quad |\text{light}\,) &= 0\,, \\ p_{b\text{-tag}}(\text{ no tag}\,|\text{light}\,) &= 1\,. \end{aligned}$$

6.1.6 Ranking of the permutations

As said in Section 6.1.2, in the usual case that there is more than one possible association of the measured particles to the model particles, at first each permutation is fitted individually. Then the permutations are ranked according to the fit result. In the no-b-tagging mode, the permutations are simply ranked by their likelihood after the fit. In case b-tagging information is used, the permutations are ranked by their event probability.

6.2 Application of the KLFitter to the *Wt*-channel

One of the major pieces of the work in this thesis is the application of the KLFitter to the lepton+jets decay mode of the *Wt*-channel, according to the scheme described in Section 6.1.2. This section describes the way this is realized.

6.2.1 The likelihood function for the *Wt*-channel

Figure 6.2 shows the decay chain of the *Wt*-channel in the lepton+jets mode, with its final state consisting of one b quark, two light quarks, one charged lepton and one neutrino. The model parameters that are

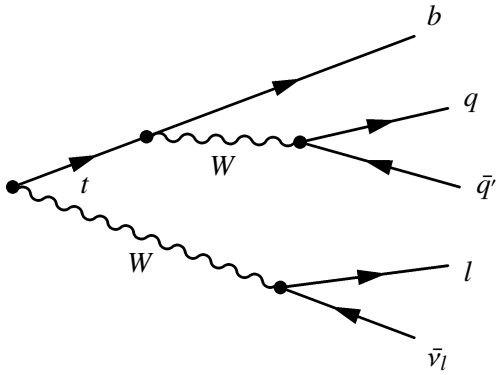


Figure 6.2: Decay chain of the *Wt*-channel in the lepton+jets mode, showing one b quark, two light quarks, one lepton and one neutrino as the final-state particles

used to describe this final state are the energy, E, pseudorapidity, η, and azimuthal angle, ϕ, of the three quarks and the lepton[1] as well as the three Cartesian components of the neutrino momentum $p_{\nu,x}$, $p_{\nu,y}$ and $p_{\nu,z}$, adding up to 15 parameters in total. The one parameter per particle that is missing to fully obtain all final-state particles' four-momenta is determined by fixing the mass of the b quark to the nominal value of 4.7 GeV and all other masses to zero. All final-state particles and their model parameters are associated to the corresponding measured objects and observables as described in Section 6.1.2. The z component of the neutrino momentum, $p_{\nu,z}$, is the only parameter that does not have a measured value. Its initial value is calculated from the measured lepton momentum and neutrino transverse momentum using the W mass constraint in the way that is described in Appendix A. All other parameters are initialized to their respective measured value.

The decay chain contains three decay vertices that can be used for mass constraints in the kinematic fit: one W that decays into the lepton and the neutrino; one W that decays into the two light quarks; and the top quark, which decays into the b quark and one of the W bosons. The following likelihood function was used for the fit of this decay chain:

$$\mathcal{L} = \mathcal{L}_{\mathrm{TF}} \cdot \mathcal{L}_{\mathrm{kin}} \,, \tag{6.1}$$

where

$$\mathcal{L}_{\mathrm{TF}} = \prod_{i \,\in\, \{E_b,\, E_q,\, E_{q'},\, E_\ell,\, p_{\nu,x},\, p_{\nu,y}\}} W(i_{\mathrm{meas}}|i) \tag{6.2}$$

and

$$\begin{aligned} \mathcal{L}_{\mathrm{kin}} = \; & BW(m_{qq'}|M_W,\Gamma_W) \cdot BW(m_{\ell\nu}|M_W,\Gamma_W) \\ & \cdot BW(m_{bW}|M_t,\Gamma_t) \,. \end{aligned} \tag{6.3}$$

The first term, $\mathcal{L}_{\mathrm{TF}}$, is the product of the TFs for the b-quark energy[2] E_b, the energies of the two light quarks E_q and $E_{q'}$, the lepton energy E_ℓ and the x and y component of the neutrino momentum $p_{\nu,x}$ and $p_{\nu,y}$. The z component of the neutrino momentum does not have a TF, as it does not have a measured

[1] For the muon, in fact p_{T}, η and ϕ are used rather than E, η and ϕ. Without loss of generality in the following only the lepton energy is referred to, for the sake of simplicity.

[2] The expression "TF for a (model) parameter" is always to be taken synonymous for "TF for the observable that is associated to the parameter".

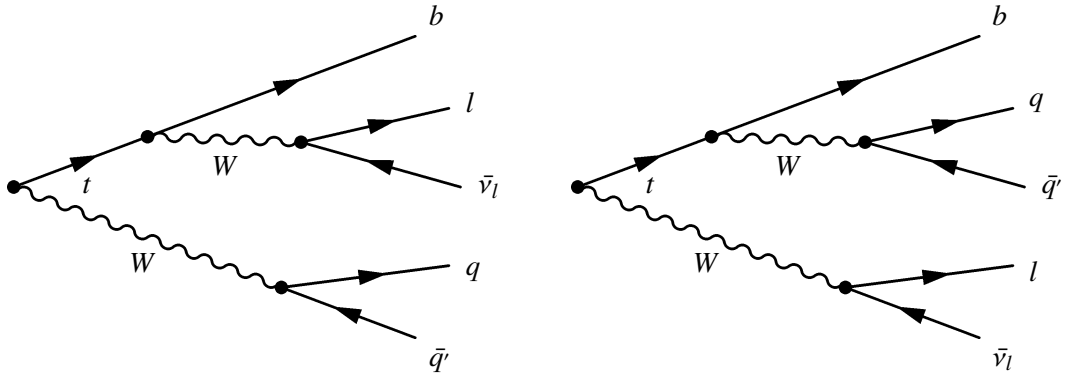


Figure 6.3: Decay chains of the *Wt*-channel in the lepton+jets mode for semileptonic (left) and hadronic (right) top-quark decay. It is impossible to transform one into the other by only interchanging the three quark labels, which would be equivalent to obtaining both fit hypotheses with one single likelihood function by only changing the association of the measured jets to the final-state quarks.

value. It only enters the fit via the mass constraints. Also, there are no TFs for the jet angles in the likelihood function. This is because, as recommended by the KLFitter developers, these are not used as fit parameters. As the jet angles are measured much more precisely than the jet energies, including them in the fit would improve the fit results only very little, while it would consume a lot more CPU time. The same holds for the angles of the charged lepton, for which the KLFitter doesn't even provide TFs.

The second term, $\mathcal{L}_{\text{kin}}$, consists of the three kinematic constraints, one for the top quark and one for each W boson. All are realized by Breit-Wigner functions (see Section 6.1.4). The values for the pole mass and the width of the W and the top quark that were used are $M_W = 80.4$ GeV, $\Gamma_W = 2.1$ GeV, $M_t = 172.5$ GeV and $\Gamma_t = 1.5$ GeV.

6.2.2 Permutations

The event selection requires exactly one measured charged lepton (see Section 5.3.2), which makes its association to the model lepton unambiguous. The association of the measured missing transverse momentum to the neutrino is unique anyway. Ambiguities in the association of the measured particles to the model particles therefore only arise from the association of the measured jets to the model quarks. For a given set of three jets, the association to the model quarks is uniquely determined by the choice of the jet that is associated to the b quark, as the two light quarks are considered indistinguishable. Hence, there are three different permutations in the three-jet bin. For higher jet bins, this has to be multiplied by the number of possibilities to draw three out of the N present jets. This amounts to 12 and 30 permutations in the four- and five-jet bins, respectively.

In the five-jet bin, only the four highest-p_{T} jets are passed to the fitter, reducing the actual number of fitted permutations back to 12. This follows a recommendation from the KLFitter developers, who use an equivalent procedure for the six-jet bin in top quark pair production. The argument is that the removal of a substantial part of the combinatorial background has a larger impact on the efficiency than the fact that sometimes also the correct permutation is removed. The corresponding study has not been repeated for the *Wt*-channel, however. In addition to the efficiency argument this procedure reduces the CPU consumption, as fewer fits need to be performed.

6.2.3 Semileptonic and hadronic top decay mode

The kinematic part of the likelihood, $\mathcal{L}_{\text{kin}}$, as given by equation 6.3 is not completely well-defined. It is ambiguous in the choice of the W boson that is used for the top-quark mass constraint: the leptonically decaying one or the hadronically decaying one (in the following simply referred to as the leptonic W and the hadronic W, respectively). Choosing one W for the top-quark mass constraint can also be perceived as choosing the decay mode of the top quark: semileptonic in case the leptonic W is used, hadronic in case the hadronic W is used. This ambiguity cannot be resolved just by the association of the measured jets to the model quarks. Rather the two top-quark decay modes pose two distinct fit hypotheses, which must be fitted separately using two distinct implementations of $\mathcal{L}_{\text{kin}}$:

$$\begin{aligned}\mathcal{L}_{\text{kin, lepTop}} = \; & BW(m_{qq'}(E_q, E_{q'})|M_W, \Gamma_W) \cdot BW(m_{\ell\nu}(E_\ell, p_{\nu,x}, p_{\nu,y}, p_{\nu,z})|M_W, \Gamma_W) \\ & \cdot BW(m_{bW_{\text{lep}}}(E_b, E_\ell, p_{\nu,x}, p_{\nu,y}, p_{\nu,z})|M_t, \Gamma_t) \end{aligned} \tag{6.4}$$

for the semileptonic top-quark decay mode hypothesis and

$$\begin{aligned}\mathcal{L}_{\text{kin, hadTop}} = \; & BW(m_{qq'}(E_q, E_{q'})|M_W, \Gamma_W) \cdot BW(m_{\ell\nu}(E_\ell, p_{\nu,x}, p_{\nu,y}, p_{\nu,z})|M_W, \Gamma_W) \\ & \cdot BW(m_{bW_{\text{had}}}(E_b, E_q, E_{q'})|M_t, \Gamma_t) \end{aligned} \tag{6.5}$$

for the hadronic top-quark decay mode hypothesis. In the following, the term "top decay mode" is usually used as a shortform for "top-quark decay mode", and the same term "(semileptonic or hadronic) top decay mode" is used for the two KLFitter modes that use the two different fit hypotheses. The decay chains for both cases are shown in figure 6.3.

Having realized that the Wt-channel is actually *two* hypotheses in terms of the KLFitter, the problem arises how to deal with this in data analysis. It seems desirable to make a decision concerning the top decay mode for each event. However, it will have to be evaluated whether this is actually possible. One aspect in this regard is that a likelihood, as provided by the KLFitter, does not present an absolute goodness-of-fit measure like a χ^2. It may be suited to test which one of two events better matches a given hypotheses, but it is quite unclear whether it is also directly applicable for comparing two distinct hypotheses, even if the two hypotheses look very similar at first glance. The answer to this question can therefore only be given after having a closer look at the fit results, which is done in Chapter 7.

6.2.4 Structure of the fit

In this section some aspects of the structure of the kinematic fit that arise from the likelihood function defined by equations 6.2, 6.4 and 6.5 are analyzed. This provides a useful basis for the study of the performance of the fit in the next chapter. Furthermore, another important issue about the fit emerges, which is discussed afterward in Section 6.2.5.

Constraint and parameter structure. The first thing to do in order to get a feeling for the fit and its general structure is to have a closer look at the parameters, in particular at which and how many constraints act on which parameters. Overall there are seven fit parameters: three quark energies, the charged lepton energy and the three components of the neutrino momentum. They can be grouped according to the mass constraints that act on them. The leptonic W mass constraint always acts on the charged-lepton energy and the three components of the neutrino momentum. This group is referred to as the leptonic parameters in the following. The next group is given by the two light-quark energies, which the hadronic W mass constraint always acts on. Finally there is the b-quark energy, which is the only

parameter which the top-quark mass constraint always acts on. In the semileptonic top decay mode the top-quark mass constraint also acts on the leptonic parameters, in the hadronic top decay mode it acts on the light-quark energies instead. In addition all parameters but the neutrino z momentum have a transfer function, which can be interpreted as additional constraints which always pull the parameters toward their initial values. Including the TFs in the counting one finds that there are two constraints acting on the b-quark energy regardless of the top decay mode, two and three constraints acting on the light-quark energies for the semileptonic and hadronic top decay mode, respectively, three and two constraints acting on the leptonic parameters except for the neutrino z momentum, two and one constraint acting on the neutrino z momentum. An overview of the constraint structure is given in table 6.2. It stands out that in the hadronic top decay mode only one constraint acts on the neutrino z momentum. This looks worrying, and indeed it will turn out later that there is a problem with the fit at this point.

Decomposition into two pieces. The next thing to notice is that not only the parameters group themselves according to the mass constraints, but in fact the whole likelihood function can be split up into two independent parts which only depend on entirely disjoint subsets of the fit parameters:

$$\begin{aligned}\mathcal{L}_{\text{lepTop}} = &\left(BW(m_{qq'}(E_q, E_{q'})|M_W, \Gamma_W) \cdot \prod_{i \in \{E_q,\, E_{q'}\}} W(i_{\text{meas}}|i)\right) \\ &\cdot \left(BW(m_{\ell\nu}(E_\ell, p_{\nu,x}, p_{\nu,y}, p_{\nu,z})|M_W, \Gamma_W) \cdot \prod_{i \in \{E_\ell,\, p_{\nu,x},\, p_{\nu,y}\}} W(i_{\text{meas}}|i)\right. \\ &\left.\cdot BW(m_{bW_{\text{lep}}}(E_b, E_\ell, p_{\nu,x}, p_{\nu,y}, p_{\nu,z})|M_t, \Gamma_t) \cdot W(E_{b,\text{meas}}|E_b)\right), \end{aligned} \tag{6.6}$$

$$\begin{aligned}\mathcal{L}_{\text{hadTop}} = &\left(BW(m_{\ell\nu}(E_\ell, p_{\nu,x}, p_{\nu,y}, p_{\nu,z})|M_W, \Gamma_W) \cdot \prod_{i \in \{E_\ell,\, p_{\nu,x},\, p_{\nu,y}\}} W(i_{\text{meas}}|i)\right) \\ &\cdot \left(BW(m_{qq'}(E_q, E_{q'})|M_W, \Gamma_W) \cdot \prod_{i \in \{E_q,\, E_{q'}\}} W(i_{\text{meas}}|i)\right. \\ &\left.\cdot BW(m_{bW_{\text{had}}}(E_b, E_q, E_{q'})|M_t, \Gamma_t) \cdot W(E_{b,\text{meas}}|E_b)\right). \end{aligned} \tag{6.7}$$

In either top decay mode the first part (given by the first line of the equation 6.6 and 6.7, respectively) consists of the prompt W mass constraint and the TFs of the corresponding parameters. The second part (second and third line of each equation) consists of the mass constraint of the W from the top-quark decay, the top-quark mass constraint and again the corresponding TFs. One may say that the fit decomposes into two uncorrelated branches, which correspond to independent fits of a top quark and a W boson.

Constraint counting. Another thing to do is to check how strongly the fit parameters are constrained overall by the fit, by just counting how many times the system is overdetermined. The overall number of mass constraints which act on the seven fit parameters is three, which naively implies a three-fold overdetermined system. However, the leptonic W mass constraint is used in order to calculate the initial value of the neutrino z momentum for the fit. In about 70% of the events this initial value is an exact solution of the W mass constraint (see Appendix A); additionally, for a large fraction of the remaining events the solution is close to exact. Therefore this is considered to be the normal case for this discussion. In this situation, the leptonic W mass constraint is already perfectly satisfied – more

	TF	$m_{W_{\text{lep}}}$	$m_{W_{\text{had}}}$	m_t lep/had	total lep/had
E_b	×	-	-	×/×	2 / 2
E_q	×	-	×	- /×	2 / 3
$E_{q'}$	×	-	×	- /×	2 / 3
E_ℓ	×	×	-	×/ -	3 / 2
$p_{\nu,x}$	×	×	-	×/ -	3 / 2
$p_{\nu,y}$	×	×	-	×/ -	3 / 2
$p_{\nu,z}$	-	×	-	×/ -	2 / 1

Table 6.2: Overview of the constraints and the fit parameters they act on. "Total" denotes the total number of constraints that act on a parameter (including TFs), "lep" and "had" denote the semileptonic and hadronic top decay mode, respectively.

precisely: the BW function is at its maximum – after the neutrino z momentum has been calculated, while all other parameters on which the constraint acts still have their initial values. That is, the leptonic W mass constraint then has lost all its actual constraining power, one may say that it is eaten up by determining the neutrino z momentum. Therefore the full system is effectively not three-fold but only two-fold overdetermined.

Given that the full fit decomposes into two inpendent branches, one should repeat the counting exercise for each branch independently. In both top decay modes, the top-quark branch contains two of the mass constraints and the prompt W branch contains the remaining one. In the semileptonic top decay mode, the leptonic W mass constraint belongs to the top-quark branch, such that effectively each branch is one-fold overdetermined. In the hadronic top decay mode, in contrast, the top-quark branch is two-fold overdetermined, while the prompt W branch only contains the leptonic W mass constraint is therefore effectively not overdetermined at all! This indeed poses a problem, which is discussed now in Section 6.2.5.

6.2.5 Treatment of the leptonic W in the hadronic top decay mode

As just seen, the prompt W branch in the hadronic top decay mode is generally not overdetermined at all, as the only mass constraint is eaten up by the neutrino z momentum. The following solution was used for the implementation of the kinematic fit in the scope of this thesis. Regardless of the number of exact solutions, $p_{\nu,z}$ was always fixed to its initial value during the fit. In case the initial $p_{\nu,z}$ value was an exact solution, i.e. the W mass constraint is perfectly satisfied, there is nothing to fit any more. Consequently the remaining four parameters are also fixed to their initial values and the fit is only performed on the top-quark branch in this case. In case there was no exact solution (see Appendix A), the W mass constraint is not yet satisfied completely after $p_{\nu,z}$ was calculated. The system is therefore still one-fold overdetermined and the fit of the prompt W branch can be performed as usual with the four remaining parameters.

Chapter 7

KLFitter performance and validation

In the previous chapter the implementation of a kinematic fit to the *Wt*-channel topology within the framework of the KLFitter package was described. The aim of this chapter is to validate that the implementation was done correctly and to show that the fit works as intended, i.e. that the behavior of its performance depending on various parameters is understood and that it actually fulfills the three main aims of kinematic fitting that were formulated in the beginning of Chapter 6: to correctly assign the measured objects to the final-state model particles, to provide improved estimates for the observables that correspond to the fit parameters with respect to their measured values; and to provide a test for the fit hypothesis.

For this purpose, in the first section of this chapter some suitable subsets of the signal MC sample are defined, which are used in the two subsequent sections when the performance of the KLFitter is analyzed. In the first of these two sections, the efficiency of the KLFitter to pick the correct permutation as the best-ranked permutation, referred to as the *fit efficiency* in the following, is studied. After that, the improvement of the fit parameters with respect to their measured values is evaluated. The qualitative behavior of both is studied depending on the decay mode of the top quark (semileptonic or hadronic), the jet bin (three-, four- or five-jet bin) and the selection (tag selection at different working points or pretag selection). For the fit efficiency, the impact of the *b*-tagging mode (working-point mode at different working points or no-*b*-tagging mode) is studied. Before looking at the observed results, some expectations are formulated that arise from the structure of the fit. These expectations are then compared to the actually observed behavior. Finally, in the last section of the chapter, distributions of the basic output variables of the KLFitter are shown and it is discussed to what extent they are suitable for use in a hypothesis test.

All results that are shown in this chapter were obtained with the MC@NLO MC sample. The same studies have also been performed with the AcerMC sample, and all relevant conclusions that are presented in this section apply there as well. The results from this cross-check are therefore not stated explicitly in general, but only in a few cases for which it appeared appropriate.

7.1 Structure of the signal sample

In order to determine the efficiency of the fit to select the correct jet permutation, it is necessary to know first whether a given event actually fulfills the fit hypothesis. Furthermore, an improvement of the fit parameters can only be expected for the correct permutation. In other words, performance studies only make sense when an appropriate subset of the full signal sample is used. Therefore it is advisable to have a closer look at the structure of the signal sample after the full selection before proceeding to the actual performance studies. This is done in this section.

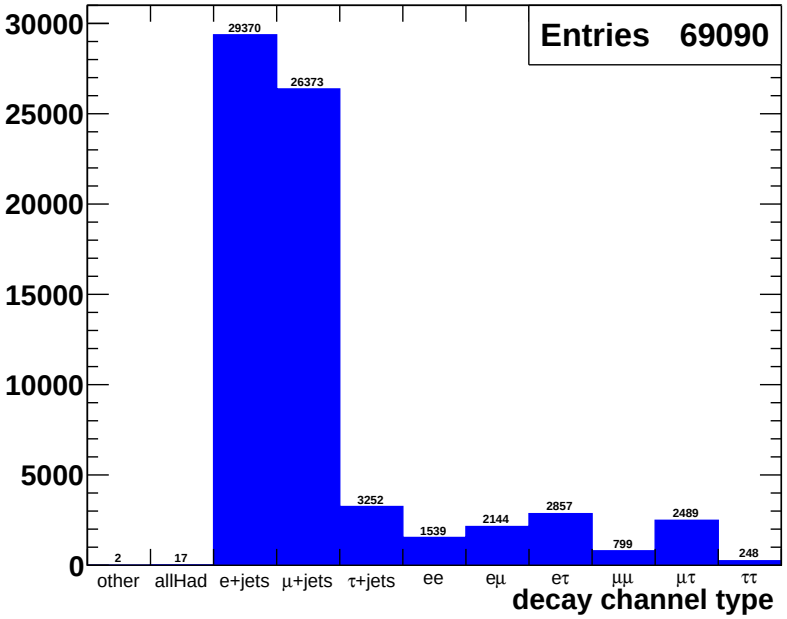


Figure 7.1: Composition of true decay modes in the signal sample in the three- to five-jet bins in the pretag selection. "Other" means that an event could not be classified by the event classification.

7.1.1 Composition of decay modes

First of all one needs to have a look at which kind of events actually pass the event selection. Although the selection aims at selecting the lepton+jets decay mode of the *Wt*-channel, it is unavoidable that also some events from other decay modes get selected. The distribution of the true decay modes of the events in the three- to five-jet bins in the pretag selection is shown in figure 7.1. It can be seen that most of the events are indeed electron+jets or muon+jets decays. The tau+jets decays provide the next largest contribution. Given the very low number of all-hadronic decays surviving the selection and the fact that a tau+jets event with a hadronic tau decay would look very similar to an all-hadronic decay, it can be assumed that most of the selected tau+jets events are leptonic tau decays. Therefore, the selected tau+jets events are considered to fulfill the signal definition given in Section 3.4.2, which at the same time means that they fulfill the fit hypothesis for the KLFitter. Still, there is also a non-negligible number of all kinds of dilepton events (plus a few all-hadronic or unclassified decays), that add up to a fraction of about 15%. This means that 15% of all selected *Wt* events do not contain a correct permutation that could be found by the fitter, i.e. they do not fulfill the fit hypothesis. Therefore, from a KLFitter point of view they might as well be regarded as background.

The assumption that most of the tau+jets events do fulfill the fit hypothesis was not checked in practice. Rather for the rest of this chapter (and also in Section 8.3), selected "true lepton+jets decays" always denote direct electron+jets or muon+jets events only.

7.1.2 Matched events

In order to fulfill the fit hypothesis not only theoretically but also in practice, an event not only needs to be a true lepton+jets decay, but also all final-state particles must have been reconstructed and the reconstructed particles must fulfill the respective object definition. The next step therefore is to have a closer look at the selected true lepton+jets decays.

Matching of measured particles to truth particles

One has to make sure that for each final-state particle a measured (i.e. reconstructed) object is present. For the neutrino this is trivial, as the missing transverse momentum is measured for every event and a cut on the minimum $p_{\mathrm{T}}^{\mathrm{miss}}$ is part of the event selection. For the other particles, this was done using a simple ΔR-based truth matching with a cut at $\Delta R = 0.4$ as described in Appendix B.1.

Matching of the charged lepton. For the lepton the situation is straightforward. Due to the event selection, there is exactly one measured lepton in each selected event. The distribution of ΔR between the measured lepton and the true final-state lepton for all true lepton+jets events in the three- to five-jet bins in the pretag selection is shown in figure 7.2. One can see that the matching efficiency for the lepton, defined as the number of events in which the measured lepton matches the true one divided by the total number of events, is virtually 100%. In the following, this efficiency is simply assumed to be indeed exactly 100% and the successful lepton matching is added to the "true lepton+jets" requirement instead, for the sake of simplicity.

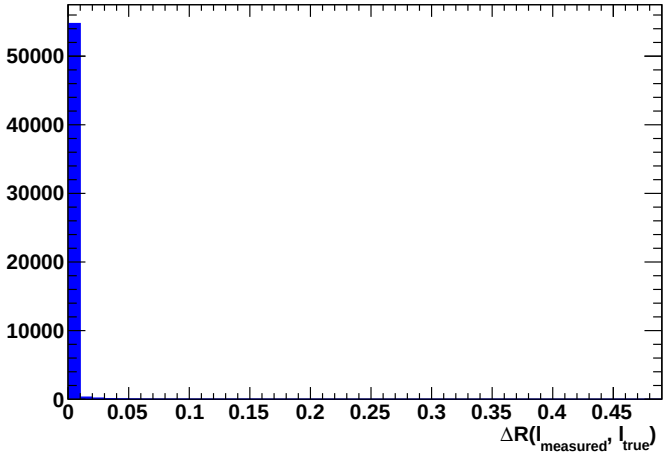


Figure 7.2: Distribution of the ΔR distance between the measured lepton and the true final-state lepton.

Quark-jet matching. Now the same matching has to be done for each of the measured jets. This procedure is referred to as *quark-jet matching* in the following. In contrast to the matching of the lepton, for jets it can happen that more than one measured jet matches a true quark or vice versa. The full description of the procedure that was applied in case of such ambiguities as well as the ΔR-distributions for the quarks can be found in Appendix B.1.
For the evaluation of the quark-jet matching, no distinction was drawn between the two light quarks, giving rise to six different possible combinations of quarks that could be matched to measured jets in a given event. The distribution of the pretag-selected true lepton+jets events in the three- to five-jet bins corresponding to these six categories is shown in figure 7.3.

Matching efficiencies

One can now define the *matching efficiency* as the fraction of events in which all three quarks can be matched to a measured jet. Such events that have all three quarks matched to a jet are referred to as *matched events* in the following. An overview of the matching efficiency, depending on the selection, the top-quark decay mode and the jet bin, is given in table 7.1.

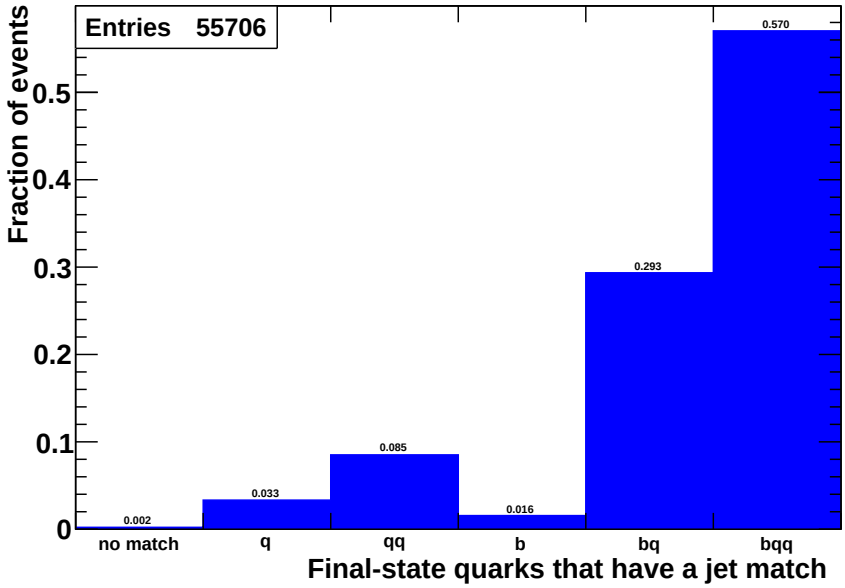


Figure 7.3: Distribution of the pretag-selected true lepton+jets events in the three- to five-jet bins corresponding to the six possible quark-matching categories. The overall fraction of events for which all three quarks can uniquely be matched to jets is about 57%.

Matching efficiencies in the pretag selection. For the pretag selection, the overall matching efficiency in the three- to five-jet bins is about 57%. It is lowest in the three-jet bin, where it is only about 51% for the pretag selection, and increases with the number of jets up to about 68% in the five-jet bin. The reason is quite apparent: the more jets fulfill the object definition, the higher is the probability that the three jets that correspond to the true quarks are among them. For all jet bins in the pretag selection the matching efficiency is about two to three percentage points higher for semileptonic top-quark decays than for hadronic top-quark decays. Presumably, this is related to differences in the kinematics of the hadronic W boson, which is the prompt W in the former and the one from the top-quark decay in the latter case. However, this has not been investigated any further.

Matching efficiency	All-jet bin lep / had	Three-jet bin lep / had	Four-jet bin lep / had	Five-jet bin lep / had
Pretag selection	0.584 / 0.557	0.522 / 0.502	0.641 / 0.608	0.696 / 0.666
Tag selection (ϵ = 60%)	0.627 / 0.596	0.571 / 0.554	0.686 / 0.637	0.73 / 0.67
Tag selection (ϵ = 70%)	0.623 / 0.595	0.573 / 0.559	0.675 / 0.638	0.73 / 0.66
Tag selection (ϵ = 80%)	0.611 / 0.593	0.568 / 0.559	0.662 / 0.639	0.71 / 0.66

Table 7.1: Overview of the matching efficiency for semileptonic/hadronic top-quark decays, depending on the jet bin and the selection. Typical uncertainties on the efficiencies range from 0.3 percentage points in the three-jet bin in the pretag selection to one percentage point in the five-jet bin in the tag selection. For the latter, only 2 digits of the corresponding values are displayed.

Matching efficiencies in the tag selection. When moving from the pretag to the tag selection, the obvious expectation is that the matching efficiency should increase, as the tag requirement strongly enriches such events, in which at least the jet that originates from the b quark is present (i.e. fulfills the object definition). Furthermore it removes events with two b-tagged jets, which should further increase the matching efficiency at least in the three-jet bin. It is possible to estimate an upper limit for the expected increase in matching efficiency due to the tag selection from figure 7.3. Assuming a perfect b-tagger and perfect matching and neglecting events with more than one b-tagged jet, the b-tagging requirement would remove exactly all the events in which the b quark has no match. Removing these bins from the histogram, which together amount to an event fraction of about 12%, and scaling it up such that that it has unit integral again yields a matching efficiency of 64.6% for this idealized scenario. What one observes from table 7.1 is that the actual increase is only about half the size. The overall matching efficiency (i.e. for the all-jet bin) rises by three to four percentage points to about 61%, depending on the working point. The differences between the individual working points are generally small, the difference between the highest and lowest efficiency per jet bin and top-quark decay mode is mostly of the order of one percentage point. The relation between the individual jet bins is mostly the same as for the pretag selection. The same holds for the relation between the top-quark decay modes. The only clearly diverging behavior can be seen in the five-jet bin, where for the hadronic top-quark decay there is basically no increase compared to the pretag selection, which at the same time leads to an increased difference between the two top-quark decay modes of five to six percentage points.

Summary of matched events

In summary it can be stated that about 43% of all true lepton+jets events actually do not fulfill the fit hypothesis, depending on the top-quark decay mode and the event selection that is considered. This needs to be taken into account when evaluating the performance of the KLFitter. For this reason, only matched events were used for the study of both the efficiency to find the correct permutation and the resolution improvement of the fit parameters.

7.1.3 Fit-truth matching and good-fit events

For the evaluation of the fit efficiency, the model quarks of the best-fit permutation need to be matched to the true quarks. This is referred to as *fit-truth matching* in the following. For the fit-truth matching, again a simple ΔR matching approach, with the same cut value of $\Delta R = 0.4$, was used.[1] The exact procedure, mainly to disentangle the ambiguity in the assignment of the light quarks, is explained in Appendix B.2.

The result of this matching can be evaluated using the same categories as for the quark-jet matching. Figure 7.4 shows the distribution of all pretag-selected matched events in the three- to five-jet bins corresponding to these categories. The plot shows a matching efficiency (i.e. a fit efficiency) of 61.5%. Such events, for which all three best-fit quarks can be matched to the true quarks, are referred to as *good-fit events* in the following. The fit efficiency can then be defined as the fraction of good-fit events in a given selection of matched events.

[1] One may argue that it would have been sufficient to simply reuse the results from the quark-jet matching in order to identify whether the KLFitter found the corrrect permutation. However, at the time this was implemented, the jet angles were still among the fit parameters, therefore a separate matching was indeed appropriate. When the jet angles were then removed from the fit, things were just left as they were, as the extra effort of simplifying the procedure would have been without any benefit.

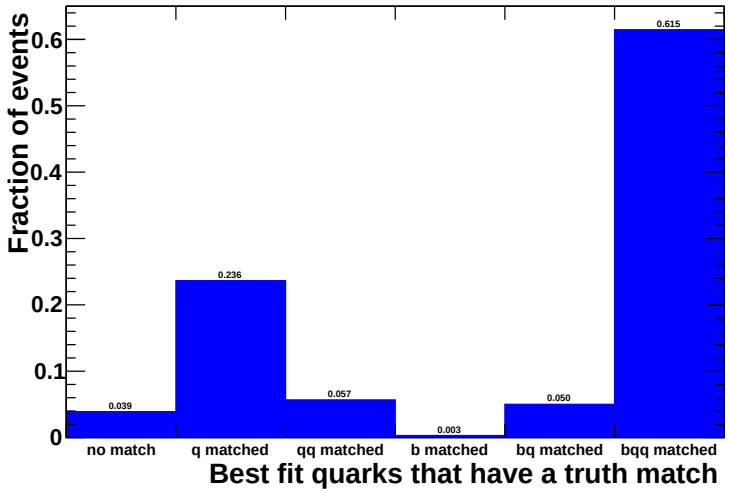


Figure 7.4: Distribution of the fit-truth matching results corresponding to the six matching categories for pretag-selected matched events in the three- to five-jet bins.

7.2 Fit efficiencies

This section is devoted to the evaluation of the fit efficiency depending on the jet bin, the top-quark decay mode, the selection and the b-tagging mode. For this purpose only matched events were used and all events were only fitted using the correct top decay mode.

In the following, first some general considerations about different aspects of the structure of the fit that have an impact on the behavior of the fit efficiency are formulated. Then the fit efficiencies are analyzed, first for the no-b-tagging mode, then for the working-point mode. In both cases, based on the said general considerations, some expectations are formulated prior to looking at the actual numbers.

7.2.1 General considerations concerning the fit structure

In this section the structure of the fit is analyzed in detail, such that on this basis expectations for the fit efficiencies can be formulated later on. In particular, any sources of potential differences in fit efficiency between the two top decay modes as well that the impact of the b-tagging working-point mode on the fit efficiency are discussed.

Sensitivity to wrong permutations depending on the top decay mode

The two parts of the likelihood function that directly depend on the association of the measured jets to the quarks and therefore provide sensitivity to wrong permutations are the hadronic W mass constraint and the top-quark mass constraint. The W mass constraint acts in the same way for both top decay modes and therefore does not play a role in the discussion of the differences between them. The difference rather lies in the way the top-quark mass constraint works.

In the hadronic top decay mode, the top-quark mass constraint acts on all three jets, but it is basically blind to which of the jets is associated to the b quark. The only small differences arise from the fact that the b quark is taken to be massive, while the light quarks are considered massless, and that the TF for the b-quark energy is different from the one for the light-quark energies. Both are expected to be of secondary importance, however. Therefore one can say that in the hadronic top decay mode, the top-quark mass constraint alone only helps to pick the correct set of three quarks, but not to pick the

correct b-quark jet. In the semileptonic top decay mode, the situation is exactly complementary. Here, the top-quark mass constraint only acts on the b quark, but does not depend at all on the choice of the two light-quark jets. Therefore one can say that in the semileptonic top decay mode, the top-quark mass constraint alone only helps to pick the correct b-quark jet, but not to pick the correct set of three quarks.

Together with the hadronic W mass constraint, both modes provide sensitivity against any wrong permutation. Some wrong permutations do fulfill one of the two relevant mass constraints in terms of correct quark-jet association. Obviously, the overall sensitivity against such permutations is reduced, as it is provided by one mass constraint alone. The permutations that fulfill the hadronic W mass constraint, and therefore rely only on the top-quark mass constraint, are the same for both top decay modes. Any differences in fit efficiency that can be traced back to such permutations can be interpreted as an indication for a different discrimination power of the semileptonic and the hadronic top-quark mass constraint.[2] In contrast, the type and number of permutations that rely only on the W mass constraint differs between both top decay modes. For the hadronic top decay mode the number of such permutations is always two, for the semileptonic mode it depends on the jet bin. Assuming approximately equal discrimination power of both top-quark mass constraints, this should be the major source of efficiency difference between the two top decay modes.

Considerations related to b-tagging

General assumptions. For all considerations about b-tagging in this chapter, some simplifying assumptions are made. Firstly, infinite light-quark rejection is assumed, i.e. only true b quarks can get tagged. Secondly the quark-jet matching of the b quark is considered perfect, i.e. the jet that is matched to the true b quark is always indeed the true b-quark jet.[3] Thirdly, it is assumed that an event always contains zero or one, but never more true b quarks. Basically, these assumptions mean that any b-tagged jet must automatically be the true b-quark jet from the top-quark decay. Events that violate these assumptions are assumed to play only a subordinate role.

Extra-suppression of wrong permutations in the working-point mode. As explained in Section 6.1.5, the b-tagging working-point mode assigns a b-tagging weight to each permutation, which is then used in the ranking of the permutations. This weight is composed of one b-tagging weight for each model quark. The correct permutation associates the b-tagged jet (if present) to the true b quark and the two untagged jets to the two true light quarks. According to Section 6.1.5, the b-tagging weight of this permutation is then given by $\epsilon \cdot (1 - 1/R)^2$ for a working point with efficiency ϵ and rejection R. For any wrong permutation, one can now define a suppression factor as the ratio of its b-tagging weight to that of the correct permutation. In the following, a classification of wrong permutations is developed, such that this suppression factor depends only on the type of wrong permutation.

Classification of wrong permutations

It is useful and instructive to classify the wrong permutations into three types, depending on what type of quark the true b-quark jet is associated to. This is done in the following paragraphs. For each type of permutation first the native (i.e. in the no-b-tagging mode) sensitivity of the KLFitter depending on the top decay mode and then the extra-suppression that is obtained when running in the b-tagging working-point mode are discussed. The discussion of the latter assumes that there is exactly one b-tagged jet in

[2] This holds in the same way for permutations against which both constraints are sensitive.
[3] Here and in the following "true b-quark jet" is used to denote "the jet that originates from the true b quark".

Type	Associated flavor	Multiplicity in the 3-/4-/5-jet bin	$N^{\text{lepTop}}_{\text{insensitive}}$	$N^{\text{hadTop}}_{\text{insensitive}}$	b-tagging suppression
First type	b	- / 2 / 5	all	-	-
Second type	none	- / 3 / 12	-	-	moderate
Third type	light	2 / 6 / 12	-	2	strong

Table 7.2: Overview of the different types of permutations and their properties. "Associated flavor" denotes the flavor of the model quark that is associated to the true b-quark jet, $N^{\text{lepTop}}_{\text{insensitive}}$ and $N^{\text{hadTop}}_{\text{insensitive}}$ denote the number of permutations of the given type to which the semileptonic and hadronic top-quark mass constraint are insensitive, respectively, "b-tagging suppression" denotes the additional suppression in the working-point mode.

the event, which approximately corresponds to the tag selection. An overview of the different types of permutations and their discussed properties is presented in table 7.2.

First type of permutation. The first type of wrong permutations denotes those in which the true b-quark jet is associated to the b quark, while a wrong set of light-quark jets is associated to the light quarks. Such permutations obviously do not occur in the three-jet bin. In the four- and five-jet bins there are two and five of them, respectively (the missing third and sixth one, respectively, is the correct permutation). For the semileptonic top decay mode, the top-quark mass constraint provides no sensitivity to such permutations, as they all assign the correct jet to the b quark. For the hadronic top decay mode, in contrast, the top-quark mass constraint does provide sensitivity, as such permutations never use the correct set of three jets. The native discrimination power of the KLFitter against such permutations is therefore higher in the hadronic top decay mode.

The working-point mode applies the same b-tagging weight to these permutations as to the correct permutation, as the b-tagged jet is correctly assigned to the b quark. Therefore the working-point mode does not provide any extra sensitivity to these permutations at all. This means in particular that the semileptonic top decay mode still relies only on the hadronic W mass constraint in this case.

Second type of permutation. The second type of wrong permutation denotes those in which the b-quark jet is not used at all. Like the first type, also this type of wrong permutations does not occur in the three-jet bin. In the four- and five-jet bins there are three and twelve of them, respectively. In both top decay modes, the top-quark mass constraint is sensitive to such permutations, as neither the correct jet is used for the b quark nor the correct set of three jets is used. Any difference in efficiency to reject these permutations between the two modes can therefore be regarded as an indication for a difference in discrimination power of the top-quark mass constraints. For the time being it is assumed that the hadronic and the semileptonic top decay mode are equally sensitive to such permutations, i.e. the two top-quark mass constraints are assumed to be approximately equally powerful.

In the working-point mode, the b-tagging weight for such permutations is $(1-\epsilon)\cdot(1-1/R)^2$, as an untagged jet is associated to the b quark, while all light quarks are still associated to untagged jets as well. The resulting suppression factor with respect to the correct permutation is given by $(1-\epsilon)/\epsilon$. For the used working points at 60%, 70% and 80% efficiency this amounts to suppression factors of 1.5, 2.3 and 4, respectively. It is noteworthy that all working points provide only a rather moderate suppression against such permutations.[4] They are therefore also referred to as the "moderately suppressed type" of permutations in the context of the working-point mode in the following.

[4] At 50% efficiency, which was a quite common choice in the top working group until the first half of 2011, there would be

Third type of permutation. Finally, the b-quark jet may be associated to a light quark. The number of this type of permutation in the three, four- and five-jet bins is two, six and twelve, respectively. In the three-jet bin, this is the only type of wrong permutation. In the semileptonic top decay mode, the top-quark mass constraint is sensitive to all of them, as none of these permutations associates the correct jet to the b quark. For the hadronic top decay mode, the two permutations to which the top-quark mass constraint is not sensitive are always of this type.

Besides the b-tagged jet being associated to a light quark, these permutations always come along with an untagged jet being associated to the b quark. The b-tagging weight for such permutations is $(1-\epsilon)\cdot(1-1/R)\cdot(1/R)$, and the corresponding suppression factor is given by $(1-\epsilon)/\epsilon\cdot R/(1-R)$. For the used working points at 60%, 70% and 80% efficiency, this amounts to 519, 223 and 88, respectively. One can see that the factor, being dominated by the rejection, decreases toward higher b-tagging efficiencies. For all working points the suppression of such permutations is still much stronger than for the second type, approximately by a factor R. Therefore, they are sometimes also referred to as the "strongly suppressed type" of permutation.

7.2.2 Expectations of the fit efficiencies in the no-b-tagging mode

General expectations. Based on the considerations in the previous section, some expectations of the behavior of fit efficiency can now be formulated, depending on the jet bin.

In the three-jet bin, there are only two wrong permutations. Both are of the third type. More specifically, these are exactly the two permutations against which the hadronic top-quark mass constraint is not sensitive, while there are no permutations against which the semileptonic top-quark mass constraint is insensitive. Consequently, one can expect that the fit efficiency for the semileptonic top decay mode is higher than for the hadronic one in the three-jet bin.

In the four-jet bin, the number of wrong permutations to which the top-quark mass constraint is insensitive is two for both top decay modes. Similar performance may therefore be expected, given the assumption holds that both top-quark mass constraints are approximately equally powerful.

In the five-jet bin, the number of permutations to which the semileptonic top-quark mass constraint is insensitive increases to five. The naive expectation would therefore be that the efficiency for the hadronic top decay mode should be higher, now. Taking into account that in practice only four out of five jets are passed to the KLFitter, one might rather expect that the relation between the fit efficiencies for the two modes is similar to the four-jet bin. However, the special treatment of the five-jet bin makes any such prediction much more difficult and speculative.

Quantitative estimate for the three-jet bin in the hadronic top decay mode. In addition to the general considerations above, there is also a way to provide even a quantitative estimate for fit efficiency in the three-jet bin of the hadronic top decay mode, which can be used to probe the KLFitter performance. As pointed out, in the three-jet bin of the hadronic top decay mode the only sensitivity to the only two wrong permutations is given by the hadronic W mass constraint. A rough quantitative estimate for the fit efficiency in this situation can therefore be obtained from the invariant mass distributions of all possible two-quark combinations without performing a fit. This distribution is shown in figure 7.5 for all combinations, for the true combination, for all non-true combinations and for the best combination, where "best" denotes the combination with an invariant mass closest to the W mass. The true combinations, as well as the best combinations, are centered around the true W mass, while the combinatorial

no extra suppression of these permutations at all, below 50% such permutations would even be enhanced with respect to the correct one.

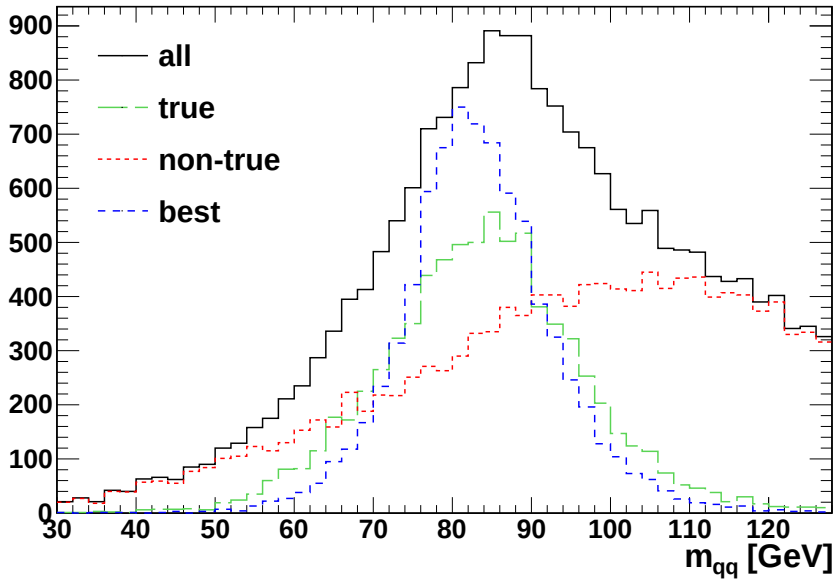


Figure 7.5: Distribution of quark-pair invariant masses in matched three-jet events with hadronic top-quark decay in the pretag selection. Shown are the distributions for all possible two-quark combinations in an event ("all"), for the true combination (i.e. the one that corresponds to the true W decay quarks, "true"), for all non-true combinations ("non-true"), and for the best combination (i.e. the one with an invariant mass closest to the W mass, "best").

background (i.e. the non-true combinations) has a much broader distribution and has typically higher invariant masses. However, the combinatorial background still has a substantial overlap with true combinations in the region around the true W mass. Consequently, the distribution of the best combination has a much stronger pronounced peak, with higher bin contents close to the mean value, than the one for the true combination, which directly shows that in many events a non-true combination must get picked as the best one. Doing the counting, one finds that in 65% of the events, the best combination is indeed the true combination.

If the KLFitter performs properly, it should therefore yield an efficiency of at least these 65%. If the KLFitter performs better than that, there are three obvious aspects in which the KLFitter differs from this simple approach and which may cause the improvement. The most apparent one is that the KLFitter actually performs a fit: in order to evaluate how "close" to the W mass a quark-pair invariant mass is, it does not rely on simply comparing the absolute difference in mass. It rather probes how compatible (according to the TFs) the measured jet energies are with those energies that would be necessary in order to actually reach the W mass. The two other aspects are the only details in which the hadronic top-quark mass constraint does in fact treat the b quark differently from the light quarks: the TFs and the quark mass. Any one of these three aspects would be a clear indication that the implementation of the KLFitter works well and that the TFs do their job.

7.2.3 Fit efficiencies in the no-b-tagging mode

Efficiencies in the pretag selection

General overview. An overview of the observed fit efficiencies in the no-b-tagging mode, based on the pretag selection and on the tag selection at all three working points, depending on the jet bin and the top decay mode is given in table 7.3. The overall efficiency (i.e. for all jet bins) in the pretag selection

Fit efficiency (no-b-tagging mode)	All-jet bin lep / had	Three-jet bin lep / had	Four-jet bin lep / had	Five-jet bin lep / had
Pretag selection	0.656 / 0.573	0.853 / 0.701	0.526 / 0.490	0.33 / 0.35
Tag selection (ϵ = 60%)	0.666 / 0.583	0.860 / 0.709	0.53 / 0.49	0.33 / 0.35
Tag selection (ϵ = 70%)	0.670 / 0.591	0.857 / 0.711	0.52 / 0.49	0.32 / 0.34
Tag selection (ϵ = 80%)	0.682 / 0.593	0.856 / 0.709	0.52 / 0.48	0.32 / 0.34
Combinatorial probability	-	0.333	0.083	0.033

Table 7.3: Overview of the fit efficiency of the KLFitter for semileptonic/hadronic top-quark decays in the no-b-tagging mode, depending on the jet bin and the selection. In addition the combinatorial probability for picking the correct permutation just by chance, depending on the jet bin, is given for comparison. Typical uncertainties on the efficiencies range from 0.4 percentage points in the three-jet bin in the pretag selection to 1.3 percentage points in the five-jet bin in the tag selection. For uncertainties above 0.9 percentage points, only 2 digits of the corresponding values are displayed.

is about 66% for the semileptonic top decay mode and about 57% for the hadronic top decay mode.

In the three-jet bin, the semileptonic and hadronic top decay modes yield fit efficiencies of about 85% and 70%, respectively. As expected, the efficiency is higher for the semileptonic decay mode.

In the four-jet bin, the fit efficiency goes down to about 53% and 49% for the semileptonic and hadronic top decay modes, respectively. The obvious reason for the decrease is statistics: the number of wrong permutations in the four-jet bin is eleven, compared to two in the three-jet bin. Therefore, the probability that a wrong permutation by chance looks better than the correct one naturally increases, irrespective of the type of permutations. The difference between the two top decay modes goes down to about four percentage points in the four-jet bin. Given the considerations presented above, this remaining difference can be interpreted such that the semileptonic top-quark mass constraint is presumably slightly more powerful than the hadronic top-quark mass constraint.

In the five-jet bin, the efficiencies are 33% and 35% for the semileptonic and hadronic top decay modes, respectively. Again, a decrease is anticipated due to the number of wrong permutations increasing from 11 to 29. Also the fact that now the efficiency is higher for the hadronic top decay mode is not unexpected, as the number of wrong permutations against which only the hadronic W mass constraint is sensitive is now five for the semileptonic mode compared to still two for the hadronic mode. However, these considerations have to be taken with care due to the special treatment of the five-jet bin.

Quantitative rating of the efficiencies. The question is now how well these efficiencies compare to what is possible in principle, i.e. whether they indicate that the KLFitter fitter works as intended. The comparison of the fit efficiencies to the combinatorial probabilities to pick up the correct permutation by random guessing, which are given in the last row of table 7.3, shows that in each jet bin the fit efficiencies are much higher than the combinatorial ones. Also the decrease toward higher jet bins is by far not as dramatic as for the combinatorial probabilities. In this regard, the KLFitter performs well, though this certainly does not provide a complete answer to the question.

A better probe for the KLFitter efficiency is given by the three-jet bin for the hadronic top decay mode, as discussed above. The efficiency one can achieve here by just looking at the invariant masses of all possible quark pairs was shown to be about 65%. The KLFitter achieves 70%. Also the cross-check using the AcerMC sample yields almost exactly the same numbers (66% and 70%). One can therefore

be confident that this gain in efficiency is indeed a systematic effect, though not very large in percentage points. This can be viewed as a very strong indication that the implementation of the KLFitter for the *Wt*-channel indeed works properly.

Dependence of the selection.

Comparing now pretag and tag selection in table 7.3, the observed differences are generally small. The most outstanding feature is in the overall fit efficiency, which increases for the tag selection with respect to the pretag selection for both top decay modes, and it increases further toward higher efficiency working points. In contrast, for each individual jet bin all differences between the selections are compatible with being statistical fluctuations; the largest difference between any two selections amounts to one percentage point. One can conclude that the impact of the selection on the fit efficiency is negligible. This is a good thing to know when looking at the fit efficiency in the *b*-tagging working-point mode, in order to disentangle effects from usage of *b*-tagging and from the selection. The observed increase in the overall efficiency can not be explained by the behavior of the individual jet bins, however, but only by a different composition of the full selection from the individual jet bins, notably by a higher fraction of three-jet events.

7.2.4 Fit efficiencies in the working-point mode in the tag selection

Now the effect of using the *b*-tagging working-point mode and its dependence on the working point is analyzed. In this section only tag selection is considered, where for the selection the same respective working point is always used as for the KLFitter. The advantage of the tag selection is that exactly one *b*-tagged jet is always present and therefore the effect of the working-point mode can be studied quite cleanly. This of course relies on the observation from the previous section, that the fit efficiency is virtually independent of the event selection.

From the considerations in Section 7.2.1, one can again extract some expectations of the fit efficiencies. These are now formulated jet bin by jet bin and then directly compared to the respective observed numbers, before proceeding to the next jet bin. The observed numbers are shown in table 7.4, depending on the working point, the jet bin and the top decay mode. For easier comparison, the first row of the table shows the efficiencies in the no-*b*-tagging mode in the pretag selection, which is the same as in table 7.3.

The three-jet bin

Expectations. The three-jet bin provides a main benchmark for the benefit of the working-point mode, as here the only two wrong permutations are of the strongly suppressed type. A significant increase in the fit efficiency compared to the no-*b*-tagging mode is anticipated. If such an increase is not observed, this would mean that even the strong suppression provides only little additional sensitivity to these permutations. In this case the working-point mode would presumably be only of limited use.

Assuming the working-point mode works as intended, the efficiencies in the three-jet bin may well end up somewhat close to 100%, at least well above 90%, given that they were already around 85% (for the semileptonic top decay mode) in the no-*b*-tagging mode. The efficiencies should decrease toward higher efficiency working points, following the behavior of the suppression factor. For the hadronic top decay mode, with its lower native sensitivity to the wrong permutations in the three-jet bin, the dependence of the fit efficiency on the working point efficiency should be higher for the hadronic top decay mode. In general, the missing native sensitivity of the hadronic top decay mode is expected to

Fit efficiency (WP-mode, tag selection)	All-jet bin lep / had	Three-jet bin lep / had	Four-jet bin lep / had	Five-jet bin lep / had
No-b-tagging mode (pretag sel.)	0.656 / 0.573	0.853 / 0.701	0.526 / 0.490	0.33 / 0.35
Working-point mode (ϵ = 60%)	0.777 / 0.809	0.965 / 0.961	0.65 / 0.71	0.44 / 0.49
Working-point mode (ϵ = 70%)	0.792 / 0.820	0.959 / 0.947	0.67 / 0.72	0.46 / 0.53
Working-point mode (ϵ = 80%)	0.807 / 0.813	0.951 / 0.927	0.69 / 0.71	0.47 / 0.54

Table 7.4: Overview of the fit efficiency in the working-point mode in the tag selection, depending on the top decay mode, the jet bin and the b-tagging working point. For the line without b-tagging, the pretag selection was used. Typical uncertainties on the efficiencies range from 0.4 percentage points in the three-jet bin in the no-b-tagging mode to 1.4 percentage points in the five-jet bin in the working-point mode. For uncertainties above 0.9 percentage points, only 2 digits of the corresponding values are displayed.

be compensated to some extent by the additional sensitivity due to the working point mode, such that the efficiency difference between the two top decay modes is smaller than in the no-b-tagging mode. The higher the additional sensitivity, i.e. the higher the achieved efficiency is, the smaller the efficiency difference. In particular, toward higher efficiency working points, going along with lower efficiencies, the difference between the top decay mode should show an increase.

In case the efficiencies in the three-jet bin indeed turn out to be somewhat close to one, as carefully anticipated, this enables some assumptions that are particularly helpful in order to formulate expectations about the higher jet bins. This is the reason why in contrast to the no-b-tagging mode, here the observed values are evaluated jet-bin by jet-bin.

Observed efficiencies. We start with the 60% efficiency working point, which is the one with the strongest suppression. As can be seen in table 7.4, the observed efficiencies in the three-jet bin are now 96.5% and 96.1% for the semileptonic and hadronic top decay modes, respectively. This is indeed a substantial improvement in particular for the hadronic top decay mode, as anticipated. The efficiency difference between the two top decay modes amounts to only half a percentage point, which indicates that the efficiency is indeed mostly driven by the b-tagging. At the 70% working point, the efficiency drops by about one half percentage point to 95.9% for the semileptonic top decay mode and by about 1.5 percentage points to 94.7% for the hadronic top decay mode. At the 80% working point, the efficiency for the hadronic top decay mode drops by another two percentage points to 91.6%, while for the semileptonic top decay mode it only drops by a bit less than one percentage point to 95.1%. One can conclude that the observed efficiencies nicely fulfill all the basic expectations: very high overall efficiencies, which decrease toward higher b-tagging efficiencies, a smaller difference than in the no-b-tagging mode between the two top decay modes, which increases toward higher efficiencies due to a stronger dependence on the working point in the hadronic top decay mode.

The four-jet bin

Expectations. The four-jet bin has eleven wrong permutations. Six of those are of the strongly suppressed type. From the results in the three-jet bin, we expect the efficiency to reject these to be close to one, at least well above 90% at all working points. Also the difference between the two top decay modes regarding these six permutations can be expected to be as small as in the three-jet bin. In

fact, as the hadronic top-quark mass constraint is insensitive to only two out of these six permutations, in contrast to two out of two in the three-jet bin, the difference can be expected to be even smaller here. Of the remaining five permutations, three are of the moderately suppressed type. Both decay modes furthermore have full native sensitivity to these permutations. The observed fit efficiencies in the no-b-tagging mode indicate that this sensitivity is slightly higher for the semileptonic top decay mode. Given the additional sensitivity by the b-tagging, even if it is only moderate, this difference will be less pronounced, now. Two wrong permutations remain, which are of the type to which the working-point mode provides no extra sensitivity. Furthermore, these are those permutations against which the semileptonic top-quark mass constraint is insensitive. These two permutations can therefore be expected to be the main source of difference in fit efficiency between the two top decay modes. This implies that the fit efficiency should now be higher for the hadronic than for the semileptonic top decay mode. A strong working point dependence of the efficiency difference between the two top decay modes is not expected, as the presumed main source of the difference is not affected at all by the b-tagging. The fit efficiency itself is expected to increase with the working point efficiency: the efficiency against the strongly suppressed permutation was shown to depend rather little on the working point, while the suppression against the moderately suppressed permutations increases toward higher b-tagging efficiency.

The four-jet bin is particularly interesting in some regards. Assume that not only the strongly suppressed, but also the moderately suppressed permutations can be rejected with about 100% efficiency and can therefore be neglected. For the former ones this is an approximation, for the latter ones it is probably a rather optimistic assumption. In this scenario, the four-jet bin in the working-point mode looks like three-jet bin in the no-b-tagging mode, but with the roles changed between the two top decay modes: there are two remaining wrong permutations; to these, the semileptonic top decay mode is only sensitive via the hadronic W mass constraint, while for the hadronic top decay mode also the top-quark mass constraint is sensitive. The fit efficiency for the semileptonic top decay mode can therefore be probed against the 70% for the hadronic mode in the no-b-tagging mode three-jet bin and the 65% from the simple diquark invariant mass example. Any efficiency significantly lower than those, particular any efficiency below 65%, indicates that the moderately suppressed permutations in fact cannot be neglected. In the same way the efficiency for the hadronic top decay mode can be compared to the no-b-tagging mode three-jet bin of the semileptonic top decay mode. In addition, the presumed difference in discrimination power between the semileptonic and hadronic top-quark mass constraint, as concluded from the no-b-tagging mode four-jet bin, should also show an effect here. In particular a smaller difference between the two top decay modes than in the no-b-tagging mode three-jet bin would be an indication for that.

Observed efficiencies. Table 7.4 shows that the observed efficiencies in the four-jet bin are indeed higher for the hadronic top decay mode, now, as expected. Also they mostly increase with the working point efficiency. The only exception to this is in the 80% working point in the hadronic top decay mode, where the efficiency is lower than for the 70% working point. This unexpected behavior might, however, be a statistical fluctuation related to the selection rather than being due to the b-tagging mode (the statistical uncertainty in the four-jet bin in the tag selection is about 0.9 percentage points). An indication for this comes from the no-b-tagging mode four-jet bin of the hadronic top decay mode (see table 7.3): already here the 80% selection has a lower efficiency by one percentage point than the three other selections, which all agree within 0.1 percentage points. Further support for the assumption that this number is an outlier comes from the cross-check with the AcerMC sample, where the 80% working point does show at least a slight increase with respect to the 70% working point. Apart from this

presumed outlier, the efficiency difference between the two decay modes ranges from 5 to 6 percentage points, i.e. the dependence on the working point is small as expected, though based on only two working points this is certainly not very conclusive. Support comes again from the AcerMC sample, where only a small dependence is observed as well.

The highest efficiency that is achieved for the semileptonic top decay mode is 69%, using the 80% working point. This is a bit lower than the 70% for the hadronic top decay mode three-jet bin in the no-b-tagging mode, but still higher than the 65% for the diquark invariant mass method (see Section 7.2.2). For the hadronic mode the highest efficiency is 72%, using the 70% working point. This is significantly lower than the 85% for the semileptonic top decay mode three-jet bin in the no-b-tagging mode. Both numbers indicate that indeed the moderately suppressed permutations are not rejected with 100% efficiency. Furthermore, both numbers further feed the assumption that the semileptonic top-quark mass constraint is slightly stronger than the hadronic one. As the efficiency difference between the two top decay modes is actually so much smaller than in the no-b-tagging mode three-jet bin, this may indicate other effects being involved in addition. For instance the kinematics of the jets may differ between the three-jet bin and the four-jet bin situation, such that the hadronic W mass constraint yields different efficiencies.

The five-jet bin

Expectations. Predictions for the five-jet bin are again difficult due to its special treatment. Basically there are only two general expectations that one can state with some confidence. Firstly the overall efficiency should again be lower than in the four-jet bin. Secondly the efficiency for the hadronic top decay mode should be higher than for the semileptonic mode, due to the composition of the wrong permutations of different types.

Observed efficiencies. Looking at table 7.4 one observes that indeed all fit efficiencies are lower than in the four-jet bin and the efficiencies are higher for the hadronic top decay mode than for the semileptonic one. Apart from that, the five-jet bin looks pretty much like a perfect four-jet bin, just with lower total efficiencies: the efficiencies for both top decay modes increase with the working point efficiency, while the difference between them is rather constant. This looks intriguingly like a consequence of the special treatment of the five-jet bin as some kind of pseudo-four-jet-bin. However, this is certainly not conclusive, although it is reproduced to some extent with the AcerMC sample (the efficiency difference between the top decay modes is not that stable there; rather it varies between four and seven percentage points).

7.2.5 Fit efficiencies in the working-point mode in the pretag selection

Finally, in this section the fit efficiencies for the working-point mode in the pretag selection are analyzed. Table 7.5 shows these fit efficiencies. The first row of the table again shows the efficiencies in the pretag in the no-b-tagging mode, which are the same as in tables 7.3 and 7.4. Keeping the simplifying assumptions from Section 7.2.1, pretag selection can be considered to be composed of events with zero or one b-tagged jet, where the relative fraction of each type is given by the working point efficiency. In particular in the three-jet bin this assumption should mostly be valid for matched events. In events with no b-tagged jets, all permutations retrieve the same b-tagging weight, therefore for these events the fit efficiency is the same as in the no-b-tagging mode. For events with one b-tagged jet, the fit efficiencies are the same as for the tag selection in the working-point mode. The fit efficiencies in the individual jet bins in the pretag selection should therefore be approximately the weighted means of the

efficiencies with and without using b-tagging. Besides events that contain more than one b-tagged jet, this approximation neglects the fact that the efficiencies for the no-b-tagging mode were obtained on the full pretag sample and the efficiencies on only the events without b-tagged jets may differ slightly. But as those efficiencies were shown to change only slightly for the tag selection, it can be assumed that on the complementary set of events, i.e. for the zero-b-tagged-jets events, the same holds.

Fit efficiency (WP-mode, pretag selection)	All-jet bin lep / had	Three-jet bin lep / had	Four-jet bin lep / had	Five-jet bin lep / had
No-b-tagging mode	0.656 / 0.573	0.853 / 0.701	0.526 / 0.490	0.33 / 0.350
Working-point mode (ϵ = 60%)	0.730 / 0.719	0.917 / 0.859	0.610 / 0.635	0.42 / 0.46
Working-point mode (ϵ = 70%)	0.745 / 0.737	0.924 / 0.869	0.631 / 0.659	0.44 / 0.49
Working-point mode (ϵ = 80%)	0.747 / 0.737	0.923 / 0.868	0.639 / 0.658	0.44 / 0.50

Table 7.5: Overview of the fit efficiency in the working-mode in the pretag selection, depending on the top decay mode, the jet bin and the b-tagging working point. Typical uncertainties on the efficiencies range from 0.4 percentage points in the three-jet bin to one percentage point in the five-jet bin For the latter, only 2 digits of the corresponding values are displayed.

Fit efficiency (estimate from weighted mean)	Three-jet bin lep / had	Four-jet bin lep / had	Five-jet bin lep / had
Working-point mode (ϵ = 60%)	0.923 / 0.860	0.599 / 0.620	0.39 / 0.44
Working-point mode (ϵ = 70%)	0.928 / 0.876	0.626 / 0.653	0.42 / 0.47
Working-point mode (ϵ = 80%)	0.933 / 0.883	0.654 / 0.665	0.44 / 0.50

Table 7.6: Estimation of the fit efficiencies in the working-point mode in the pretag selection by a weighted mean of the observed fit efficiencies in the no-b-tagging mode in the pretag selection and in the working-point mode in the tag selection, depending on the top decay mode, the jet bin and the working point. For all values the same number of digits as for the observed values is displayed

Table 7.6 shows the numbers that one obtains from this naive estimation. One can see that the estimation works remarkably well, in particular in the three-jet bin, where the impact of two-b-tagged-jets events is least. The naive estimate tends to overestimate the working point dependence of the efficiencies, in particular the difference between the 70% and the 80% working point. The effect is more pronounced in the four- and five-jet bin than in the three-jet bin. This holds for both top decay modes in the same way.

7.2.6 Summary of fit efficiencies

In summary it can be said that the obtained efficiencies and their behavior depending on the jet bin and the b-tagging working point (in case b-tagging was used) are generally understood from considerations about the structure of the kinematic fit and the permutations. Additionally in some cases the efficiencies could also be probed quantitatively. All results indicate that the KLFitter performs well and that all its

components do their jobs. In particular it can be stated that the KLFitter indeed does fulfill the aim of correctly assigning the measured objects to the final-state model particles.

7.3 Parameter resolutions before and after the fit

The second aspect of KLFitter performance which is topic of this chapter is the improvement of the fit parameter values with respect to their measured values due to the fit. This is examined in this section.

7.3.1 Approach

Any improvement due to the fit can only be expected when the fit hypothesis is actually fulfilled. Consequently, only matched events are used for the study of the parameter resolutions, and only the fit of the correct permutation using the correct top-quark decay mode is considered. Both the measured values of the fit parameters and their final values after the fit are then compared to the respective true values, where the true value of a parameter is defined via the association of the truth particles to the measured and the fit particles, respectively. For the fit particles, the association of the quarks is given by the quark-fit matching, while for the lepton and the neutrino it is trivial; the association of measured particles is given via their association to the fit particles. The measured value of the z component of the neutrino momentum is given by its initial value as used for the fit. For the sake of simplicity, the talk is about the *measured value*, the *fit value* and the *true value* of the fit parameters in the following.

For any fit parameter P, the measured/fit ΔP is defined as the difference between the measured/fit value $P_{\text{meas/fit}}$ and the true value P_{true} of the parameter, respectively:

$$\Delta P = P_{\text{meas/fit}} - P_{\text{true}} \,.$$

The neutrino momentum is parametrized in Cartesian coordinates within the KLFitter. In addition to these, also the transverse momentum and the azimuthal angle are considered in this section. For the charged lepton, the fit parameter is given by its energy in case it is an electron and by its transverse momentum in the muon channel. Both the lepton energy and the transverse momentum are considered, here, but without distinguishing between the electron and the muon channel.

A Gaussian was fitted to all the considered ΔP distributions. The used fit range is derived from an iterative procedure. The first fit is performed within a 1.5 RMS range around the mean value of the histogram. Each subsequent fit is performed on the 1.5σ range around the mean value of the previous one, until the range converges (which normally happens after a few iterations, as the range is binned). The width (σ) and the mean of all final fits along with their statistical errors are then analyzed, providing measures for the resolution and the bias of the measured and the fit values.

7.3.2 Expectations

In order to come up with some expectations concerning the parameter widths, one has to look again at which parameters are constrained by the same mass constraint and compare how well they are determined before the constraint is applied. "Determined" here primarily means "measured", where the measurement precision in terms of the KLFitter is defined by the TF, but it may also include another mass constraint that already acts on a parameter. The basic idea is that when a mass constraint acts on two or more parameters, the one that is known least precisely will be changed – presumably improved – most by the constraint, while a parameter that is measured very precisely will mostly help to improve the other parameters. If a constraint acts on two parameters that are measured approximately equally

precisely, both will profit by approximately the same amount, in contrast. Based on this consideration, some concrete expectations for the different parameters can now be formulated, depending on the top decay mode.

Leptonic parameters. The leptonic parameters are always constrained by the leptonic W mass. However, the leptonic W mass constraint is eaten up by the neutrino z momentum (see Section 6.2.4), while the remaining parameters are not changed at all (the case when there is no exact solution for the neutrino z is again neglected for now). In the hadronic top decay mode that's all; the leptonic parameters get fixed to their initial values, no improvement can be expected for them. In the semileptonic top decay mode the leptonic parameters are further constrained by the top-quark mass constraint, together with the b-quark energy. Among all these parameters the lepton energy/momentum is measured most precisely, while the neutrino parameters have the poorest precision. Some improvement can therefore be expected for the neutrino parameters, while the lepton energy/momentum is expected to change only a little bit.

Light-quark energies. The two light-quark energies are always constrained by the hadronic W mass. As they both use the same TF, i.e. they are measured equally precisely, they can be expected to profit by the same amount from the W mass constraint. For the semileptonic top decay mode that's all for these parameters. In contrast to the leptonic parameters in the hadronic top decay mode, here some improvement due to the W mass constraint is expected. In the hadronic top decay mode the light-quark energies are further constrained by the top-quark mass constraint, together with the b-quark energy. Here they are the more precisely determined parameters. Firstly the TF for b-quark energy is a bit wider than the one for the light-quark energies due to neutrinos and/or muons escaping the jet; secondly the light-quark energies have already gained additional precision from the W mass constraint. Therefore the additional improvement of the light-quark energies, and hence the difference in improvement between the two top decay modes, can be expected to be rather small.

b-quark energy. The b-quark energy always gets constrained by the top-quark mass constraint and nothing else. In the semileptonic top decay mode the leptonic parameters enter the constraint in addition. The b-quark energy is measured less precisely than the lepton energy/momentum, but still much more precisely than the neutrino parameters. Therefore the b-quark energy will only obtain a reduced benefit from the top-quark mass constraint compared to the neutrino parameters. The leptonic W mass constraint, which additionally acts on the leptonic parameters, does not play a role in this discussion, as it is serves only to determine the neutrino z momentum.

In the hadronic top decay mode instead the light-quark energies enter the constraint in addition to the b-quark energy. As discussed above, the light-quark energies can be expected to be measured more precisely than the b-quark energy here. Thus, the major benefit from the top-quark mass constraint should go to the b-quark energy. In particular the improvement of the b-quark energy should certainly be larger in the hadronic top decay mode than in the semileptonic decay mode.

7.3.3 Results in the pretag selection

Figures 7.6, 7.7, 7.8, 7.9 and 7.10 show the distributions of both the measured (top row) and the fit (second and third row) ΔP for $P \in \{E_q, E_{q'}, E_b, p_{\nu,z}, p_{\nu,x}, p_{\nu,y}, p_{\nu,\mathrm{T}}, \phi_\nu, E_\ell, p_{\ell,\mathrm{T}}\}$, based on the fit of the correct permutation in all matched events in the pretag selection. The fit ΔP is shown for the semileptonic (second row) and the hadronic (third row) top decay mode separately. The width and the mean of all final fits are summarized along with their statistical errors in table 7.7. The effect of the kinematic fit on the widths and the means is discussed in the following, parameter by parameter.

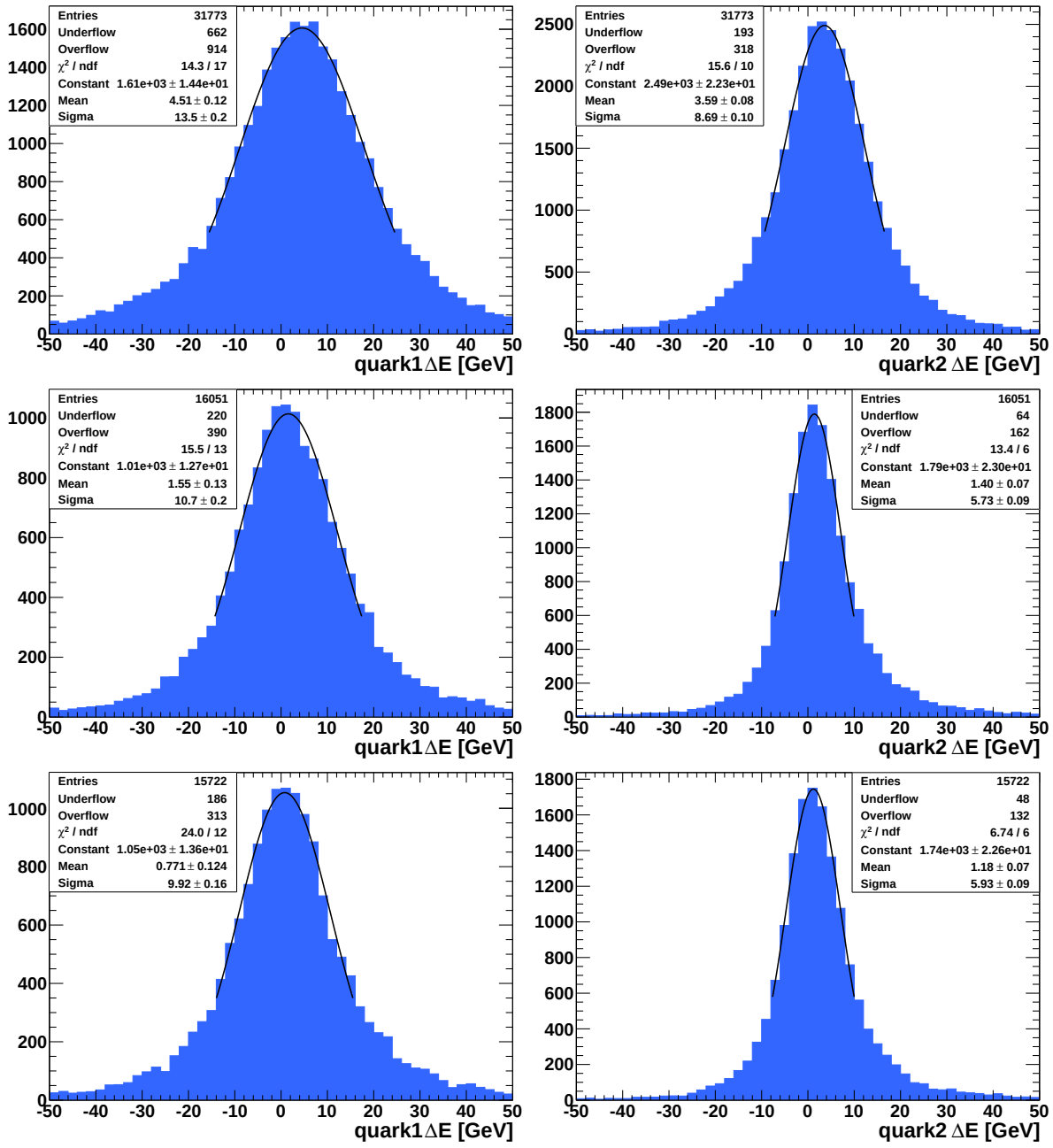


Figure 7.6: ΔP distributions for the energy of the first (i.e. higher-energetic, left column) and second (i.e. lower-energetic, right column) light quark before the fit (top row), after the semileptonic top decay mode fit (middle row) and after the hadronic top decay mode fit (bottom row) in the pretag selection. For the definition of ΔP, details about which events are used and how the Gaussian fit is done, see Section 7.3.1.

Light-quark energies

Figure 7.6 shows the ΔP distribution for the energies of the light quarks, where the first quark (left column) is always the higher energy one and the second quark (right column) is the lower energy one. The width for the first quark is 13.5 GeV before the fit and gets reduced to about 10-11 GeV after the fit. The width is smaller by 0.75 GeV for the hadronic top decay mode. The mean value decreases from 4.5 GeV to 1.6 GeV and 0.8 GeV in the semileptonic and hadronic top decay modes, respectively. For the second quark, the width is 8.7 GeV before the fit and about 5.8 GeV after the fit. The difference between the two top decay modes is only 0.2 GeV here, which is hardly significant however compared to the statistical errors. The mean value decreases from 3.6 GeV to 1.4 GeV and 1.2 GeV, respectively. Again the difference between the two top decay modes is hardly significant. In summary the KLFitter both reduces the bias and improves the resolution of the light-quark energies significantly, where the difference between the two top decay modes is moderate for the first quark and small for the second quark. This matches very nicely the expectation formulated in Section 7.3.2.

b-quark energy

The ΔP distributions for the energy of the b quark are shown in figure 7.7 (left). The first striking point about the b-quark energy is that before the fit the distribution is strongly asymmetric with a long negative tail (note the underflow bin), while after the fit this asymmetry is mostly gone. As explained in Section 6.1.3, this asymmetry can be traced back to semileptonic decays of hadrons into muons within the b-quark jets and is taken into account by the asymmetric shape of the TF for E_b. The disappearance of the negative tail can therefore be seen as a particular success of the TFs – even more as the capability of employing asymmetric TFs is a unique feature of the likelihood approach that is used in the KLFitter, compared to a χ^2 fit, which inherently assumes symmetric errors.

Also it is visible even by eye that the improvement is substantially larger in the hadronic top decay mode. Having a look at the numbers, the width is reduced by the fit from 13 GeV to 11 GeV and 6 GeV in the semileptonic and hadronic mode, respectively. The bias on E_b goes down from -1.9 GeV before the fit to 0.41 GeV and 0.38 GeV after the fit, respectively. It is interesting that there is virtually no difference between the top decay modes concerning the mean, while the performance on the resolution is substantially better in the hadronic top decay mode. Overall the observed results nicely match the expectation of a moderate improvement in the semileptonic top decay mode and a significantly larger improvement in the hadronic top decay mode.

Neutrino z momentum

The ΔP distributions for the z component of the neutrino momentum are shown in figure 7.7 (right). The mean values of all three distributions are compatible with zero bias compared to the respective statistical errors. The behavior of the widths is quite surprising. What one observes is that the value before the fit is 25 GeV, while after the fit it is 27 GeV for the semileptonic and 20 GeV for the hadronic top decay mode. In other words: in the semileptonic top decay mode, where some improvement is expected, the width increases instead, while in hadronic top decay mode, where fit does not change this parameter at all, an improvement is observed. Assuming that the fixing of the parameter is implemented correctly and indeed happens as intended, this leads to the conclusion that the $p_{\nu,z}$ resolution must differ between the top decay modes already before the fit, e.g. due to different kinematics of the leptonic W boson between the two top decay modes. In this case the observed width after the fit in the semileptonic top decay mode would probably even indeed be an improvement with respect to before the fit. This idea gets supported by a cross-check that was performed by fixing $p_{\nu,z}$ to its initial value also in the semileptonic top decay

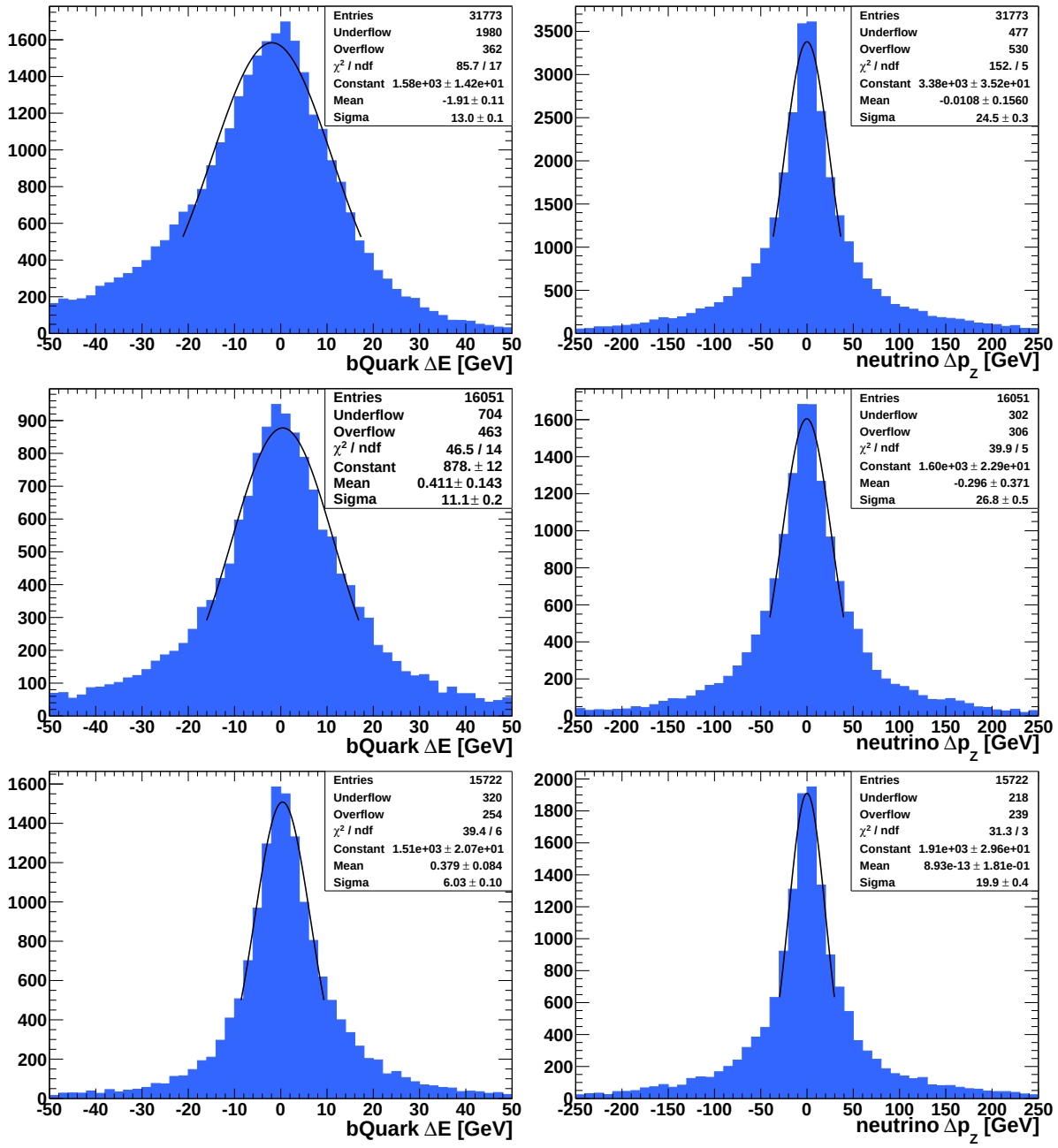


Figure 7.7: ΔP distributions for the energy of the b quark (left column) and the z component of the neutrino momentum (right column) before the fit (top row), after the semileptonic top decay mode fit (middle row) and after the hadronic top decay mode fit (bottom row) in the pretag selection. For the definition of ΔP, details about which events are used and how the Gaussian fit is done, see Section 7.3.1.

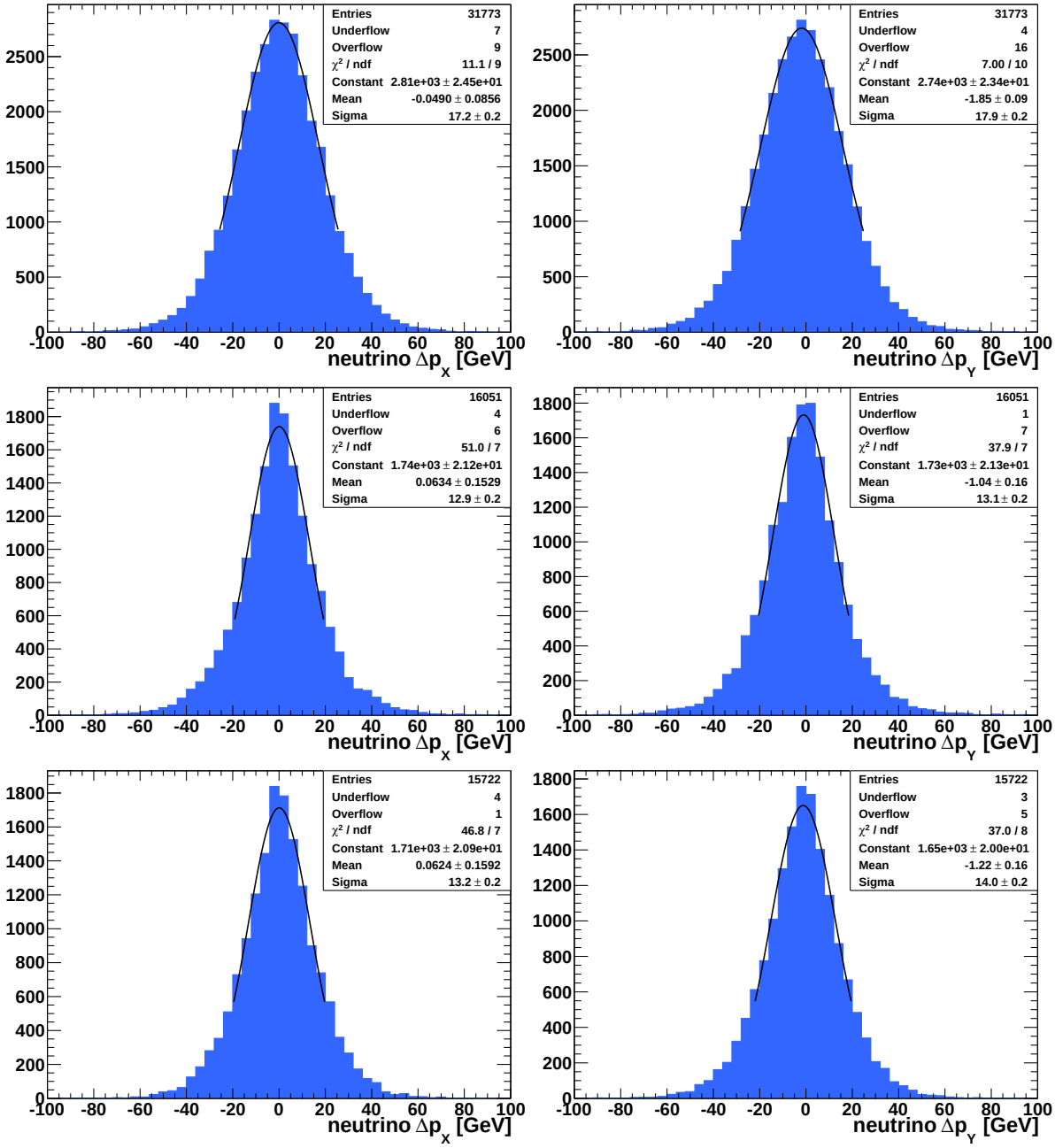


Figure 7.8: ΔP distributions for the x (left column) and y (right column) component of the neutrino momentum before the fit (top row), after the semileptonic top decay mode fit (middle row) and after the hadronic top decay mode fit (bottom row) in the pretag selection. For the definition of ΔP, details about which events are used and how the Gaussian fit is done, see Section 7.3.1.

mode, which did not lead to any apparent improvement of the resolutions, but rather the fit efficiencies tended to decrease slightly.

Neutrino x and y momentum

Figure 7.8 shows the ΔP distributions for the x and y components of the neutrino momentum. For the mean value of the x component there is no significant bias in any of the three distributions. For the y component, there is a bias of -1.85 GeV before the fit. Even after the fit, a bias of -1 GeV and -1.2 GeV remains for the semileptonic and hadronic top decay modes, respectively. At first glance this is very surprising, as neither there seems to be any apparent reason for the x and the y component to behave differently, nor for any bias at all. However, this behavior is in fact observed in the very same way also with the AcerMC sample, so there should be a reason, possibly related to the detector. A likely explanation is in fact given by the acceptance gaps of the muon spectrometer at the positions of the main detector supports. A muon escaping through these gaps would indeed give rise to considerable fake $p_{\mathrm{T}}^{\mathrm{miss}}$ in the negative y direction, while the x component cancels out on average. Still it is a bit surprising that this seems to have such a clearly observable effect.

Concerning the widths of the distributions, $p_{\nu,x}$ and $p_{\nu,y}$ behave mostly the same, just as expected, though still the y component has a slightly larger width than the x component in all three distributions. Before the fit, the width amounts to about 17.2 GeV and 17.9 GeV for $p_{\nu,x}$ and $p_{\nu,y}$, respectively, after the fit this gets reduced to about 12.9 GeV and 13.1 GeV for the semileptonic top decay mode and 13.2 GeV and 13.9 GeV for the hadronic top decay mode.

As expected, the improvement due to the fit is larger for the semileptonic top decay mode, both for the means and the widths. In fact, for the hadronic mode it would have rather been expected that there is hardly any improvement at all, as $p_{\nu,x}$ and $p_{\nu,y}$ are only changed by the fitter in case there is no exact solution for $p_{\nu,z}$, which happens in about 30% of all events. Therefore, one can again suspect that for the hadronic top decay mode the width was narrower than the average already before the fit. At the same time that would mean an even stronger improvement in the semileptonic mode than indicated by the numbers. This would furthermore give a quite consistent picture with the situation for $p_{\nu,z}$.

Neutrino p_{T} and ϕ

Figure 7.9 shows the ΔP distributions for the transverse momentum and the azimuthal direction of the neutrino momentum. The transverse momentum before the fit shows a large bias of 9 GeV. After the fit this gets reduced to 4 GeV and 5 GeV for the semileptonic and hadronic top decay modes, respectively. A bias in the transverse momentum should also show effect in the distributions for $p_{\nu,x}$ and $p_{\nu,y}$. As the transverse momentum is invariant against the sign of $p_{\nu,x}$ and $p_{\nu,y}$, it does not show up as a bias there, but rather increases the width of the distributions compared to the width of the distribution for $p_{\nu,\mathrm{T}}$. This is in fact observed. The width of the $p_{\nu,\mathrm{T}}$ distribution before and after the semileptonic and hadronic top decay mode fit amounts to 15.6 GeV, 12.1 GeV and 12.3 GeV, respectively, which is significantly smaller than for the individual components in all cases. For the azimuthal direction of the neutrino momentum no pronounced bias is present in any of the distributions. The width improves from 0.246 rad before the fit to 0.217 rad and 0.209 rad, respectively. Again, a clear improvement is visible and the semileptonic top decay mode looks better, in agreement with the expectation. The only exception is given by the resolution of the azimuthal direction, where the hadronic top decay mode looks slightly better. Compared to the statistical errors this is hardly significant, however. Assuming again that the hadronic top decay mode looked better already before the fit, the fit even must have performed better for the semileptonic top decay mode.

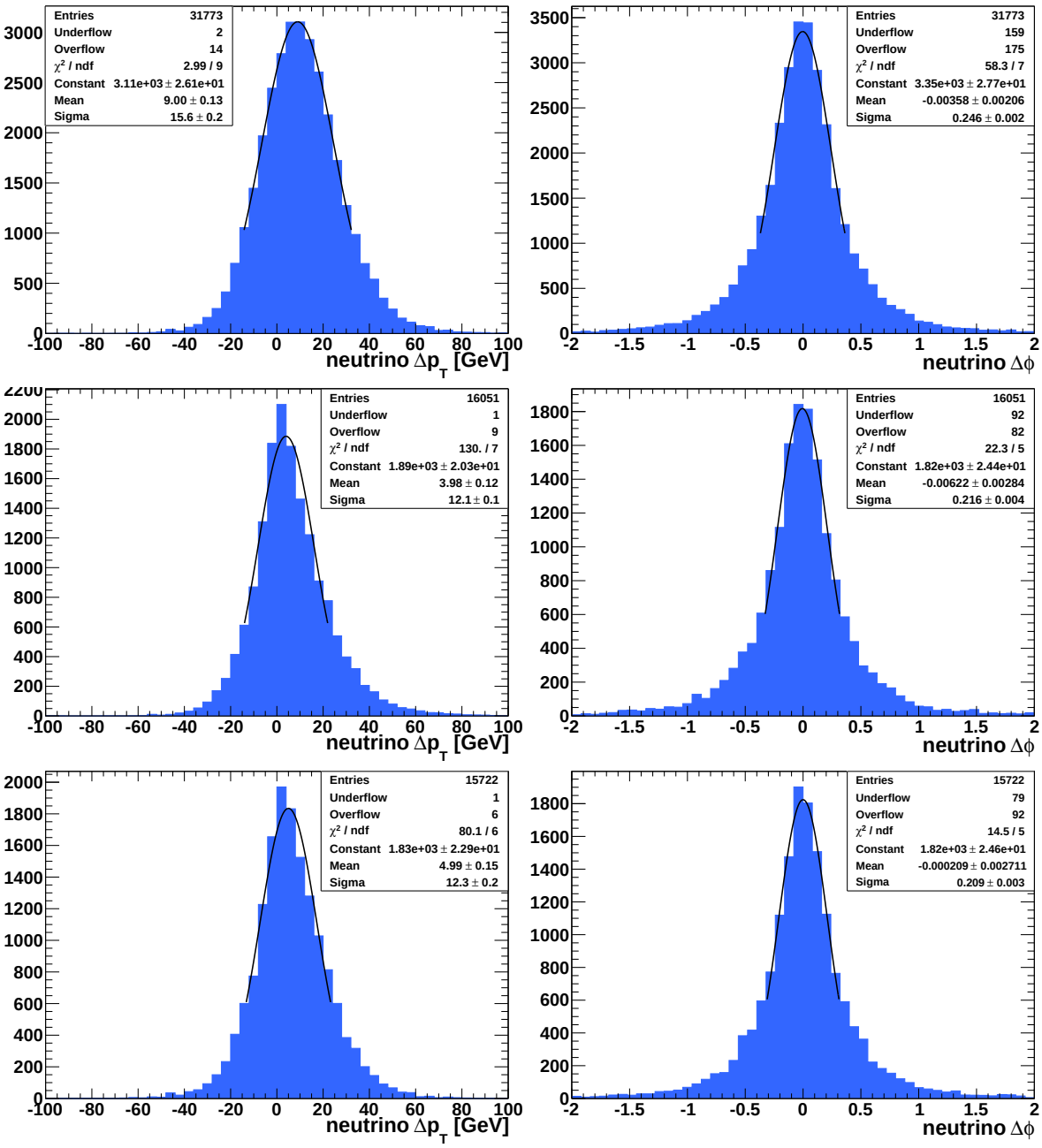


Figure 7.9: ΔP distributions for the transverse momentum (left column) and the azimuthal direction (right column) of the neutrino before the fit (top row), after the semileptonic top decay mode fit (middle row) and after the hadronic top decay mode fit (bottom row) in the pretag selection. For the definition of ΔP, details about which events are used and how the Gaussian fit is done, see Section 7.3.1.

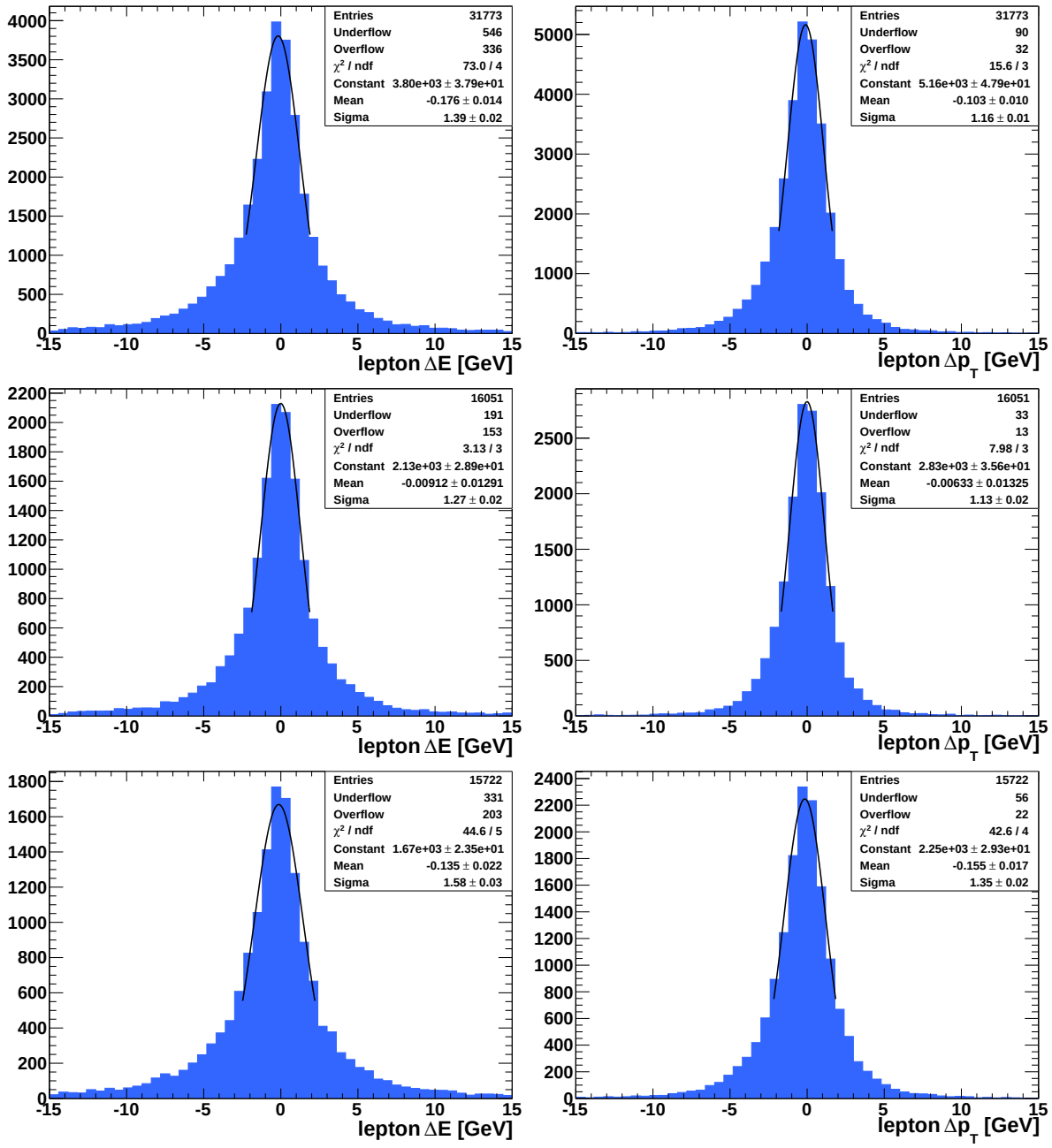


Figure 7.10: ΔP distributions for the energy (left column) and transverse momentum (right column) of the lepton before the fit (top row), after the semileptonic top decay mode fit (middle row) and after the hadronic top decay mode fit (bottom row) in the pretag selection. For the definition of ΔP, details about which events are used and how the Gaussian fit is done, see Section 7.3.1.

Lepton energy and transverse momentum

Figure 7.10 shows the ΔP distributions for the lepton energy and transverse momentum. A small but, due to the even smaller errors, significant bias can be seen before the fit and after the hadronic top decay mode fit; only in the semileptonic top decay mode does the bias disappear after the fit. The widths even increase in three out of the four plots after the fit. The differences with respect to before the fit are small but significant. While only a small effect of the fit was expected, the increasing widths were certainly not. As the lepton is still the by far most precise part of the fit and the differences are in fact almost negligible compared to the precision of the jets and the neutrino, this is considered to be unproblematic.

Mean	Before the fit	Leptonic fit	Hadronic fit
E_b [GeV]	-1.91 ± 0.11	0.41 ± 0.14	0.379 ± 0.084
E_q [GeV]	4.51 ± 0.12	1.55 ± 0.13	0.77 ± 0.12
$E_{q'}$ [GeV]	3.589 ± 0.077	1.397 ± 0.071	1.176 ± 0.075
E_ℓ [GeV]	-0.176 ± 0.014	-0.009 ± 0.013	-0.135 ± 0.022
$p_{l,\mathrm{T}}$ [GeV]	-0.103 ± 0.010	-0.006 ± 0.013	-0.155 ± 0.017
$p_{\nu,x}$ [GeV]	-0.049 ± 0.086	0.06 ± 0.15	0.06 ± 0.16
$p_{\nu,y}$ [GeV]	-1.852 ± 0.090	-1.04 ± 0.16	-1.22 ± 0.16
$p_{\nu,z}$ [GeV]	-0.01 ± 0.16	-0.30 ± 0.37	0.00 ± 0.18
$p_{\nu,\mathrm{T}}$ [GeV]	9.00 ± 0.13	3.98 ± 0.12	4.99 ± 0.1501
ϕ_ν [mrad]	-0.0036 ± 0.0021	-0.0062 ± 0.0028	-0.0002 ± 0.0027
Width	**Before the fit**	**Leptonic fit**	**Hadronic fit**
E_b [GeV]	12.97 ± 0.13	11.07 ± 0.18	6.03 ± 0.10
E_q [GeV]	13.50 ± 0.15	10.67 ± 0.16	9.92 ± 0.16
$E_{q'}$ [GeV]	8.689 ± 0.098	5.734 ± 0.086	5.930 ± 0.094
E_ℓ [GeV]	1.393 ± 0.017	1.266 ± 0.023	1.576 ± 0.027
$p_{l,\mathrm{T}}$ [GeV]	1.165 ± 0.013	1.126 ± 0.016	1.350 ± 0.021
$p_{\nu,x}$ [GeV]	17.19 ± 0.21	12.92 ± 0.18	13.21 ± 0.19
$p_{\nu,y}$ [GeV]	17.92 ± 0.20	13.09 ± 0.20	13.95 ± 0.20
$p_{\nu,z}$ [GeV]	24.53 ± 0.29	26.84 ± 0.49	19.95 ± 0.39
$p_{\nu,\mathrm{T}}$ [GeV]	15.61 ± 0.16	12.11 ± 0.14	12.27 ± 0.20
ϕ_ν [mrad]	0.2461 ± 0.0022	0.2165 ± 0.0037	0.2089 ± 0.0034

Table 7.7: Overview of the means (upper half of the table) and widths (lower half) of the Gaussian fits to all the ΔP distributions before the fit, after the semileptonic top decay mode fit ("Leptonic fit") and after the hadronic top decay mode fit ("Hadronic fit") in the pretag selection. For the definition of ΔP, details about which events are used and how the Gaussian fit is done see Section 7.3.1.

7.3.4 Impact of tag selection

Tables 7.8 and 7.9 show the means and the widths of the ΔP distributions that are obtained in the tag selection, depending on the working point. In addition, the numbers for the pretag selection are shown again for easier comparison. One finds that the differences between the selections are mostly in the order of the statistical errors, while any larger differences are very rare. Therefore one can conclude confidently that the choice of the selection does not seem to have any systematic effect.

Semileptonic top decay mode				
Mean	Pretag	Tag at ϵ = 60%	Tag at ϵ = 70%	Tag at ϵ = 80%
E_b [GeV]	0.41 ± 0.14	0.27 ± 0.19	0.26 ± 0.18	0.17 ± 0.19
E_q [GeV]	1.55 ± 0.13	1.43 ± 0.18	1.40 ± 0.17	1.34 ± 0.16
$E_{q'}$ [GeV]	1.397 ± 0.071	1.427 ± 0.091	1.47 ± 0.090	1.48 ± 0.10
E_ℓ [GeV]	-0.009 ± 0.013	-0.039 ± 0.025	0.008 ± 0.021	-0.007 ± 0.019
$p_{l,\mathrm{T}}$ [GeV]	-0.006 ± 0.013	-0.003 ± 0.018	0.007 ± 0.018	-0.006 ± 0.018
$p_{\nu,x}$ [GeV]	0.06 ± 0.15	-0.05 ± 0.20	0.03 ± 0.19	0.24 ± 0.20
$p_{\nu,y}$ [GeV]	-1.04 ± 0.16	-1.13 ± 0.20	-1.02 ± 0.20	-1.03 ± 0.20
$p_{\nu,z}$ [GeV]	-0.30 ± 0.37	-0.56 ± 0.50	-0.31 ± 0.48	-0.26 ± 0.50
$p_{\nu,\mathrm{T}}$ [GeV]	3.98 ± 0.12	3.25 ± 0.18	3.32 ± 0.18	3.28 ± 0.18
ϕ_ν [mrad]	-0.0062 ± 0.0028	-0.0083 ± 0.0039	-0.0079 ± 0.0037	-0.0036 ± 0.0024
Hadronic top decay mode				
Mean	Pretag	Tag at ϵ = 60%	Tag at ϵ = 70%	Tag at ϵ = 80%
E_b [GeV]	0.379 ± 0.084	0.38 ± 0.11	0.36 ± 0.11	0.24 ± 0.10
E_q [GeV]	0.77 ± 0.12	0.77 ± 0.16	0.71 ± 0.16	0.77 ± 0.13
$E_{q'}$ [GeV]	1.176 ± 0.075	1.14 ± 0.10	1.203 ± 0.093	1.16 ± 0.094
E_ℓ [GeV]	-0.135 ± 0.022	-0.092 ± 0.028	-0.105 ± 0.027	-0.102 ± 0.027
$p_{l,\mathrm{T}}$ [GeV]	-0.155 ± 0.017	-0.121 ± 0.022	-0.135 ± 0.022	-0.133 ± 0.023
$p_{\nu,x}$ [GeV]	0.06 ± 0.16	0.40 ± 0.21	0.08 ± 0.20	0.16 ± 0.20
$p_{\nu,y}$ [GeV]	-1.22 ± 0.16	-0.93 ± 0.21	-1.07 ± 0.21	-0.89 ± 0.21
$p_{\nu,z}$ [GeV]	0.00 ± 0.18	-0.02 ± 0.24	0.19 ± 0.23	0.08 ± 0.24
$p_{\nu,\mathrm{T}}$ [GeV]	4.99 ± 0.15	4.97 ± 0.20	4.91 ± 0.19	4.89 ± 0.20
ϕ_ν [mrad]	-0.0002 ± 0.0027	0.0002 ± 0.0035	0.0014 ± 0.0035	-0.0021 ± 0.0036

Table 7.8: Comparison of the means of the Gaussian fits to all the ΔP distributions after the semileptonic top decay mode fit (upper half of the table) and after the hadronic top decay mode fit (lower half) in the pretag selection and in the tag selection at all three different working points. For the definition of ΔP, details about which events are used and how the Gaussian fit is done see Section 7.3.1.

Semileptonic top decay mode				
Width	Pretag	Tag at ϵ = 60%	Tag at ϵ = 70%	Tag at ϵ = 80%
E_b [GeV]	11.07 ± 0.18	11.58 ± 0.23	11.57 ± 0.23	11.67 ± 0.24
E_q [GeV]	10.67 ± 0.16	10.79 ± 0.23	10.71 ± 0.22	10.58 ± 0.18
$E_{q'}$ [GeV]	5.734 ± 0.086	5.66 ± 0.11	5.71 ± 0.11	5.52 ± 0.12
E_ℓ [GeV]	1.266 ± 0.023	1.375 ± 0.031	1.270 ± 0.027	1.265 ± 0.029
$p_{l,\mathrm{T}}$ [GeV]	1.126 ± 0.016	1.143 ± 0.021	1.143 ± 0.021	1.144 ± 0.021
$p_{\nu,x}$ [GeV]	12.92 ± 0.18	12.76 ± 0.22	12.77 ± 0.24	12.84 ± 0.25
$p_{\nu,y}$ [GeV]	13.09 ± 0.20	12.86 ± 0.24	12.94 ± 0.24	12.81 ± 0.24
$p_{\nu,z}$ [GeV]	26.84 ± 0.49	26.92 ± 0.65	30.41 ± 0.56	26.94 ± 0.61
$p_{\nu,\mathrm{T}}$ [GeV]	12.11 ± 0.14	10.89 ± 0.24	10.82 ± 0.23	10.82 ± 0.24
ϕ_ν [mrad]	0.2165 ± 0.0037	0.2184 ± 0.0047	0.2168 ± 0.0045	0.2199 ± 0.0051
Hadronic top decay mode				
Width	Pretag	Tag at ϵ = 60%	Tag at ϵ = 70%	Tag at ϵ = 80%
E_b [GeV]	6.03 ± 0.10	5.97 ± 0.13	6.00 ± 0.13	6.18 ± 0.12
E_q [GeV]	9.92 ± 0.16	9.90 ± 0.21	9.89 ± 0.20	9.92 ± 0.19
$E_{q'}$ [GeV]	5.930 ± 0.094	5.88 ± 0.11	5.86 ± 0.12	5.82 ± 0.12
E_ℓ [GeV]	1.576 ± 0.027	1.551 ± 0.035	1.575 ± 0.035	1.550 ± 0.034
$p_{l,\mathrm{T}}$ [GeV]	1.350 ± 0.021	1.323 ± 0.026	1.339 ± 0.026	1.357 ± 0.028
$p_{\nu,x}$ [GeV]	13.21 ± 0.19	13.27 ± 0.27	13.04 ± 0.23	13.10 ± 0.24
$p_{\nu,y}$ [GeV]	13.95 ± 0.20	13.23 ± 0.27	13.36 ± 0.27	13.26 ± 0.27
$p_{\nu,z}$ [GeV]	19.95 ± 0.39	20.02 ± 0.51	20.25 ± 0.51	20.40 ± 0.53
$p_{\nu,\mathrm{T}}$ [GeV]	12.27 ± 0.20	12.07 ± 0.25	12.08 ± 0.25	12.15 ± 0.25
ϕ_ν [mrad]	0.2089 ± 0.0034	0.2076 ± 0.0044	0.2110 ± 0.0042	0.2127 ± 0.0044

Table 7.9: Comparison of the widths of the Gaussian fits to all the ΔP distributions after the semileptonic top decay mode fit (upper half of the table) and after the hadronic top decay mode fit (lower half) in the pretag selection and in the tag selection at all three different working points. For the definition of ΔP, details about which events are used and how the Gaussian fit is done see Section 7.3.1.

7.3.5 Summary on improvements of fit parameters

In summary one can state that the KLFitter does fulfill the aim of improving the resolutions of the parameters and that in most regards the improvements behave as expected. It has been shown that the achieved resolutions do not depend on the selection. Furthermore, the disappearance of the long negative tail in the ΔP distribution for the b-quark energy can be considered as a particular success of the concept of the TFs.

7.4 Distributions of the basic KLFitter output variables

The first of the two aims of this section is to get familiar with the basic output variables of the KLFitter. Primarily, these are the likelihood and the event probability of a permutation. In practice, always their (natural) logarithm are shown. In addition, the *normalized event probability* is considered, which is the event probability of a permutation normalized to the sum of the event probabilities of all permutations in the event. In the following, these three variables together are referred to as the "likelihood variables".

The second aim of this section is to discuss the question how much these variables are suited for a hypothesis test, which is the last of the three main aims of kinematic fitting that has not been discussed yet in this chapter. Two aspects are focused on: firstly, the question whether true hadronic and semileptonic top decays can be distinguished from each other, secondly, whether the different sub-classes of the signal sample, (i.e. matched/unmatched events, (non-)good-fit events etc.) can be distinguished. The third obvious aspect, namely whether the KLFitter output is suited to discriminate Wt signal from background, is only discussed in Chapter 8.

7.4.1 Inclusive comparison of the three best permutations

In order to get familiar with the likelihood variables and how their distributions look like in general, the distributions of the likelihood variables for the three best permutations in the pretag selection on the full signal sample are shown in this subsection, first in the no-b-tagging mode and then in the working-point mode. For the latter, only the distributions for the 70% working point are shown exemplarily.

No-b-tagging mode

Figure 7.11 shows the distributions of the likelihood variables for the signal sample in the no-b-tagging mode. The event probability is omitted, as in the no-b-tagging mode it is identical to the likelihood. Instead, the normalized event probability is additionally shown on a logarithmic scale, as on the linear plot the shapes of distributions are difficult to compare. The first thing that stands out about the log(Likelihood) distribution is that is exhibits a peculiar multi-peak structure. The precise origin of that structure has not been investigated. However, as such a structure also shows up in the application of the KLFitter to $t\bar{t}$ production, it can be assumed that this behavior is somewhat inherent to the KLFitter. The next point that is noteworthy is that the shape of the log(Likelihood) distribution actually differs between the two top decay modes. Though overall peak structure is similar, the relative heights of the individual peaks and their precise shapes are visibly different (note that the x axis scales of the two plots differ). This is a first indication that deciding to which decay mode a given event belongs may be not straightforward. Finally, it is apparent that the distributions of the three permutations are not well-separated, but overlap substantially. This implies that the likelihood value of a permutation alone is not necessarily meaningful in order to decide how "good" a permutation is, e.g. the worst (i.e. third) permutation of a three-jet event may still have a higher likelihood than many of the best permutations.

This deficiency of the log(Likelihood) distribution is fixed by the normalized event probability. Here any value above 0.5 automatically denotes the best permutation of an event. Also the distribution of the normalized event probability shows a distinct structure. The distribution for the best permutation shows an extremely pronounced peak at a value of one, while the other two distributions have such a peak at zero. Given that the likelihood distribution ranges over several orders of magnitude, this is in fact not unexpected. The next feature that stands out is the edge in the distribution for the best permutation at 0.5, which is sort of a threshold effect, resulting from the fact that 0.5 is the upper limit for the second permutation. Furthermore it is noteworthy that the third permutation actually still has a

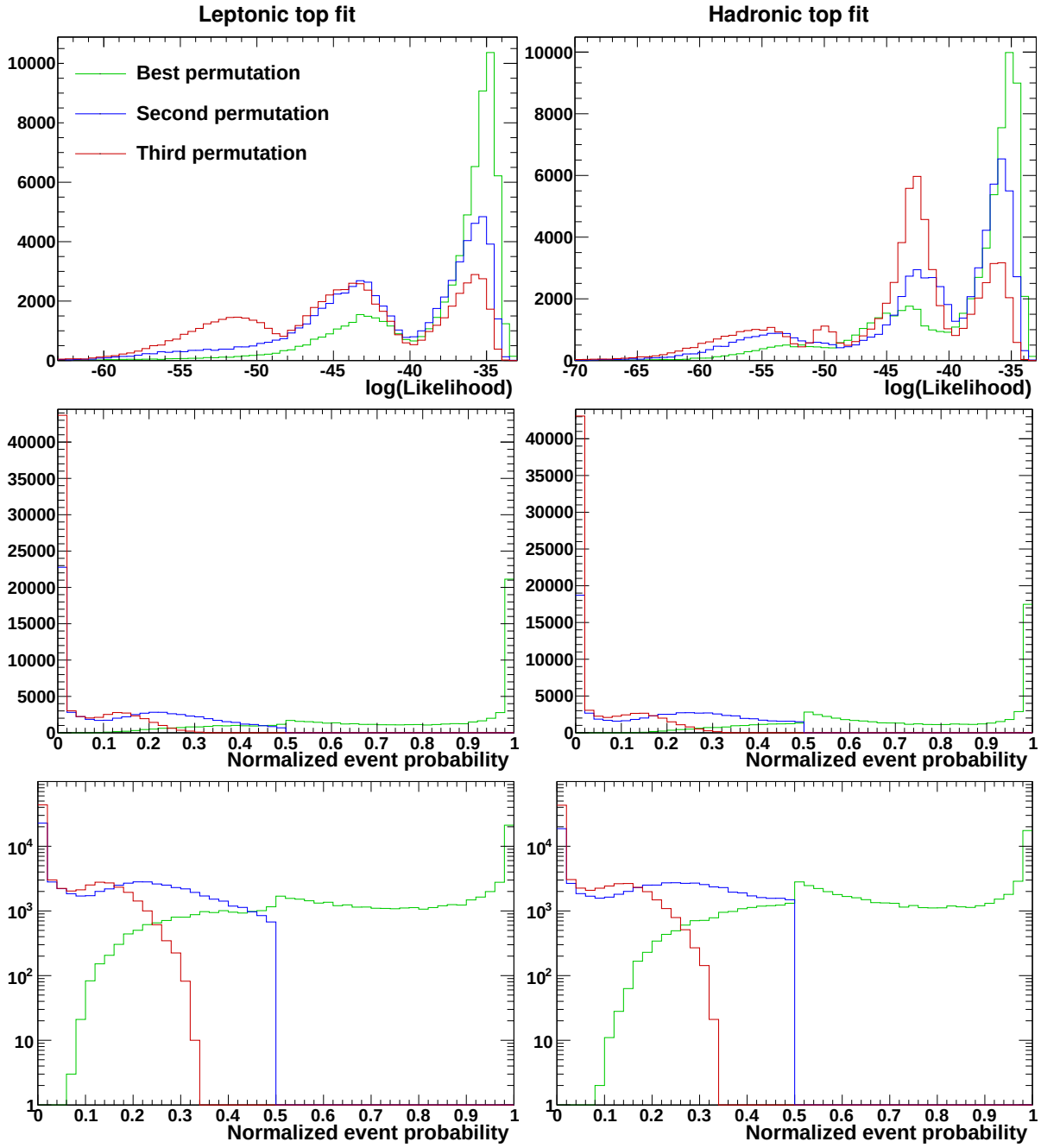


Figure 7.11: Distribution of the log(Likelihood) and the normalized event probability of the first three permutations in the no-*b*-tagging mode in the pretag selection on the signal sample for both fit hypotheses. The event probability is not shown separately as in the no-*b*-tagging mode it is identical to the likelihood.

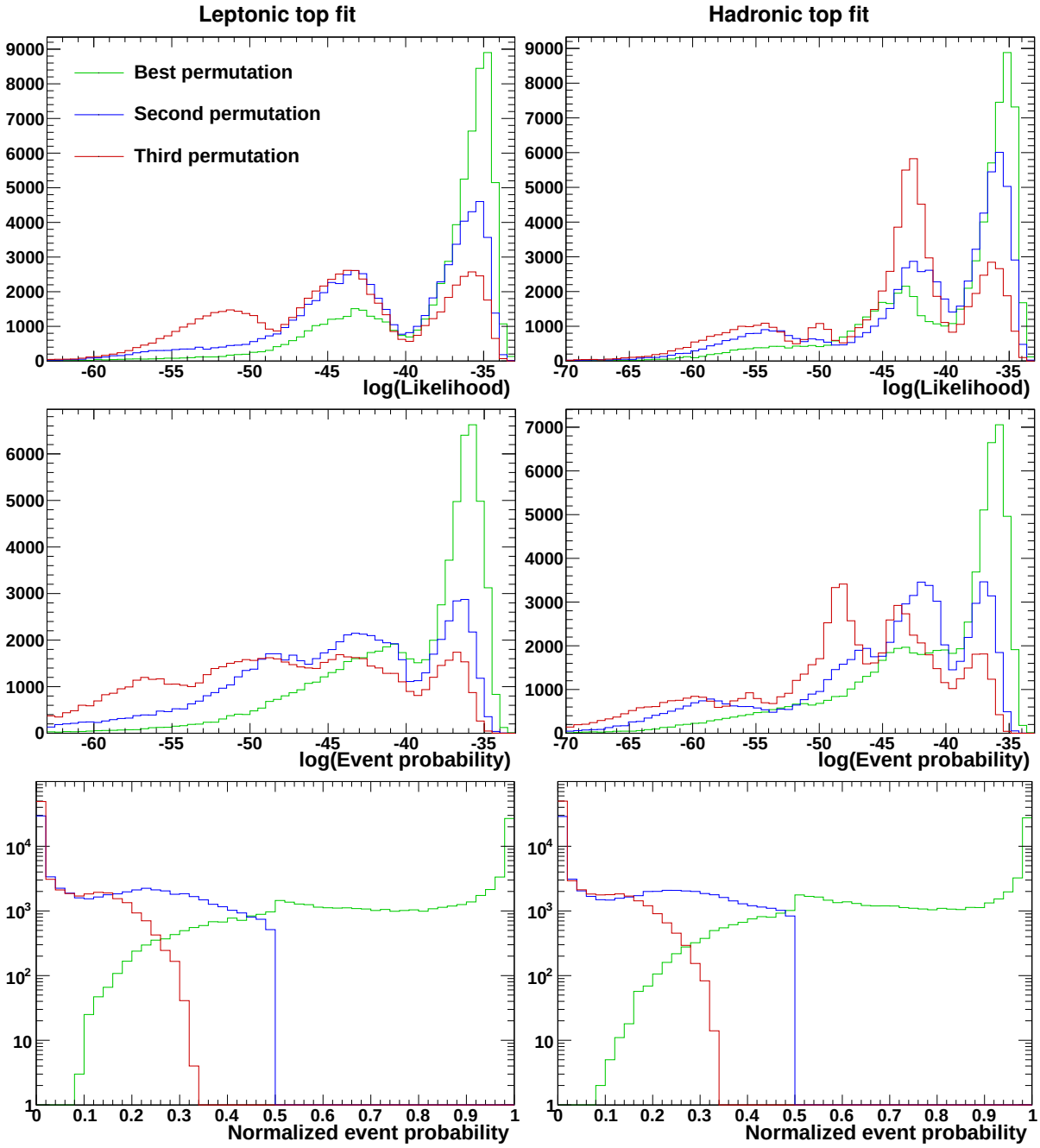


Figure 7.12: Distribution of the log(Likelihood), log(Event probability) and the normalized event probability of the first three permutations in the working-point mode in the pretag selection on the signal sample for both fit hypotheses. For the working-point mode the 70% working point was used.

significant contribution outside the peak at zero, which means that there are events for which all three best permutations have likelihoods that are at least of the same order of magnitude.

Working-point mode

In figure 7.12 again the distributions of the likelihood variables for the signal sample are shown, this time using the working point mode at 70% b-tagging efficiency. The log(Likelihood) looks very much like in the no-b-tagging mode, any differences besides the different y axis scale are hard to see. This is not too surprising, as the log(Likelihood) values themselves are still the same for each permutation, only the association of the permutations to the three distributions has changed to some extent.

The log(Event probability) distributions – unsurprisingly – look quite similar to the log(Likelihood) distributions, in particular for the best permutation. However, their peak structure, in particular for the second and third permutations, looks somewhat washed out compared to the latter. Also for these permutations the differences between the two top decay modes are more pronounced. Furthermore, the separation between the best and the two other permutations is a bit better in the log(Event probability) distribution, though the overlap is still substantial.

The normalized event probability distributions look mostly the same as for the no-b-tagging mode. The only apparent difference is that all the peaks are even more pronounced, now, and consequently the edge at 0.5 in the distribution for the best permutation is a bit smaller.

7.4.2 Comparison of the subsets of the signal sample for the best permutation

This subsection is devoted to the question whether the likelihood variables can be used in an easy way to distinguish between the two top decay modes and to find out whether the best permutation is also the correct one. In this context "easy" would typically mean that these aims can be achieved, at least to some extent, by cutting on a single variable. For this purpose the distributions of the likelihood variables for only the best permutation on different subsets of the signal sample are examined. Again first the no-b-tagging mode and then the working-point mode is evaluated and only the 70% working point is used exemplarily.

No-b-tagging mode

Figure 7.13 shows the distribution of the likelihood variables for the best permutation of each event in the no-b-tagging mode for different subsets of the signal sample in the pretag selection. As before, the event probability is omitted as it is identical to the likelihood. The full sample is split into all true lepton+jets events for which the used KLFitter top decay mode is the correct one (solid black line), all true lepton+jets events for which the used KLFitter top decay mode is the incorrect one (dashed black line), all true tau+jets events (as these are still considered to contain mostly signal; dashed orange line) and all other true decay modes (solid orange line). In addition, the true lepton+jets events for which the used KLFitter top decay mode is the correct one are again subdivided into events that could not be matched (non-matched events, blue line), events that are matched and that are good-fit events (green line), and events that are matched but for which the KLFitter did not pick the correct permutation (non-good-fit events, red line).

The most apparent observation is that again all distributions overlap substantially, regardless of whether they contain a correct permutation or not. Any easy separation of any of the shown subsets just by a simple cut on one of these distributions is therefore certainly not possible. Nevertheless some interesting features of the plots are discussed here. Firstly, it is noteworthy that true hadronic decays

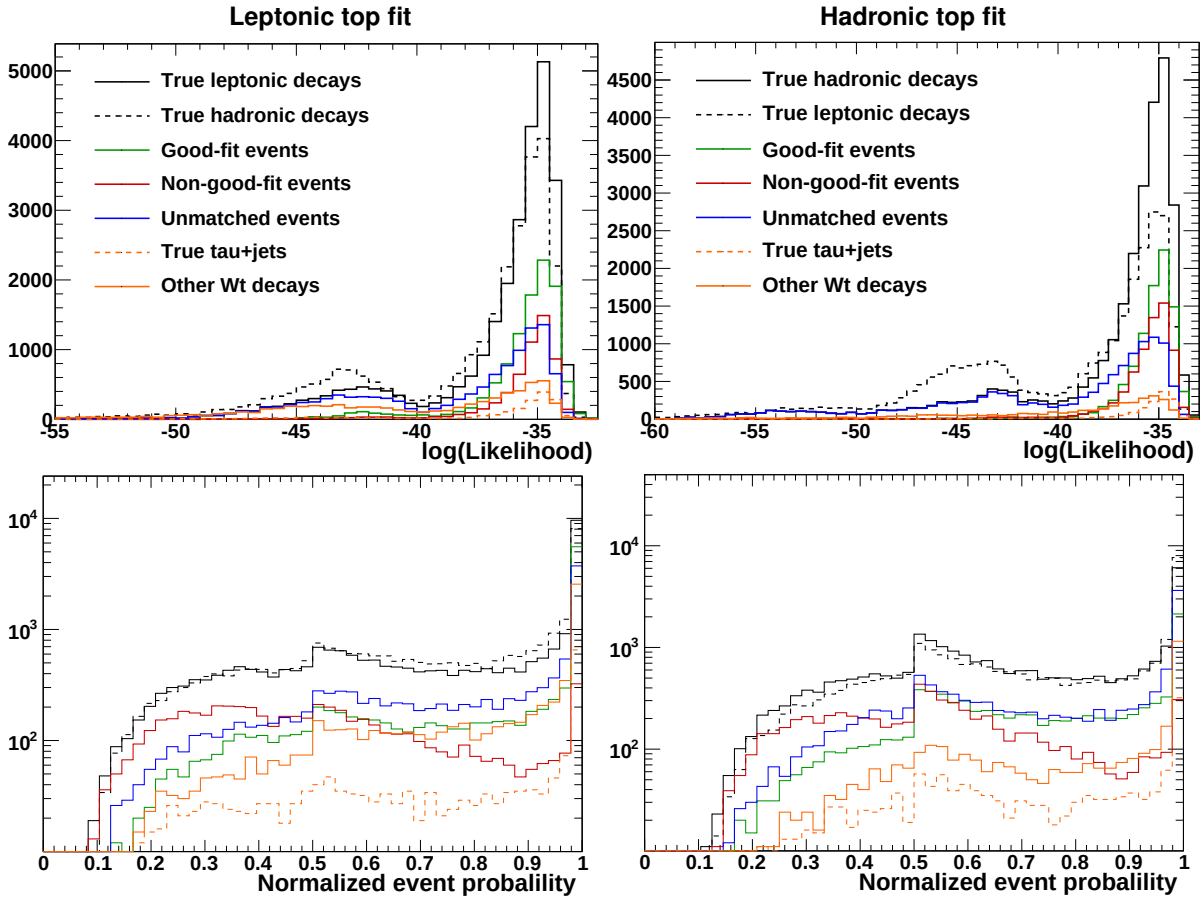


Figure 7.13: Distribution of the log(Likelihood) and the normalized event probability of the best permutation in the no-*b*-tagging mode for different subsets of the full signal sample (for detailed explanations see the text) in the pretag selection for both fit hypotheses. The event probability is not shown separately as in the no-*b*-tagging mode it is identical to the likelihood.

fitted with the semileptonic top-quark decay mode end up in the main peak of the likelihood distribution (i.e. at highest likelihoods) more frequently than vice versa. In contrast, for the "other" *Wt* decays it is just the other way around, while for the true tau+jets both distributions look roughly the same. Furthermore, in the semileptonic top decay mode, more good-fit events end up in the lower peak than in the hadronic top decay mode. For both top decay modes the "non-good-fit" events are located almost exclusively in the main peak. This is quite plausible, as a wrong permutation that is picked by the fitter needs to have a higher likelihood than the true permutation, which can be expected to be mostly in the main peak, given that almost all good-fit events are located there. Summing the good-fit events and the non-good-fit events, one can therefore state that events which contain a true permutation usually end up in the main peak. Consequently, a cut on the likelihood, keeping only the main peak would at least have a high efficiency to select these events, while the purity would still be quite low. These features of the likelihood distribution also show up in the normalized event probability distribution as differently pronounced peaks at a value of one and differently pronounced edges at 0.5, but this is not discussed in any more detail here.

Working-point mode

Figure 7.14 shows the distribution of the likelihood variables for the best permutation of each event for the same subsets of the signal sample as above in the working-point mode. In order to maximize the impact of the working-point mode this time the distributions in the tag selection are shown, using the same working point as for the working-point mode (i.e. 70%).

What can be seen is that the higher fit efficiency shows up naturally in a higher fraction of good-fit events. Regarding the shapes of the distributions the most obvious difference with respect to the no-*b*-tagging mode is that the discrimination between the two decay modes looks somewhat better, while it is still not good. Various other differences can be observed, which are all not discussed in detail, as none of them changes the big picture. The main conclusions are therefore still largely the same as in the no-*b*-tagging mode. Though some moderate improvement can be perceived, mainly regarding the discrimination between the two decay modes, any easy separation just by a cut on one of the distributions is still not possible.

7.4.3 Summary of likelihood variable distributions

In summary one can state that the likelihood variables that are provided by the KLFitter do have some discrimination power between different subsets of the signal sample as their distributions for the different subsets show a rich structure, while there is not one single variable which would be suited for a simple cut-based separation. However, it may be possible to exploit these variables by using a more advanced technique. This is done in the next chapter by combining the likelihood variables in neural networks, not only for the purpose of separating the *Wt*-channel from background, but also in order to distinguish some of the subsets of the signal sample.

This also finally answers the question from Section 6.2.3 on what to do about the two top decay modes in analysis. The answer is that no final decision can be taken regarding the top decay mode of an event. Rather a neural network is trained to distinguish between the two decay modes, and the output of this network then enters the final analysis along with the likelihood variables. For both purposes the variables from *both* top decay modes are used in order to fully exploit the available information from the KLFitter.

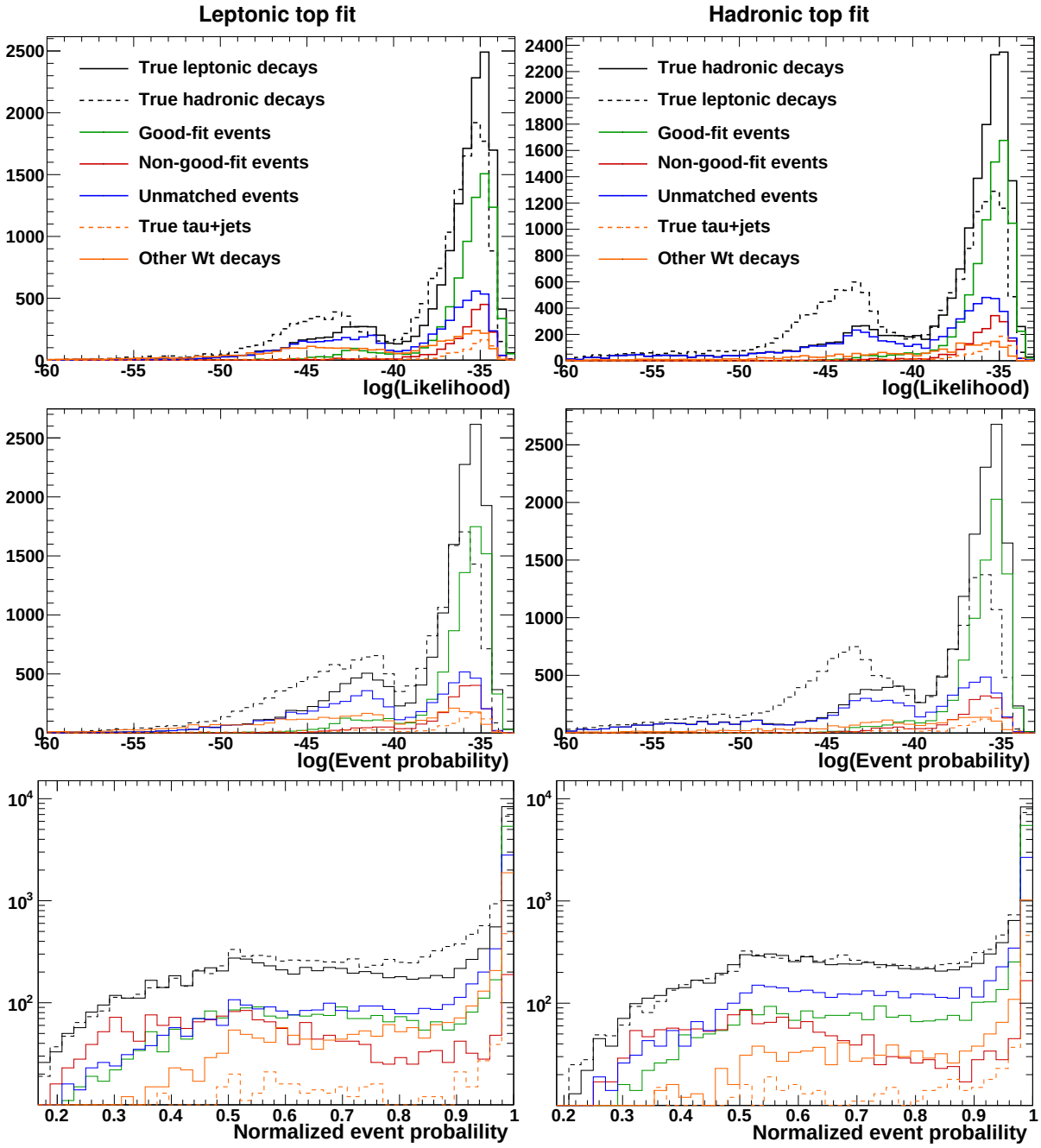


Figure 7.14: Distribution of the log(Likelihood), log(Event probability) and the normalized event probability of the best permutation in the working-point mode for different subsets of the full signal sample in the tag selection for both fit hypotheses. For both tag selection and the working-point mode the 70% working point was used.

Chapter 8

Use of the KLFitter output for analysis

In the previous chapter it has been shown that the implementation of the kinematic fit to the *Wt*-channel topology that was described in Chapter 6 was done properly and works as intended. The aim of this chapter is to demonstrate also the usefulness of the KLFitter for physics analysis. In the first section of this chapter, the tool that is used for the analysis, NeuroBayes, is introduced. In the second section, the variables that are used for analysis with NeuroBayes are described. In the third section, NeuroBayes is used to classify events within the signal sample. The fourth section finally deals with separating the *Wt* signal from background.

8.1 NeuroBayes

NeuroBayes (NB) [71] is a powerful, neural-network-based tool for multivariate analysis. Within this thesis it was used in its mode for binary classification. For each event in some input datasets, a target value – either -1 or 1, typically representing signal or background – is passed to NB along with the values of a set of input variables. NB then constructs a function that calculates one output variable from all the input variables, such that the value of the output variable is always as close to the target value as possible. In other words, NB constructs a single variable that combines the power of all the input variables to discriminate between two classes of events. The function that calculates the output variable consists of a neural network preceded by a preprocessor which prepares the input variables such that the neural network can make maximum use of them. Furthermore, the preprocessor is capable of choosing a suitable subset of input variables for the network out of a larger set. In the following the term "(neural) network" is used not only for the network itself but also for the complete function consisting of preprocessor and network together. The process of constructing the network such that it obtains its discrimination power is referred to as the *training* of the network. The datasets that are used for the training are called the training samples. During the training, individual weights can be assigned to each event, which can be used in particular for normalizing the relative fractions of signal and background events within the full training sample as desired. In this thesis the signal fraction was always adjusted to 50% this way. In the following, some relevant aspects of both the preprocessor and the actual neural network are detailed a bit further.

8.1.1 The preprocessor

During the training, the preprocessor does three things after the values of the input variables for each event have been passed to NB. First it applies a (non-linear) transformation to each individual input variable such that afterward its distribution has a Gaussian shape. Then, the variables are ranked by a sophisticated ranking procedure. Only those variables that are found to provide a significant contribution to the full separation power of the complete set of variables are passed to the neural network for the actual training process, the rest is discarded. Finally, the remaining input variables are decorrelated in

order to maximize their benefit for the neural network. In order to indicate how well the given input variables discriminate between signal and background the preprocessor provides two global figures of merit during the training: the "total correlation to target" of all variables, expressed in percent, and the "total significance" of this correlation, expressed in terms of standard deviations. In addition, for each variable four individual figures of merit are provided that arise from the ranking procedure. These are described along with the ranking itself in the following.

The ranking algorithm

The ranking is an iterative procedure. In each iteration it determines the variable that contributes least to the total significance among the remaining variables and then removes it from the list of variables (starting from the full set of input variables in the first iteration). In order to do so, it first calculates the total significance of all currently remaining variables. Then for each variable it derives the loss of significance (in terms of quadratic difference) that is caused by excluding only this variable from the calculation. The variable that causes the lowest significance loss is considered to contribute least to the total significance; this variable is ranked last among the remaining variables and removed from the list, in the following this is called the variable is "ranked out". This is repeated until no variables are left.

Figures of merit provided by the ranking

Besides the actual ranking, four quantities that characterize the importance of a variable are provided by the ranking procedure:

- *(Additional) significance*:
  The (additional) significance is the significance loss that a variable causes when it is ranked out. The quadratic sum of the additional significances of all variables equals the total significance of a set of variables. This also holds for any continuous subset of the N best-ranked variables.

- *Single significance*:
  The single significance is the total significance calculated from the variable alone. For the variable that is ranked best, this is identical to its (additional) significance.

- *Significance loss*:
  The significance loss when only this variable is excluded from the total set of variables. For the variable that is ranked last overall, this is identical to the (additional) significance.

- *Global correlation*:
  The total correlation of the variable to all others.

Although they are assigned to each individual variable, all these quantities (except for the single significance) are in fact to some extent collective properties which also depend on the other variables that are used in the preprocessing. This has some counterintuitive consequences. Most important to note is that none of the four quantities is precisely ordered along the ranking of the variables. The typical situation which contradicts intuition in this sense is when a variable with a very low (additional) significance is ranked very high, while e.g. the next lower ranked variable has a much higher (additional) significance. Here one has to keep in mind that – according to the definition of the ranking procedure – the higher-ranked variable would have caused an even higher significance loss if it had been ranked out instead of the lower-ranked variable. Apparently in such cases the higher-ranked variable only gained its high significance through the combination with other variables. As soon as these are gone, it loses

a large fraction of its discrimination power and gets ranked out next, with only a very low (additional) significance. Other typical counterintuitive situations are when a variable with a very low single significance has a very high (additional) significance or when a variable with a very high single significance is ranked out early with a very low (additional) significance. The former case again happens when a variable only gains its significance in combination with other variables, while the latter case occurs when the full separation power of a high-single-significance variable is already mostly contained in a set of other variables.

Choice of variables for network training

Large sets of input variables increase the risk of overtraining, i.e. the neural network may learn statistical fluctuations by heart in addition to learning the systematic features of the training samples. One of the protection mechanisms that NB offers against this is to restrict the set of variables that are passed to the neural network by the preprocessor to such whose contributions have a sufficient statistical significance. For this purpose, a cutoff on the (additional) significance is defined, above which a variable is considered to fulfill this criterion. For this thesis a value of 3σ was chosen. Starting from the variable that was ranked last overall, the preprocessor discards all variables in the order they were ranked out, until one variable has an (additional) significance above this cutoff value. This variable and all that were ranked higher are then passed to the network, even if some of them may have an (additional) significance below the threshold.

8.1.2 The neural network

Topology of the neural network

Neural networks are made up from nodes, also called neurons, and the connections between them. The connections are directed. From the input values that it obtains from its incoming connections, each node calculates an output value and sends it to its outgoing connections, whereat a weight is applied to each connection. The function that is used to calculate the output value is the same for each node, only the weights differ. The determination of the weights is the actual subject of the network training. In general, the nodes and the connections between them can be arranged quite arbitrarily. The topology of the network that is used by NB is rather fixed: it always is a two-layered feed-forward network. "Feed-forward" means that the nodes are arranged in consecutive layers, where each node gets input from all nodes in the previous layer and sends its output to all nodes in the next layer. "Two-layered" means that the network comprises one input layer, one intermediate layer, usually called the "hidden" layer, and one output layer. The input nodes (i.e. the nodes in the input layer) simply send the values of the input variables through their outgoing connections. Consequently, their number is given by the number of input variables, plus one bias-neuron. The output layer always consists of one single output node in the mode for binary classification. The number of hidden nodes (i.e. nodes in the hidden layer) is one of the parameters that can be defined by the user. It should be chosen as low as possible without limiting the learning capabilities of the neural network.[1]

Figures of merit and control plots provided by the neural network

NB provides a number of plots to assess the training process and the result of the training. Only the most interesting ones are outlined briefly here. Firstly, there are some control plots that indicate whether the

[1] Large numbers are not too critical, however. They may merely slow down the training and increase the risk of overtraining.

training process went smoothly. Secondly NB provides the distributions of the neural network output for signal and background in an overlay plot, which directly visualizes how good the separation between signal and background is. The neural network output is prepared by NB such that it ranges from -1 to 1 and has a linear relation to the signal purity (given a 50% overall signal fraction like in the training). A control plot that visualizes how good this linear relation is fulfilled is included in the NB output. Finally the signal efficiency and the overall efficiency that one obtains by cutting on the neural network output at any value are plotted against each other. The ratio of the area between the obtained efficiency curve and the unity line in this plot over the area below the unity line is called the *Gini index*, the plot is accordingly referred to as the *Gini plot* in the following. The Gini index is a main figure of merit for the overall separation power of the neural network and is also given in the NB output. The Gini index is bounded above by the signal fraction in the training samples, i.e. in this thesis the maximum possible value is always 50%. Such a value corresponds to perfect separation, i.e. by cutting on the neural network output first all the background gets cut away and only then also the signal is removed.

8.2 Used variables and their notation

For the training of all the neural networks in this chapter all three likelihood variables from the fits of all three permutations using both decay modes are used, amounting to 18 likelihood variables overall. Likelihood, event probability and normalized event probability are in the following written as $\mathcal{L}$, $\mathcal{L}_{\text{event}}$ and $\mathcal{L}_{\text{event, norm.}}$, respectively. The permutation ("1$^{\text{st}}$", "2$^{\text{nd}}$" or "3$^{\text{rd}}$") and fit hypothesis ("lep" or "had") to which they refer are written as arguments of the respective variable.

Furthermore, some kinematic variables of the fit particles from both fits of only the best permutation are used. These are firstly the transverse momenta of all model particles, including the two W bosons, the top quark and the sum of the top quark and the prompt W, in the following denoted as the Wt (system). From the W bosons, the top quark and the Wt, the transverse energy, the pseudorapidity, the transverse mass (calculated from their decay products), the azimuthal difference between their decay products, and the decay angle in their rest frame (i.e. the angle between the decay axis in their rest frame and their flight axis) are used as well. Finally, the energy and some momentum components of the prompt W and the top quark in the Wt rest frame as well as the mass of the Wt are used in addition. In total, 72 kinematic variables are used.

For all kinematic variables, the particle(s) as well as the fit hypothesis to which they refer are written as arguments of the respective variable. The two light quarks are written as q_1 and q_2, respectively, the charged lepton is denoted as l, the W boson that is produced in association with the top quark is written as W_{prompt}, the one from the top-quark decay is written as W_t. A particle in the rest frame of the Wt system has an (additional) subscript "Wt rest". The azimuthal difference $\Delta\phi$ has both particles from which it is calculated as arguments, the decay angle θ_{decay} and the transverse mass only have the corresponding mother particle as an argument.

8.3 Analysis within the signal sample

In this section the question is answered how well the KLFitter output is suited to distinguish between the different subsets of the signal sample. For this purpose, two neural networks are trained as described in Section 8.1. One network is trained to distinguish between the two top decay modes and one network to identify whether the KLFitter picked the correct permutation. For both networks the KLFitter is run in the working-point mode; only the three-jet bin in the tag selection is used, which is the same selection that is also considered for analysis. For both the working-point mode and the selection, only the 70%

working point is considered exemplarily. For the training of both networks, no event weights are used except for the normalization of the signal fraction to 50%.

8.3.1 Separation of semileptonic and hadronic top-quark decays

Training sample

The separation of semileptonic from hadronic top decays only makes sense when restricting at least to true lepton+jets decays. Given that the KLFitter can only be expected to achieve good separation when the true permutation is present, this should be restricted further to matched events only. In fact it turned out that the separation, estimated by the total significance provided by the preprocessor, is even better when using only good-fit events. Given the high efficiency of the KLFitter in the three-jet bin using the working-point, the restriction to only good-fit events leads to only a very low fraction of events that are lost with respect to all matched events. Therefore the good-fit events are chosen as the training sample. True hadronic top-quark decays are used as signal, true semileptonic top-quark decays are used as background. As input the 90 variables described in Section 8.2 are used.

Total correlation to target: 72.9%			Total significance: 72.3σ		
Rank	Variable	(Add.) sign.	Single sign.	Sign. Loss	Gl. Corr.
1	$\log(\mathcal{L}_{\text{event}}(1^{\text{st}}, \text{had}))$	56.9	56.9	11.5	89.3
2	$\log(\mathcal{L}_{\text{event}}(1^{\text{st}}, \text{lep}))$	38.3	44.2	11.6	83.2
3	$\mathcal{L}_{\text{event, norm.}}(1^{\text{st}}, \text{lep})$	8.7	48.6	10.7	88.7
4	$\log(\mathcal{L}(2^{\text{nd}}, \text{lep}))$	9.8	10.7	8.8	86.4
5	$\log(\mathcal{L}(2^{\text{nd}}, \text{had}))$	8.0	49.2	11.6	88.9
6	$\mathcal{L}_{\text{event, norm.}}(1^{\text{st}}, \text{had})$	9.4	8.0	9.7	81.3
7	$p_{\text{T}}(\nu, \text{lep})$	8.1	7.3	9.2	54.6
8	$\log(\mathcal{L}(3^{\text{rd}}, \text{had}))$	7.5	49.2	6.9	80.5
9	$p_{\text{T}}(W_{\text{prompt}}, \text{had})$	4.0	14.0	4.1	98.0
10	$\log(\mathcal{L}_{\text{event}}(3^{\text{rd}}, \text{lep}))$	1.0	22.0	4.0	98.5
11	$\log(\mathcal{L}(3^{\text{rd}}, \text{lep}))$	3.8	20.3	3.8	98.4
12	$p_{\text{T}}(t_{Wt\text{ rest}}, \text{had})$	3.5	12.8	3.9	98.1
13	$m(Wt, \text{had})$	2.3	2.4	4.3	97.8
14	$E(t_{Wt\text{ rest}}, \text{had})$	2.9	11.2	3.9	97.6
15	$m_{\text{T}}(t, \text{had})$	3.5	24.3	3.5	65.9

Table 8.1: Results from the preprocessing of the NB training to distinguish hadronic from semileptonic top-quark decays.

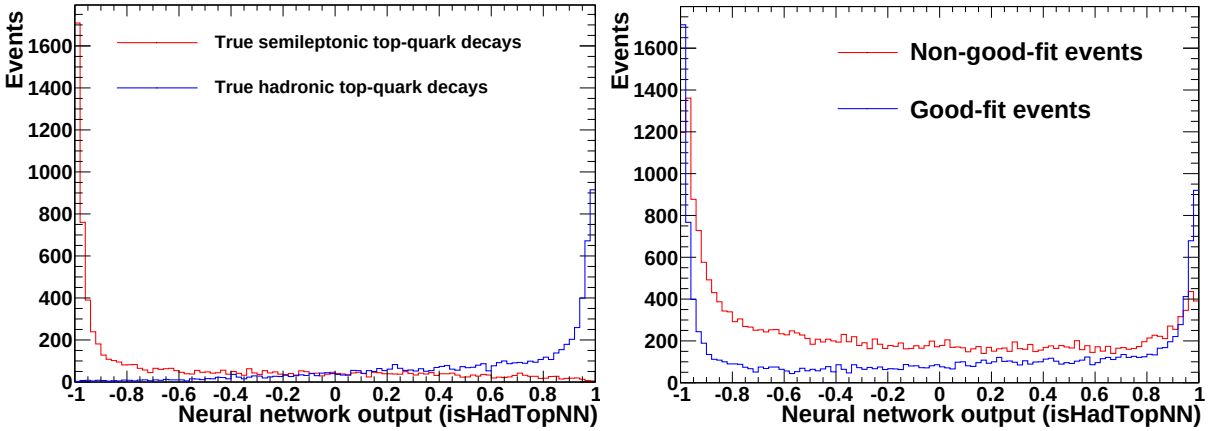
Preprocessing results

The preprocessing kept 15 variables for the network training. They are shown in table 8.1 along with the results from the preprocessing when repeated with these variables only (otherwise all figures of merit provided by the preprocessing would still refer to the full set of 90 variables). The table contains nice examples for some of the peculiar features of the ranking properties that were discussed in Section 8.1.1. Firstly, the two variables ranked 13th and 14th both have a lower (additional) significance than the variable ranked 15th (i.e. last). In particular that value is even below the cutoff value of 3σ for both of them. However, from their significance loss one can see that if they had been ranked last, they would have (additional) significances of 4.3σ and 3.9σ, respectively, which is above both the (additional) significance of the last variable and the cutoff value. Also there is a couple of variables with high single significances above 20σ, but much lower (additional) significances, most prominently the variable ranked 10th, which has a single significance of 22σ but gets ranked out with an (additional) significance of only 1σ. Another interesting aspect is that none of the variables has a significance loss of more than 12σ, not even the two variables that are ranked highest with very high additional significances way above 30σ. That is, part of their full individual separation power is taken away or hidden by the other variables, one may also say their separation power overlaps with that of some other variables. On the other hand some variables, e.g. the ones ranked 3rd and 10th, have a higher significance loss than (additional) significance, which means that they only *obtain* their full separation power due to the presence of other variables. One may also say that part of their separation power is taken away with the removal of other variables in some sense. Both aspects nicely illustrate that within a set of variables something like the individual contribution of one single variable to the full separation power is not uniquely defined. The full separation power is rather a collective quantity of the full set of variables. Therefore, a ranking such as that performed by the NB preprocessor, is always arbitrary to some extent. In particular, it is not guaranteed that it always provides the optimal combination of variables (in fact it doesn't always do so). Rather it must be regarded as a good, in particular fast and workable, approximation. This is something one should always keep in the back of one's mind when working with multivariate tools like NB.

Apart from these general aspects, the observed total correlation to target of 72.9% is an excellent result. Furthermore, the overall six highest-ranked variables and nine of the eleven highest-ranked variables are likelihood variables. Therefore, one can say that to a very large extent this high correlation is based on the likelihood variables. Even given the aspects about the collectivity of separation power that were just discussed, it can still be stated with certainty that the bulk of the observed total significance can also be achieved using only likelihood variables, e.g. the six best variables in the ranking already add up to 70.9σ compared to 72.3σ for the full set of 15 variables. This clearly indicates that the results from the kinematic fit are indeed suited to distinguish between semileptonic and hadronic top-quark decays. In order to obtain one single discriminating variable one only needs to combine the likelihood variables appropriately. Looking at the wide spectrum of likelihood variables that are kept for the network training, it furthermore confirms that keeping all likelihood variables from all three permutations of both fits was indeed a good idea.

Neural network result

The distribution of the output of the trained neural network for true semileptonic and hadronic top-quark decays using only good-fit events (i.e. the training sample) is shown in figure 8.1(a). It illustrates very clearly the excellent separation between semileptonic and hadronic top-quark decays that is achieved. Figure 8.2 shows the Gini plot for the network. A Gini index of 44% is achieved, which is quite close



(a) (b)

Figure 8.1: Output of the neural network that was trained to distinguish hadronic from semileptonic top-quark decays. Shown are the distributions for true semileptonic and hadronic decays using only good-fit events (i.e. the training sample) (a) as well as for good-fit and non-good-fit events using the full signal sample (b).

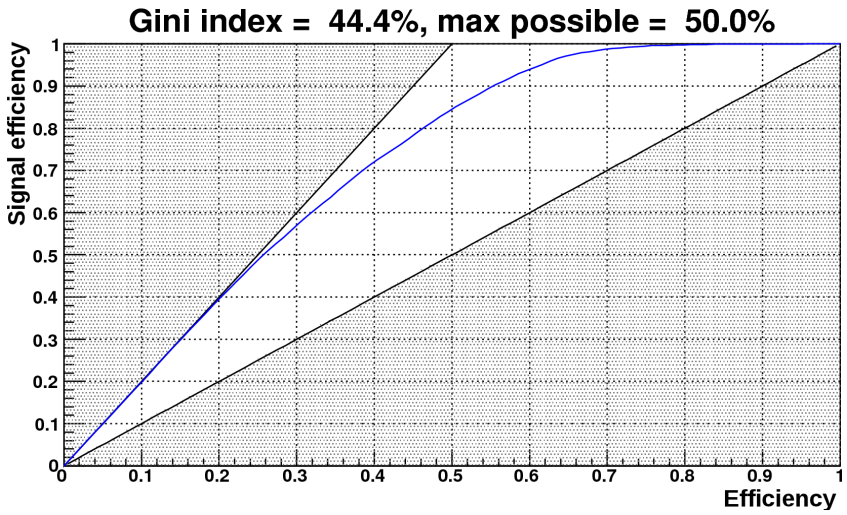


Figure 8.2: Gini plot for the neural network that was trained to distinguish hadronic from semileptonic top-quark decays.

to the maximum possible 50%, again confirming the excellent separation. One can conclude that the output variable of this neural network, in the following called "`isHadTopNN`", provides a very good test, for the actual top-quark decay hypothesis, in case one of them is actually fulfilled, i.e. in case of a matched true *Wt* event. Recalling that the `isHadTopNN` variable mostly relies on likelihood variables, this also implies that the KLFitter does fulfill the aim of kinematic fitting to provide a hypothesis test at least to this extent.

Figure 8.1(b) shows the `isHadTopNN` distribution for all good-fit events and for all non-good-fit events. The non-good-fit events show a similar peak structure as the good-fit events, but for them the peaks are less pronounced and instead the plateau between the peaks is higher. Taking the absolute value of `isHadTopNN` should therefore yield a variable that has some discrimination power between good-fit events and non-good-fit events, as this just puts the two peaks on top of each other to one side of the

distribution and the plateau to the other side. Consequently, both `isHadTopNN` and its absolute value were added to the input variables for the neural network that was trained to identify good-fit events.

8.3.2 Separation of good-fit events

For the training of the network to separate good-fit events from non-good-fit events, the `isHadTopNN` variable and its absolute value are used as input variables in addition to those that are described in Section 8.2. As the training sample this time the full signal MC sample is used. Good-fit events are used as signal, non-good-fit events are considered as background.

Preprocessing results

The preprocessing kept 25 variables for the network training. They are shown in table 8.2 along with the results from the preprocessing when repeated with these variables only. One more general feature of the preprocessing results should be mentioned, for which there was not such a pronounced example in table 8.1: variables that do not have any separation power by themselves (i.e. that have a very low single significance), but that develop a high significance when they are combined with others. Such examples are given here by the variables ranked 8th and 11th, which have single significances of 1σ and 0.4σ, but in the end provide (additional) significances of 9.5σ and 8.2σ, respectively.

The training yields a total correlation to target of 65.1%, which is not as high as in the previous case, but still a very good result. The fact that the total significance is nevertheless higher than for the decay modes, is simply because the training sample was larger this time. Both the `isHadTopNN` variable and its absolute value are among the top three variables. Both have a larger (additional) significance than single significance, in particular the plain `isHadTopNN`, where the effect is pronounced very strongly. This strongly supports the idea of reusing the output from one neural network as an input to another one. Furthermore, again the major contribution to the network comes from likelihood variables (the two `isHadTopNN` variables can also be considered to be mostly likelihood-based).

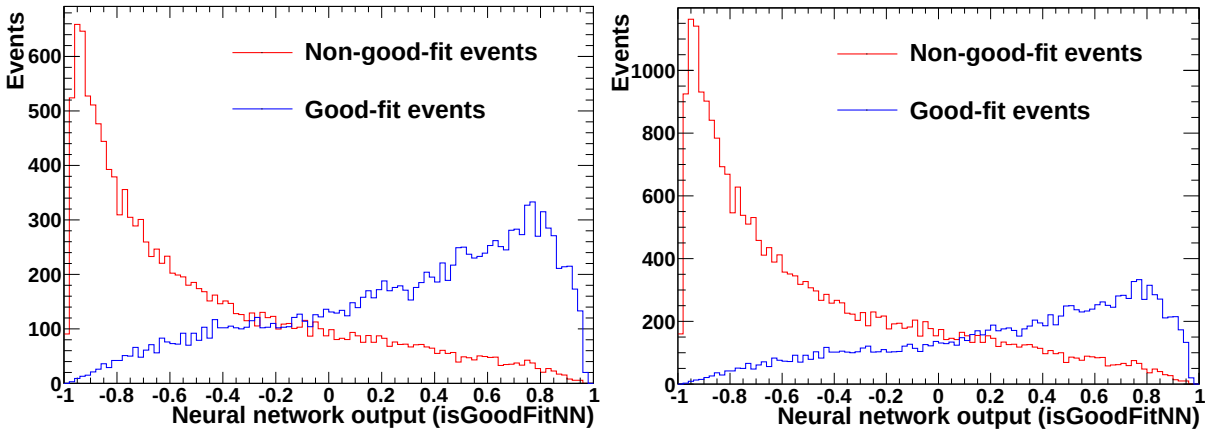
Neural network results

The distribution of the final neural network output for true good-fit events and non-good-fit events is shown in figure 8.3. Figure 8.3(a) shows the distributions with the background normalized to signal, as used for the training, figure 8.3(b) shows the same distributions with absolute normalization. One can see that indeed the separation is still good, even if not as good as for the decay modes. Figure 8.4 shows the Gini plot for the network, a Gini index of 39.5% is achieved, again confirming the good separation.

One can conclude that also in this regard the KLFitter does fulfill the aim of providing a hypothesis test. This hypothesis test again does not exactly test whether the fit hypothesis is fulfilled or not, as the network was trained to identify good-fit events rather than matched events, which are exactly those events that do fulfill the fit hypothesis (more precisely: one of the two hypotheses). But given the fit efficiency of more than 90% with the settings that are used, it is in fact quite close to this. Furthermore, one can conclude that the output variable from this network, which is called `isGoodFitNN`, should also provide some discrimination against backgrounds to the *Wt*-channel, which generally consist of non-good-fit events, too. Therefore the `isGoodFitNN` variable is also used for the training of the network to distinguish signal from background. The absolute value of `isGoodFitNN` is used as well, after this idea worked out so well with `isHadTopNN`.

Total correlation to target: 65.1%			Total significance: 97.8σ		
Rank	Variable	(Add.) sign.	Single sign.	Sign. Loss	Gl. Corr.
1	$\log(\mathcal{L}_{\text{event}}(1^{\text{st}}, \text{lep}))$	69.9	69.9	19.1	97.4
2	`isHadTopNN`	43.1	16.9	19.8	95.4
3	abs(`isHadTopNN`)	29.3	22.1	19.2	45.8
4	$\mathcal{L}_{\text{event, norm.}}(3^{\text{rd}}, \text{had})$	22.2	53.4	6.8	87.5
5	$p_{\text{T}}(q_2, \text{had})$	19.3	26.4	6.5	78.3
6	$p_{\text{T}}(Wt, \text{had})$	15.6	31.8	14.9	25.4
7	$\mathcal{L}_{\text{event, norm.}}(1^{\text{st}}, \text{lep})$	15.0	41.2	10.4	92.6
8	$\Delta\phi(l, \nu, lep)$	9.5	1.0	3.5	89.5
9	$p_{\text{T}}(\nu, \text{lep})$	9.3	3.2	3.9	93.6
10	$\log(\mathcal{L}(1^{\text{st}}, \text{lep}))$	9.7	60.2	7.5	93.3
11	$p_{\text{T}}(q_1, \text{lep})$	8.2	0.4	6.1	80.1
12	$p_{\text{T}}(b, \text{lep})$	7.1	9.7	7.3	58.1
13	$\log(\mathcal{L}_{\text{event}}(2^{\text{nd}}, \text{lep}))$	5.5	26.3	6.8	91.8
14	$\cos(\theta_{\text{decay}}(W_{\text{prompt}}, \text{had}))$	6.1	3.6	6.1	79.5
15	$p_{\text{T}}(q_2, \text{lep})$	5.6	16.1	5.9	79.7
16	$E(t_{Wt\text{ rest}}, \text{had})$	1.3	12.8	4.4	97.2
17	$m(Wt, \text{had})$	4.8	14.3	4.3	97.5
18	$\log(\mathcal{L}(1^{\text{st}}, \text{had}))$	4.4	61.8	4.2	92.7
19	$m_{\text{T}}(W_t, \text{lep})$	4.3	4.6	4.9	82.8
20	$\log(\mathcal{L}(3^{\text{rd}}, \text{had}))$	4.1	45.3	3.8	92.8
21	$\cos(\theta_{\text{decay}}(Wt, \text{lep}))$	3.8	0.7	3.8	30.5
22	$\mathcal{L}_{\text{event, norm.}}(2^{\text{nd}}, \text{had})$	3.5	48.3	3.5	77.0
23	$\Delta\phi(b, W_t, lep)$	2.6	8.5	4.3	75.3
24	$p_{\text{T}}(t, \text{lep})$	3.1	3.9	3.3	88.5
25	$p_{\text{T}}(\nu, \text{had})$	3.1	5.5	3.1	94.4

Table 8.2: Preprocessing results from the training to identify good-fit events.



(a) (b)

Figure 8.3: Output of the neural network that was trained to distinguish good-fit events from non-good-fit events. In (a) the distribution for background is normalized to signal like during the training, (b) shows the non-normalized distributions.

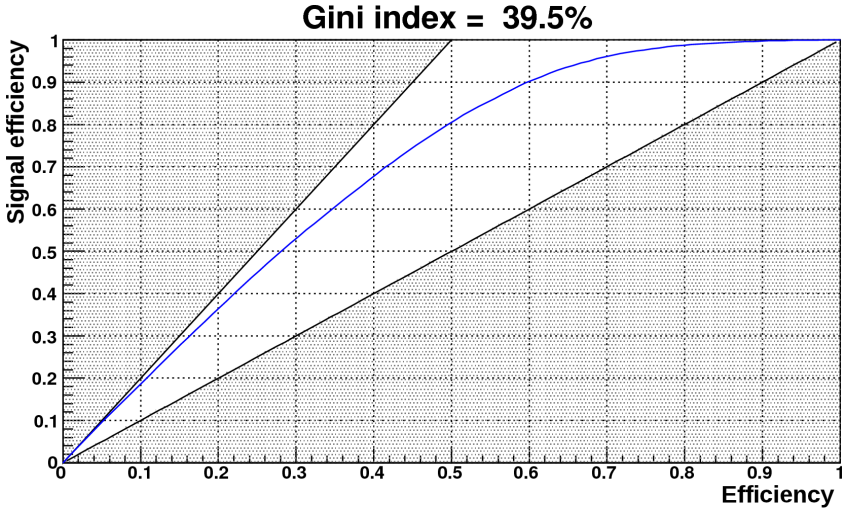


Figure 8.4: Gini plot for the neural network that was trained to distinguish good-fit events from non-good-fit events.

8.4 Separation of signal from background

This section is finally devoted to the separation of the *Wt*-channel signal from all the backgrounds. For this purpose again a neural network is trained. After outlining the training setup and describing the results from the training, these results are used to extract an estimate for the expected statistical significance of the signal.

8.4.1 NeuroBayes setup and training results

Training setup

For the training of the network all the MC samples are used as training samples. The *Wt*-channel MC (MC@NLO sample only) was used as signal and all the other MC samples are used as background. Multi-jets background was not used for the training in order not to let the neural network suffer from the large uncertainties that are connected to any multi-jets estimate. Again the KLFitter was run in the working-point mode and only the three-jet bin in the tag selection was used. Also, for both the working-point mode and the selection, again only the 70% working point was used exemplarily. Besides the normalization of the signal fraction to 50%, full event weights are applied for the training this time in order to reflect realistic analysis conditions. As input variables all the variables that are described in Section 8.2 are used as well as the `isHadTopNN` and the `isGoodFitNN` variable and the absolute values of both – another good example for the collective nature of the separation power and its sometimes counterintuitive consequences.

Preprocessing results

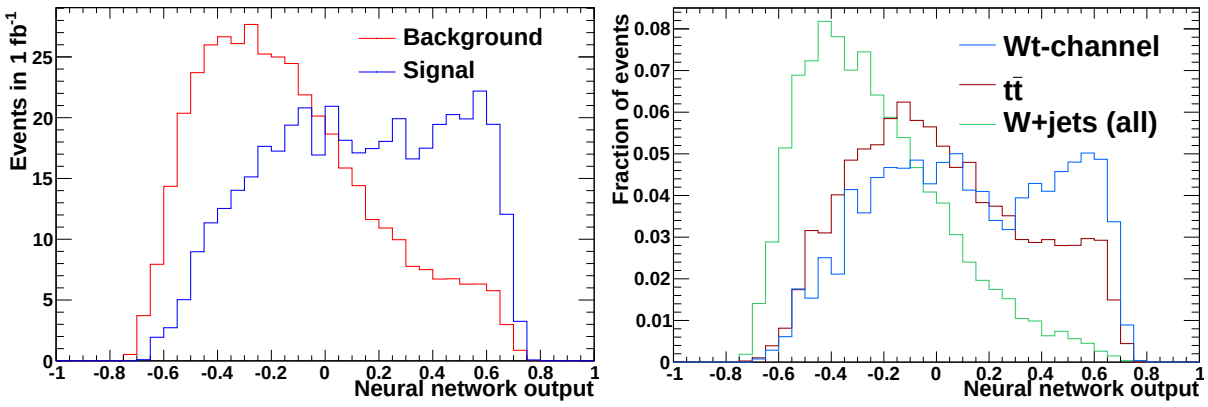
The preprocessing kept 22 variables for the network training. They are shown in table 8.3 along with the results from the preprocessing when repeated with these variables only. The total correlation to target is only 36.0%, which is substantially lower than for the two networks that were trained within the signal sample. Again, numerous likelihood variables are among the variables that were passed to the network. By adding their (additional) significances in quadrature, one finds that they still account for the major part of the total significance, even though, compared to the other two networks, the kinematic variables make up a stronger overall contribution this time. Out of the overall four neural-network-based variables, three are kept for the network training – the concept of using neural network output variables as input to other networks again turns out to be quite successful. The one of these variables that is not kept for the training is the (signed) `isGoodFitNN` variable, which was actually expected to provide some separation power. In fact, it does yield a very high single significance of 36.1σ, more than any of the variables that are taken for the network. During the training it drops out early with a very low rank and an (additional) significance of 0.8σ, however.

Neural network results

The distribution of the neural network output for signal and background is shown in figure 8.5(a), with the background normalized to signal as used for the training. One can see that the *Wt*-channel is indeed hard to separate from background – some separation is clearly visible, however. Figure 8.6 shows the corresponding Gini plot. A Gini index of 22.0% is achieved, also reflecting the lower separation power. This Gini index is of the same order of magnitude that is also achieved with non-KLFitter variables only, however.[72] This implies already that the KLFitter is indeed suitable for analysis and that it is at least competitive to a "normal" analysis. Furthermore, first attempts at combining both KLFitter and

Total correlation to target: 36.0%			Total significance: 45.3σ		
Rank	Variable	(Add.) sign.	Single sign.	Sign. Loss	Gl. Corr.
1	$\log(\mathcal{L}_{\text{event}}(1^{\text{st}}, \text{lep}))$	26.9	26.9	10.8	90.7
2	$E_{\text{T}}(W_{\text{prompt},Wt\text{ rest}}, \text{had})$	20.2	16.6	7.5	95.9
3	`isHadTopNN`	16.6	11.1	13.7	83.1
4	$\cos(\theta_{\text{decay}}(t, \text{lep}))$	11.4	0.7	4.0	64.7
5	$p_{\text{T}}(q_2, \text{had})$	10.2	13.4	5.9	51.7
6	$E_{\text{T}}(t, \text{lep})$	7.7	12.6	5.0	87.2
7	$m_{\text{T}}(W_{\text{prompt}}, \text{had})$	7.7	3.9	8.6	84.9
8	abs(`isHadTopNN`)	6.5	6.3	8.2	42.7
9	abs(`isGoodFitNN`)	6.4	11.4	9.0	60.5
10	$\log(\mathcal{L}(2^{\text{nd}}, \text{lep}))$	2.4	15.5	4.7	89.2
11	$\mathcal{L}_{\text{event, norm.}}(1^{\text{st}}, \text{lep})$	6.0	12.8	5.7	88.0
12	$\log(\mathcal{L}(3^{\text{rd}}, \text{lep}))$	3.1	14.1	5.3	96.8
13	$\log(\mathcal{L}_{\text{event}}(3^{\text{rd}}, \text{lep}))$	5.1	12.0	4.6	97.0
14	$p_{\text{T}}(b, \text{had})$	5.0	5.0	6.4	50.5
15	$p_{\text{T}}(q_1, \text{had})$	4.3	9.9	4.9	71.0
16	$\Delta\phi(l, \nu, \text{had})$	3.1	12.1	5.4	94.1
17	$p_{\text{T}}(t_{Wt\text{ rest}}, \text{had})$	4.3	14.8	4.9	97.3
18	$\log(\mathcal{L}_{\text{event}}(2^{\text{nd}}, \text{had}))$	3.6	23.4	3.7	85.2
19	$p_{\text{T}}(t, \text{lep})$	3.7	8.8	4.1	91.3
20	$\cos(\theta_{\text{decay}}(W_{\text{prompt}}, \text{lep}))$	3.3	5.9	3.6	37.3
21	$E_{\text{T}}(W_{\text{prompt}}, \text{lep})$	3.1	7.9	3.6	84.2
22	$\cos(\theta_{\text{decay}}(Wt, \text{had}))$	3.3	10.0	3.3	52.0

Table 8.3: Preprocessing results from the training to identify Wt signal.



(a) (b)

Figure 8.5: Output of the neural network that was trained to distinguish the *Wt* signal from background. In (a) the distribution for background is normalized to signal like during the training, (b) shows the area-normalized distributions for signal and only the two main backgrounds.

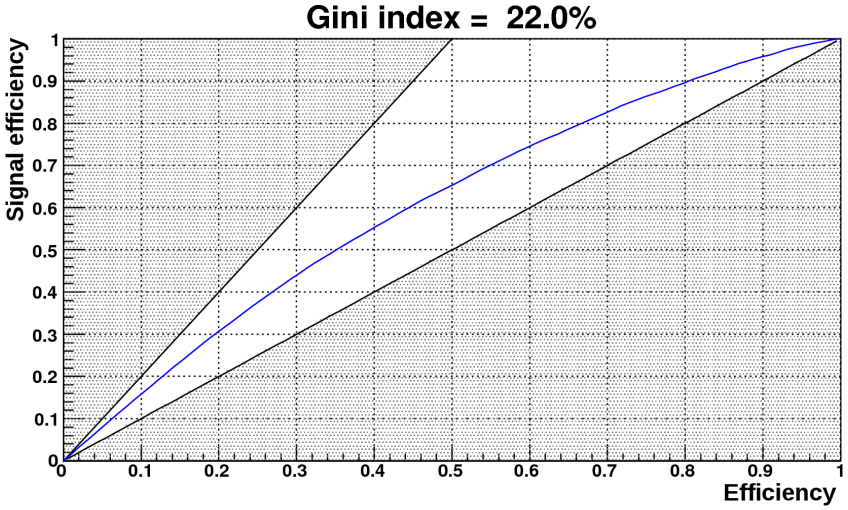


Figure 8.6: Gini plot for the neural network that was trained to distinguish the *Wt* signal from background.

non-KLFitter variables in one single network indicate that by doing so an additional increase of the Gini index by some percentage points can be achieved, e.g. an increase from about 19% to 22.5% for the 60% working point.[72] For the 70% working point this test has not been performed yet.

Figure 8.5(b) compares the area-normalized shapes of the neural network output distribution for signal and the two main backgrounds only. From this plot it becomes obvious that the reason for the bad separation is in fact the $t\bar{t}$ background, which has a similar shape as the signal, while the separation from the W+jets background looks rather good. To be fair, one must say that this biased by the fact that the largest background at the 70% working point is W+jets and therefore the network puts a larger focus on separating W+jets than $t\bar{t}$ during the training. Indeed, at the 60% working point, when $t\bar{t}$ is the largest background, or when a network is even trained against $t\bar{t}$ background only, the separation gets slightly better. Nevertheless, $t\bar{t}$ background always remains much harder to separate from signal than any other background, as it was anticipated right from the beginning, due to its final state being so similar to the *Wt*-channel.

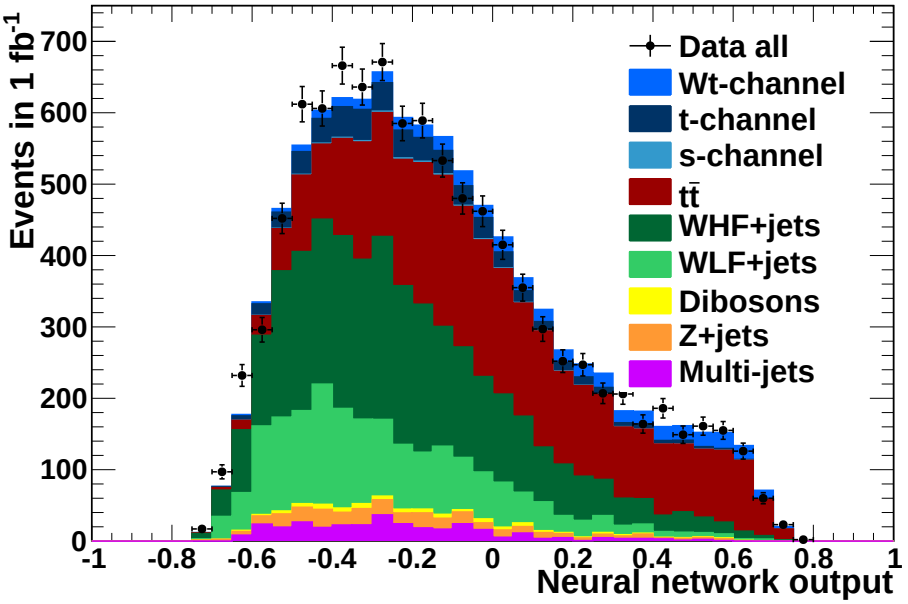


Figure 8.7: Stack plot of the neural network output distribution for signal and all backgrounds.

8.4.2 Extraction of a significance estimate

Figure 8.7 shows a stack plot of the neural network output distribution for signal and all backgrounds, including multi-jets, superimposed with real data. The multi-jets distribution concentrates at negative values, so not using it for the training turns out unproblematic in this regard. The estimation of the relative multi-jets fraction compared to data was taken over from the 60% working point, as no specific estimate for the 70% working point was available in time. Nevertheless, the agreement between data and MC is good. The largest discrepancies are at lowest network output values, where the data are underestimated. Instead, between about -0.2 and 0.1 the MC tends to overestimate the data. One might suspect from this that the true multi-jets contribution is even shifted a bit to the left with respect to the estimate, but this is speculation.

The signal contribution in figure 8.7 is well visible, in particular above a network output value of 0.1, the largest signal fraction can be anticipated by eye between about 0.3 and 0.6. Much more interesting is the question of the $S/\sqrt{B}$ ratio, which estimates the expected statistical significance. Figure 8.8 shows this ratio (including multi-jets) as a function of a cut on the neural network output keeping only the events above the cut value. One observes a maximum statistical significance of about 5.1 at cut values around zero. The $S/\sqrt{B}$ ratio without any cut is 4.6. This gain of 0.5 in significance, corresponding to a gain of 2.2 when doing the subtraction in quadrature or to an increase of statistics by a factor of 1.23, does not look very much at first glance. For a *Wt*-channel analysis this is in fact perfectly competitive result, however. Also, by the combination with non-KLFitter variables additional improvement is anticipated. Furthermore, given that at present the *Wt*-channel is still a discovery analysis, even such a seemingly small gain can be a decisive factor in order to reach 5σ in the end, as figure 8.8 nicely illustrates.

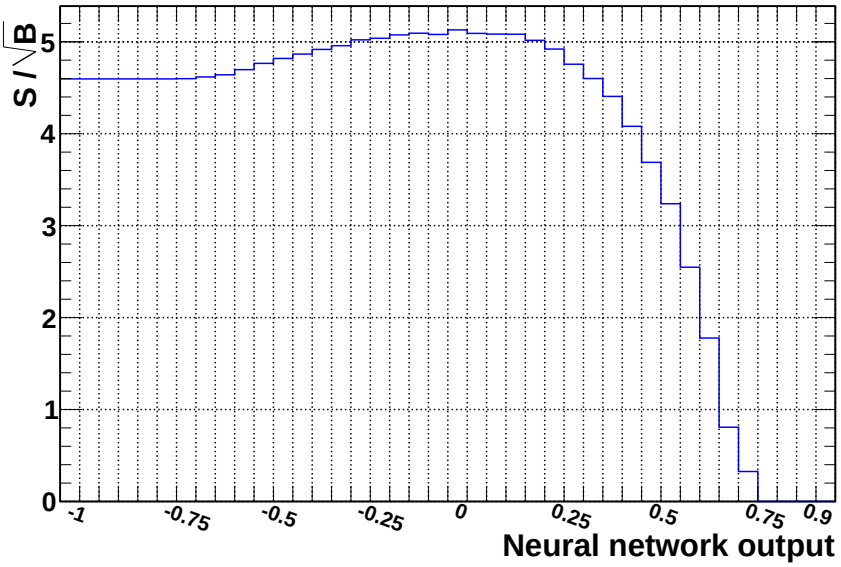


Figure 8.8: Expected statistical significance as a function of a cut on the neural network output.

8.4.3 Conclusions from the training of signal against background and prospects for the full analysis

Benefit of the KLFitter for analysis. It has been shown that a neural network trained with only KLFitter variables achieves a Gini index of 22%, and a cut on the neural network output can add 2.2σ to the expected statistical signal significance. This performance is achieved to a large extent by likelihood variables, which are the core output variables of the KLFitter and which provide the major contribution to the neural network. That includes the contribution from the outputs of the neural networks which were trained within the signal sample, as these are themselves mostly based on likelihood variables. Furthermore, the achieved performance is about the same as that of a comparable network without any KLFitter variables.[72] One can therefore conclude with certainty that the KLFitter in itself is indeed valuable for analysis: it appears to be fully competitive with non-KLFitter analyses; moreover, first studies clearly indicate that the combination of both will provide additional benefit.

Prospects for the full analysis. An expected statistical significance of 5.1σ for 1 fb^{-1} of 2011 data has been achieved by a cut-and-count approach on the neural network output. However, this significance estimate still disregards any systematic uncertainties. Taking these into account, a significance of in the order of 3σ is anticipated for the 1 fb^{-1} that was used in this thesis. The full analysis, which will of course have to include all the systematics, will instead be done on the full 2011 dataset of almost 5 fb^{-1}; and the final signal extraction will certainly employ more sophisticated statistical methods, making use of the full distribution of the network output. Furthermore it will profit from the combination of KLFitter and non-KLFitter variables. It is therefore expected that discovery will be achieved. But all this lies beyond the scope of this thesis.

Chapter 9

Summary

The work that has been presented in this thesis is part of the effort of the Bonn single top group to measure for the first time single top-quark Wt-channel production. The studies are being made in the lepton+jets decay mode in $\sqrt{s} = 7$ TeV proton–proton collision data taken with the ATLAS detector at the LHC. The general analysis strategy is to combine the discrimination power of different variables into one single variable by means of a neural network. The main objective of this thesis was to construct additional variables for the analysis that are directly sensitive to how much an event looks like Wt-channel signal, by employing a kinematic fit to the signal topology.

The implementation of the kinematic fit using the KLFitter package was presented. For this implementation, some difficulties that are inherent to the Wt-channel had to be overcome. Firstly, the semileptonic and the hadronic decay of the top quark present two distinct fit hypotheses that must be fitted independently. Furthermore, one of the kinematic constraints generally does not contribute any constraining power to overall fit, as it is used up entirely for the determination of the z momentum component of the neutrino that is part of the final state, which cannot be measured directly. As a consequence, in the hadronic top decay mode some of the fit parameters can not be improved by the fit, as no other constraint acts on them. It was shown that the implementation of the fit works as intended and that, despite these complications, it fulfills three important aims of kinematic fitting: it is capable of correctly assigning the measured objects to the final-state particles that are part of the fit hypothesis; it improves the values of the fit parameters with respect to their measured values; and it provides a test for the fit hypothesis.

The efficiency of the fit to correctly assign the measured objects to the model particles, with and without using b-tagging information in addition to the kinematic constraints, was studied as a function of the top-quark decay mode, the event selection and the number of jets. The observed efficiencies could be explained qualitatively by the internal structure of the fit. For some specific scenarios the efficiencies could also be probed quantitatively. The improvement of the fit parameters was studied depending on the top-quark decay mode. The different behavior of the individual fit parameters could be explained by considerations about the internal structure of the fit. For both studies, all results clearly indicate that the kinematic fit works as intended. In order to demonstrate that the kinematic fit can provide a test for the fit hypothesis and is thus indeed of value for the analysis, its output variables were combined in neural networks. Two networks were trained on simulated signal events. The first one was trained to distinguish semileptonic from hadronic top-quark decays; an excellent discrimination was achieved. The second one was trained to identify events where the kinematic fit succeeded to find the correct assignment of the measured objects. Again a very good discrimination was achieved. It is noteworthy that the output of the former network served as one of the best input variables to the latter. Finally a network was trained to distinguish Wt-channel signal from background, in the same way as for the full analysis; the outputs of both other neural networks were among the input variables. In general a good discrimination was achieved. The only exception is $t\bar{t}$ production, the second largest of the backgrounds, against which only moderate discrimination was achieved – but this is a general problem of Wt-channel

analyses, however.

Overall the separation from background that was achieved here using only output variables from the kinematic fit as input to the neural network is fully competitive to what is currently achieved in the full analysis of the Bonn group (based on the same neural-network tool) without any input from the kinematic fit. First studies of combining of both types of variables in one neural network indicate that this will indeed provide a significant benefit. Thus it can be stated that the kinematic fit is suitable for analysis as intended.

Besides the implementation of the kinematic fit and its validation, the full analysis setup as of late summer 2011 was presented, using data that were taken from March to June 2011 and which amount to an integrated luminosity of 1 fb^{-1}. For this setup the distribution of the neural network output on MC, including a data-driven estimate of multi-jets background, was compared to real data. Good agreement between data and MC is observed. An expected statistical significance of 5σ was extracted. This does not take into account any systematic uncertainties however. The full analysis of the Bonn group, which is ongoing, will be carried out with the full 2011 dataset of 5 fb^{-1}. It is expected that this amount of data will be sufficient to achieve a discovery.

Appendix A

Reconstruction of $p_{\nu,z}$ from the W mass constraint

While the x and y component of the momentum of a single neutrino in an event can be estimated by the missing transverse momentum, there is no such general way to estimate the z component of its momentum, $p_{\nu,z}$. However, in the situation that the neutrino comes from the leptonic decay of a W boson and the charged lepton that was produced in the same decay is known, as is the case for the lepton+jets mode of the Wt-channel, it is possible to use the W mass constraint to calculate an estimate.

Within this thesis, such an estimate was used as initial value for the parameter $p_{\nu,z}$ in the kinematic fit that is described in Chapter 6. The way how this estimate was calculated is detailed in this appendix.

A.1 The W mass as a function of $p_{\nu,z}$

Before deriving how to actually calculate $p_{\nu,z}$ from the W mass constraint, it is instructive to have a look at the lepton+neutrino invariant mass, $m_{\ell\nu}$, as a function of $p_{\nu,z}$ in general. For the sake of simplicity of the equations, we do actually not look at $m_{\ell\nu}$ itself, but at $\frac{1}{2}m_{\ell\nu}^2$. In the limit of zero lepton and neutrino masses[1] this reads

$$\begin{aligned}\tfrac{1}{2}m_{\ell\nu}^2 &= |\vec{p}_\ell| \cdot |\vec{p}_\nu| \cdot (1 - \cos(\sphericalangle(\vec{p}_\ell, \vec{p}_\nu))) \\ &= |\vec{p}_\ell| \cdot |\vec{p}_\nu| - \vec{p}_\ell \cdot \vec{p}_\nu \\ &= \sqrt{p_{\ell,\mathrm{T}}^2 + p_{\ell,z}^2} \cdot \sqrt{p_{\nu,\mathrm{T}}^2 + p_{\nu,z}^2} - p_{\ell,z} \cdot p_{\nu,z} - \vec{p}_{\ell,\mathrm{T}} \cdot \vec{p}_{\nu,\mathrm{T}}\,,\end{aligned}$$

where subscripts ℓ and ν label properties of the charged lepton and the neutrino, respectively, $\sphericalangle(\vec{p}_\ell, \vec{p}_\nu)$ denotes the angle between the lepton and the neutrino and $\vec{p}_{\ell,\mathrm{T}} \cdot \vec{p}_{\nu,\mathrm{T}} = p_{\ell,x} \cdot p_{\nu,x} + p_{\ell,y} \cdot p_{\nu,y}$.
From the cosine term in the first line of the equation it is clear that $m_{\ell\nu}$ becomes minimal when the neutrino is closest to the lepton. With $p_{\nu,z}$ as the only free parameter (i.e. with the lepton momentum, $p_{\nu,\mathrm{T}}$ and ϕ_ν fixed), one expects the minimum to be at that value of $p_{\nu,z}$ at which the lepton and the neutrino have equal polar angles. For a more quantitative analysis, we have a look at the derivatives. The first derivative reads

$$\frac{\mathrm{d}}{\mathrm{d}p_{\nu,z}}(\tfrac{1}{2}m_{\ell\nu}^2) = \frac{\sqrt{p_{\ell,\mathrm{T}}^2 + p_{\ell,z}^2}}{\sqrt{p_{\nu,\mathrm{T}}^2 + p_{\nu,z}^2}} \cdot p_{\nu,z} - p_{\ell,z}\,.$$

[1] For all practical purposes this approximation is perfectly valid for both electrons and muons.

Setting the derivative equal to zero and solving for $p_{\nu,z}$ yields

$$\frac{\mathrm{d}}{\mathrm{d}p_{\nu,z}}(\tfrac{1}{2}m_{\ell\nu}^2) = 0 \tag{A.1}$$

$$\Leftrightarrow \quad \sqrt{p_{\ell,\mathrm{T}}^2 + p_{\ell,z}^2}\cdot p_{\nu,z} = p_{\ell,z}\cdot\sqrt{p_{\nu,\mathrm{T}}^2 + p_{\nu,z}^2} \tag{A.2}$$

$$\Rightarrow p_{\ell,\mathrm{T}}^2\cdot p_{\nu,z}^2 + p_{\ell,z}^2\cdot p_{\nu,z}^2 = p_{\ell,z}^2\cdot p_{\nu,\mathrm{T}}^2 + p_{\ell,z}^2\cdot p_{\nu,z}^2 \tag{A.3}$$

$$\Leftrightarrow \quad p_{\nu,z} = \frac{p_{\nu,\mathrm{T}}}{p_{\ell,\mathrm{T}}}\cdot p_{\ell,z}\,. \tag{A.4}$$

The transition from (A.3) to (A.4) would actually also allow the opposite sign, but from (A.2) is it clear that $p_{\nu,z}$ and $p_{\ell,z}$ must have the same sign. This solution indeed implies $\theta_\nu = \theta_\ell$ as anticipated. It is particularly instructive to see how the value of $\frac{1}{2}m_{\ell\nu}^2$ that we obtain with this solution looks like:

$$\begin{aligned}
\tfrac{1}{2}m_{\ell\nu}^2\Big|_{p_{\nu,z}=\frac{p_{\nu,\mathrm{T}}}{p_{\ell,\mathrm{T}}}\cdot p_{\ell,z}} &= \sqrt{p_{\ell,\mathrm{T}}^2 + p_{\ell,z}^2}\cdot\sqrt{p_{\nu,\mathrm{T}}^2 + \frac{p_{\nu,\mathrm{T}}^2}{p_{\ell,\mathrm{T}}^2}\cdot p_{\ell,z}^2} - p_{\ell,z}\cdot\frac{p_{\nu,\mathrm{T}}}{p_{\ell,\mathrm{T}}}\cdot p_{\ell,z} - \vec{p}_{\ell,\mathrm{T}}\cdot\vec{p}_{\nu,\mathrm{T}} \\
&= p_{\ell,\mathrm{T}}\cdot\sqrt{1 + \frac{p_{\ell,z}^2}{p_{\ell,\mathrm{T}}^2}}\cdot p_{\nu,\mathrm{T}}\cdot\sqrt{1 + \frac{p_{\ell,z}^2}{p_{\ell,\mathrm{T}}^2}} - p_{\nu,\mathrm{T}}\cdot\frac{p_{\ell,z}^2}{p_{\ell,\mathrm{T}}} - \vec{p}_{\ell,\mathrm{T}}\cdot\vec{p}_{\nu,\mathrm{T}} \\
&= p_{\ell,\mathrm{T}}\cdot p_{\nu,\mathrm{T}}\cdot\left(1 + \frac{p_{\ell,z}^2}{p_{\ell,\mathrm{T}}^2} - \frac{p_{\ell,z}^2}{p_{\ell,\mathrm{T}}^2}\right) - \vec{p}_{\ell,\mathrm{T}}\cdot\vec{p}_{\nu,\mathrm{T}} \\
&= \tfrac{1}{2}m_{W,\mathrm{T}}^2\,.
\end{aligned}$$

In words: the minimal possible value for $m_{\ell\nu}$ as a function of $p_{\nu,z}$ is just the transverse mass! In fact, in the strict sense we only calculated that this is an extremal value so far. In order to clarify this, we also have a look at the second derivative:

$$\begin{aligned}
\frac{\mathrm{d}^2}{\mathrm{d}p_{\nu,z}^2}(\tfrac{1}{2}m_{\ell\nu}^2) &= \sqrt{p_{\ell,\mathrm{T}}^2 + p_{\ell,z}^2}\cdot\frac{\sqrt{p_{\nu,\mathrm{T}}^2 + p_{\nu,z}^2} - p_{\nu,z}\cdot\frac{p_{\nu,z}}{\sqrt{p_{\nu,\mathrm{T}}^2+p_{\nu,z}^2}}}{(p_{\nu,\mathrm{T}}^2 + p_{\nu,z}^2)} \\
&= \frac{\sqrt{p_{\ell,\mathrm{T}}^2 + p_{\ell,z}^2}\cdot p_{\nu,\mathrm{T}}^2}{\sqrt{p_{\nu,\mathrm{T}}^2 + p_{\nu,z}^2}^{\,3}} \;>\; 0\,.
\end{aligned}$$

We see that the second derivative is always positive, i.e. the transverse mass indeed is the minimum possible value of $m_{\ell\nu}$ as a function of $p_{\nu,z}$, and it is taken when the lepton and the neutrino have equal polar angles. Furthermore the function neither has any inflection points nor any additional extrema, but is perfectly "U"-shaped.

As a consequence for the derivation of $p_{\nu,z}$ from the W mass constraint we can expect that there will be no solution when $m_{W,\mathrm{T}}$ is larger than the W mass, exactly one solution for $m_{W,\mathrm{T}} = m_W$ and two solutions when $m_{W,\mathrm{T}}$ is smaller than m_W.

A.2 Derivation of $p_{\nu,z}$ from the W mass constraint

In order to derive how to calculate $p_{\nu,z}$ from the W mass constraint, we start by writing down the four-momentum conservation in the W decay (again neglecting the masses of both the neutrino and the charged lepton):

$$\begin{aligned} m_W^2 = p_W^2 &= (p_\ell + p_\nu)^2 \\ &= m_\ell^2 + m_\nu^2 + 2p_\ell \cdot p_\nu \\ &= 2 \cdot (|\vec{p}_\ell| \cdot |\vec{p}_\nu| - \vec{p}_\ell \cdot \vec{p}_\nu)\,. \end{aligned}$$

Now we have to bring the two summands on the right hand side to different sides of the equal sign so we can square them independently from each other. Then we split up any occurrence of $\vec{p}_\nu$ into $\vec{p}_{\nu,T}$ and $p_{\nu,z}$, put everything to one side and sort by powers of $p_{\nu,z}$:

$$\begin{aligned} & m_W^2 = 2 \cdot (|\vec{p}_\ell| \cdot |\vec{p}_\nu| - \vec{p}_\ell \cdot \vec{p}_\nu) \\ \Leftrightarrow \quad & |\vec{p}_\ell| \cdot |\vec{p}_\nu| = \tfrac{1}{2} m_W^2 + \vec{p}_\ell \cdot \vec{p}_\nu \\ \Leftrightarrow \quad & |\vec{p}_\ell|^2 \cdot |\vec{p}_\nu|^2 = (\tfrac{1}{2} m_W^2 + \vec{p}_{\ell,\mathrm{T}} \cdot \vec{p}_{\nu,\mathrm{T}} + p_{\ell,z} \cdot p_{\nu,z})^2 \\ \Leftrightarrow \quad & |\vec{p}_\ell|^2 \cdot (|\vec{p}_{\nu,\mathrm{T}}|^2 + p_{\nu,z}^2) = (\tfrac{1}{2} m_W^2 + \vec{p}_{\ell,\mathrm{T}} \cdot \vec{p}_{\nu,\mathrm{T}})^2 + (m_W^2 + 2 \cdot \vec{p}_{\ell,\mathrm{T}} \cdot \vec{p}_{\nu,\mathrm{T}}) \cdot p_{\ell,z} \cdot p_{\nu,z} + p_{\ell,z}^2 \cdot p_{\nu,z}^2 \\ \Leftrightarrow \quad & 0 = (|\vec{p}_\ell|^2 - p_{\ell,z}^2) \cdot p_{\nu,z}^2 - (m_W^2 + 2 \cdot \vec{p}_{\ell,\mathrm{T}} \cdot \vec{p}_{\nu,\mathrm{T}}) \cdot p_{\ell,z} \cdot p_{\nu,z} \\ & \qquad - (\tfrac{1}{2} m_W^2 + \vec{p}_{\ell,\mathrm{T}} \cdot \vec{p}_{\nu,\mathrm{T}})^2 + |\vec{p}_\ell|^2 \cdot |\vec{p}_{\nu,\mathrm{T}}|^2 \\ \Leftrightarrow \quad & p_{\nu,z}^2 - 2 \cdot \alpha \cdot \frac{p_{\ell,z}}{p_{\ell,\mathrm{T}}^2} \cdot p_{\nu,z} + \frac{|\vec{p}_\ell|^2 \cdot |\vec{p}_{\nu,\mathrm{T}}|^2 - \alpha^2}{p_{\ell,\mathrm{T}}^2} = 0\,, \end{aligned}$$

where $\alpha = \frac{1}{2} m_W^2 + \vec{p}_{\ell,\mathrm{T}} \cdot \vec{p}_{\nu,\mathrm{T}}$.
This is a quadratic equation for $p_{\nu,z}$. The solution reads

$$p_{\nu,z} = \frac{\alpha \cdot p_{\ell,z}}{p_{\ell,\mathrm{T}}^2} \pm \frac{\sqrt{p_\ell^2 \cdot (\alpha^2 - p_{\nu,\mathrm{T}}^2 \cdot p_{\ell,\mathrm{T}}^2)}}{p_{\ell,\mathrm{T}}^2}\,. \tag{A.5}$$

It can be shown that the sign of the discriminant (and hence the number of solutions) is indeed directly related to the size of the transverse mass with respect to the W mass, as expected from the considerations in the previous section. First, we can translate a positive discriminant into a condition for α:

$$\begin{aligned} & p_\ell^2 \cdot (\alpha^2 - p_{\nu,\mathrm{T}}^2 \cdot p_{\ell,\mathrm{T}}^2) \geq 0 \\ \Leftrightarrow \quad & \alpha^2 \geq p_{\nu,\mathrm{T}}^2 \cdot p_{\ell,\mathrm{T}}^2 \\ \Leftrightarrow \quad & |\alpha| \geq p_{\nu,\mathrm{T}} \cdot p_{\ell,\mathrm{T}}\,. \end{aligned}$$

If $\alpha \geq 0$, we have $|\alpha| = \alpha$ and can go on straightforwardly:

$$\begin{aligned} \alpha \;=\; & \tfrac{1}{2} m_W^2 + \vec{p}_{\ell,\mathrm{T}} \cdot \vec{p}_{\nu,\mathrm{T}} \geq p_{\nu,\mathrm{T}} \cdot p_{\ell,\mathrm{T}} \\ \Leftrightarrow \quad & m_W^2 \geq 2 \cdot (p_{\nu,\mathrm{T}} \cdot p_{\ell,\mathrm{T}} - \vec{p}_{\ell,\mathrm{T}} \cdot \vec{p}_{\nu,\mathrm{T}}) \;=\; m_{W,\mathrm{T}}\,. \end{aligned}$$

This means that for positive α the relation between m_W and $m_{W,\mathrm{T}}$ is indeed connected to the sign of the

discriminant in the expected way. For $\alpha < 0$, i. e. $|\alpha| = -\alpha$, we get

$$\begin{aligned} -\alpha &= -\tfrac{1}{2}m_W^2 - \vec{p}_{\ell,\mathrm{T}} \cdot \vec{p}_{\nu,\mathrm{T}} \geq p_{\nu,\mathrm{T}} \cdot p_{\ell,\mathrm{T}} \\ \Leftrightarrow \qquad & m_W^2 \leq -2 \cdot (p_{\nu,\mathrm{T}} \cdot p_{\ell,\mathrm{T}} - \vec{p}_{\ell,\mathrm{T}} \cdot \vec{p}_{\nu,\mathrm{T}}) \,. \end{aligned}$$

As the right hand side is always negative this is never true, i.e. for negative α the discriminant can never be greater than or equal to zero and thus there is never a solution for $p_{\nu,z}$. What's still missing is the relation to $m_{W,\mathrm{T}}$ for this case. This can be obtained by

$$\begin{aligned} & \alpha = \tfrac{1}{2}m_W^2 + \vec{p}_{\ell,\mathrm{T}} \cdot \vec{p}_{\nu,\mathrm{T}} \; < \; 0 \\ \Leftrightarrow \; & m_W^2 < -2 \cdot \vec{p}_{\ell,\mathrm{T}} \cdot \vec{p}_{\nu,\mathrm{T}} \\ & \quad < -2 \cdot \vec{p}_{\ell,\mathrm{T}} \cdot \vec{p}_{\nu,\mathrm{T}} + 2 \cdot p_{\nu,\mathrm{T}} \cdot p_{\ell,\mathrm{T}} \\ & \quad = m_{W,\mathrm{T}}^2 \,. \end{aligned}$$

Having shown this, we now have confirmed that, as anticipated at the end of the previous section, there are two solutions for $p_{\nu,z}$ when $m_{W,\mathrm{T}} < m_W$, there is no solution when $m_{W,\mathrm{T}} > m_W$ and there is exactly one solution when $m_{W,\mathrm{T}} = m_W$. In the following it is discussed how to obtain a unique estimate for $p_{\nu,z}$ in each case.

A.2.1 One solution

The one-solution case is the only one for which a unique estimate for $p_{\nu,z}$ is trivial and straightforward. However, in practice this case basically never occurs. It is discussed here anyway, as some of the approaches to deal with the no-solution case will fall back on the one-solution case.

In fact, there is not just one, but there are rather two straightforward ways to write down the $p_{\nu,z}$ estimate in the one-solution case. Either one just uses equation (A.5) without the square root term or one makes use of the insight from Section A.1 that the one-solution case is equivalent to $m_{W,\mathrm{T}} = m_W$ and that $m_{\ell\nu}$ becomes equal to $m_{W,\mathrm{T}}$ (and hence to m_W in this case) when $\theta_\nu = \theta_\ell$. Of course, both approaches must be equivalent, as is shown explicitly in the following.

We start by using equation (A.5):

$$p_{\nu,z} \;=\; \frac{\alpha \cdot p_{\ell,z}}{p_{\ell,\mathrm{T}}^2} \;=\; \frac{(\frac{1}{2}m_W^2 + \vec{p}_{\ell,\mathrm{T}} \cdot \vec{p}_{\nu,\mathrm{T}}) \cdot p_{\ell,z}}{p_{\ell,\mathrm{T}}^2} \,. \tag{A.6}$$

Then we need the condition that the discriminant vanishes:

$$\begin{aligned} & \alpha^2 - p_{\nu,\mathrm{T}}^2 \cdot p_{\ell,\mathrm{T}}^2 = 0 \\ \Leftrightarrow \qquad & \alpha = p_{\nu,\mathrm{T}} \cdot p_{\ell,\mathrm{T}} \,. \end{aligned}$$

The negative sign for α can be excluded, as in Section A.2 it was shown that this would imply the no-solution case. By inserting this in (A.6) we obtain

$$p_{\nu,z} \;=\; \frac{p_{\nu,\mathrm{T}} \cdot p_{\ell,\mathrm{T}} \cdot p_{\ell,z}}{p_{\ell,\mathrm{T}}^2} \;=\; \frac{p_{\nu,\mathrm{T}}}{p_{\ell,\mathrm{T}}} \cdot p_{\ell,z}$$

in accordance with the results from Section A.1.

A.2.2 No solution

The case of having no exact solution at all, which seems to be the most difficult one at first glance, turns out to be actually more convenient than the two-solution case in the end, as there is no ambiguity to resolve here. A few approaches to construct an estimate for $p_{\nu,z}$ are presented in the following. One criterion that each of these solutions should fulfill, is that in the limit of $m_{W,\mathrm{T}} \to m_W$ they should yield the same $p_{\nu,z}$ as the one-solution case.

The minimum mass approach

The no-solution case occurs when the lowest possible value for $m_{\ell\nu}$ as a function of $p_{\nu,z}$, namely the transverse mass, is higher than m_W. It is therefore obvious that choosing $p_{\nu,z}$ such that $m_{\ell\nu}$ gets minimal, should be a good idea. As shown, this is the case when the lepton and the neutrino have equal polar angles, i.e. for

$$p_{\nu,z} = \frac{p_{\nu,\mathrm{T}}}{p_{\ell,\mathrm{T}}} \cdot p_{\ell,z}\,,$$

yielding $m_{\ell\nu} = m_{W,\mathrm{T}}$.

As this equation for $p_{\nu,z}$ is the same as for the one-solution case, it will obviously also give the same result for $m_{W,\mathrm{T}} \to m_W$.

$\vec{p}_{\nu,\mathrm{T}}$ adjustment approaches

This group of approaches is based on the idea that the no-solution case can only occur due to a mismeasurement, as the transverse mass calculated from the true lepton and neutrino (transverse) momenta is bounded above by the W mass. As the lepton is measured quite precisely, one can assume that the mismeasured quantity is $\vec{p}_{\nu,\mathrm{T}}$. To be more precise, this "mismeasurement" is rather due to the fact that the missing transverse momentum only provides a rather rough estimate for $\vec{p}_{\nu,\mathrm{T}}$.

The idea is therefore to reestimate $\vec{p}_{\nu,\mathrm{T}}$ such that $m_{W,\mathrm{T}} = m_W$. With this reestimated value for $\vec{p}_{\nu,\mathrm{T}}$, there is then exactly one solution for $p_{\nu,z}$, which is used as the estimate. Furthermore, in the limit of $m_{W,\mathrm{T}} \to m_W$ the continuous transition to the one-solution case is obvious.

The minimum deviation approach. The most rigorous approach that follows this idea is to reestimate both $p_{\nu,x}$ and $p_{\nu,y}$ such that $m_{W,\mathrm{T}} = m_W$ and at the same time the absolute difference between the measured and the reestimated transverse momentum, given by $|\Delta\vec{p}_{\nu,\mathrm{T}}| = \sqrt{(\vec{p}_{\nu,\mathrm{T}} - \vec{p}\,'_{\nu,\mathrm{T}})^2}$, is minimized. However, this minimization cannot be performed analytically, but requires a numerical solution, i.e. a fit needs to be performed. This extra effort has not been undertaken in the scope of this thesis.

The $p_{\nu,\mathrm{T}}$ adjustment approach. In a simplified approach, one may keep ϕ_ν constant and only adjust $p_{\nu,\mathrm{T}}$ such that $m_{W,\mathrm{T}} = m_W$:

$$\begin{aligned}
\tfrac{1}{2}m_{W,\mathrm{T}}^2 &= p_{\ell,\mathrm{T}} \cdot p_{\nu,\mathrm{T}} - \vec{p}_{\ell,\mathrm{T}} \cdot \vec{p}_{\nu,\mathrm{T}} \\
&= p_{\ell,\mathrm{T}} \cdot p_{\nu,\mathrm{T}} \cdot (1 - \cos(\Delta\phi(\vec{p}_{\ell,\mathrm{T}}, \vec{p}_{\nu,\mathrm{T}}))) \\
\tfrac{1}{2}(m'_{W,\mathrm{T}})^2 &= p_{\ell,\mathrm{T}} \cdot p'_{\nu,\mathrm{T}} - \vec{p}_{\ell,\mathrm{T}} \cdot \vec{p}\,'_{\nu,\mathrm{T}} \\
&= p_{\ell,\mathrm{T}} \cdot p'_{\nu,\mathrm{T}} \cdot (1 - \cos(\Delta\phi(\vec{p}_{\ell,\mathrm{T}}, \vec{p}_{\nu,\mathrm{T}}))) \\
&= \tfrac{1}{2}m_W^2 \\
\Rightarrow \qquad \frac{p'_{\nu,\mathrm{T}}}{p_{\nu,\mathrm{T}}} &= \frac{\frac{1}{2}m_W^2}{\frac{1}{2}m_{W,\mathrm{T}}^2} \\
\Leftrightarrow \qquad p'_{\nu,\mathrm{T}} &= \frac{m_W^2}{m_{W,\mathrm{T}}^2} \cdot p_{\nu,\mathrm{T}} \,.
\end{aligned}$$

With this reestimated value for $p_{\nu,\mathrm{T}}$, the $p_{\nu,z}$ estimate looks like

$$\begin{aligned}
p_{\nu,z} &= \frac{p'_{\nu,\mathrm{T}}}{p_{\ell,\mathrm{T}}} \cdot p_{\ell,z} \\
&= \frac{m_W^2}{m_{W,\mathrm{T}}^2} \cdot \frac{p_{\nu,\mathrm{T}}}{p_{\ell,\mathrm{T}}} \cdot p_{\ell,z} \,.
\end{aligned}$$

No ϕ_ν adjustment approach. In the same way, one may think of adjusting only ϕ_ν instead, keeping $p_{\nu,\mathrm{T}}$ constant. The solution would again look like

$$p_{\nu,z} = \frac{p'_{\nu,\mathrm{T}}}{p_{\ell,\mathrm{T}}} \cdot p_{\ell,z} \,.$$

However, as only ϕ_ν was modified, while $p'_{\nu,\mathrm{T}} = p_{\nu,\mathrm{T}}$, this leads to the same estimate for $p_{\nu,z}$ as the minimum mass approach.

The discriminant ignorance approach

This is a very pragmatic approach. As the square root term of equation (A.5) is what keeps us from getting a solution, the term is simply ignored, leaving a unique estimate for $p_{\nu,z}$:

$$p_{\nu,z} = \frac{\alpha \cdot p_{\ell,z}}{p_{\ell,\mathrm{T}}^2} \,.$$

Also here it is apparent, that for $m_{W,\mathrm{T}} \to m_W$ this gives the same result as the one-solution case, as the square-root term then really vanishes.

Choice for the kinematic fit

Except for the minimum deviation approach, all the discussed approaches were tried out in the course of this thesis, and the resulting $p_{\nu,z}$ estimates were compared to the true values. The best resolution was obtained with the $p_{\nu,\mathrm{T}}$ adjustment approach. Therefore this approach was used in the end to calculate the initial values for the kinematic fit in the no-solution case. For the $p_{\nu,x}$ and $p_{\nu,y}$ parameters, the measured

values were kept as the initial values rather than using the reestimated values, however, in order to stay consistent with the TFs.

A.2.3 Two solutions

The good point about the two-solution case is that there are exact solutions that fulfill the W mass constraint. The bad point about the two-solution case is that it is not known which is the "correct" one, i.e. which one is closer to the true neutrino p_z. The problem is that this ambiguity cannot be resolved by kinematic considerations. The approach followed within this thesis is to make a guess by choosing the solution that has the smaller $|p_{\nu,z}|$. Comparisons to truth showed that this is the "correct" choice in about 60% of the cases.

Appendix B

Truth matching

Truth matching in the scope of this thesis was mainly used to find out the true jet permutation and to compare the best KLFitter permutation with the true one in the context of KLFitter performance evaluation. For this purpose high efficiency and purity were not required; more important was to get an overall feeling for the performance, to be able to compare different settings and to track changes. To fulfill these moderate requirements, a ΔR matching approach was used. The exact procedure used for determining the true jet permutation and for comparing the KLFitter-favored permutation to the true one is described in detail in the following.

B.1 Quark-jet matching

This section describes the procedure that was used for the quark-jet matching, i.e. to uniquely assign measured jets to the true final-state quarks in true lepton+jets events. As explained, a ΔR matching approach was used for this purpose, which is detailed in the following.

Simple ΔR mapping

Figure B.1 shows the distributions of the distances, ΔR, of each of the final-state quarks to all measured jets, to only the respective closest jet and to all but the closest jet. A clear separation between the distributions of the distance to the closest jet and the distributions of the distances to all but the closest jet can be seen for all three quarks. The crossing point of the two distributions is located between $\Delta R = 0.4$ and $\Delta R = 0.5$ in all cases, the minimum of the inclusive distribution is located around $\Delta R = 0.4$ for one of the light quarks and between 0.4 and 0.5 for the two other quarks. This motivated the cut value for mapping a jet and a quark to each other to be chosen at $\Delta R = 0.4$. No further optimization for this cut value was performed, considering the given low demands of the truth matching.

Resolving ambiguities

The distribution of the number of jets that pass the ΔR cut for each quark is shown in figure B.2. It can be seen that for each of the quarks this is more than one jet in between 1% and 2% of all events. The same also holds vice versa (though not shown here explicitly). Thus, the mapping between the quarks and the jets provided by the ΔR cut alone is generally not unique. These mappings can rather be regarded as matching candidates. Therefore, in order to finally match a jet to a quark, additional criteria are applied. In case the mapping between a quark and a jet is one-to-one, i.e. neither the quark nor the jet is mapped to any other jet or quark, respectively, the jet is considered as matched to this quark. In case one jet is mapped to more than one quark, but none of these quarks is mapped to any other jet, the jet gets matched to that quark that has the smallest ΔR distance to the jet. All the other quarks that were mapped to this jet are considered unmatched. The same holds the other way around in case one quark is

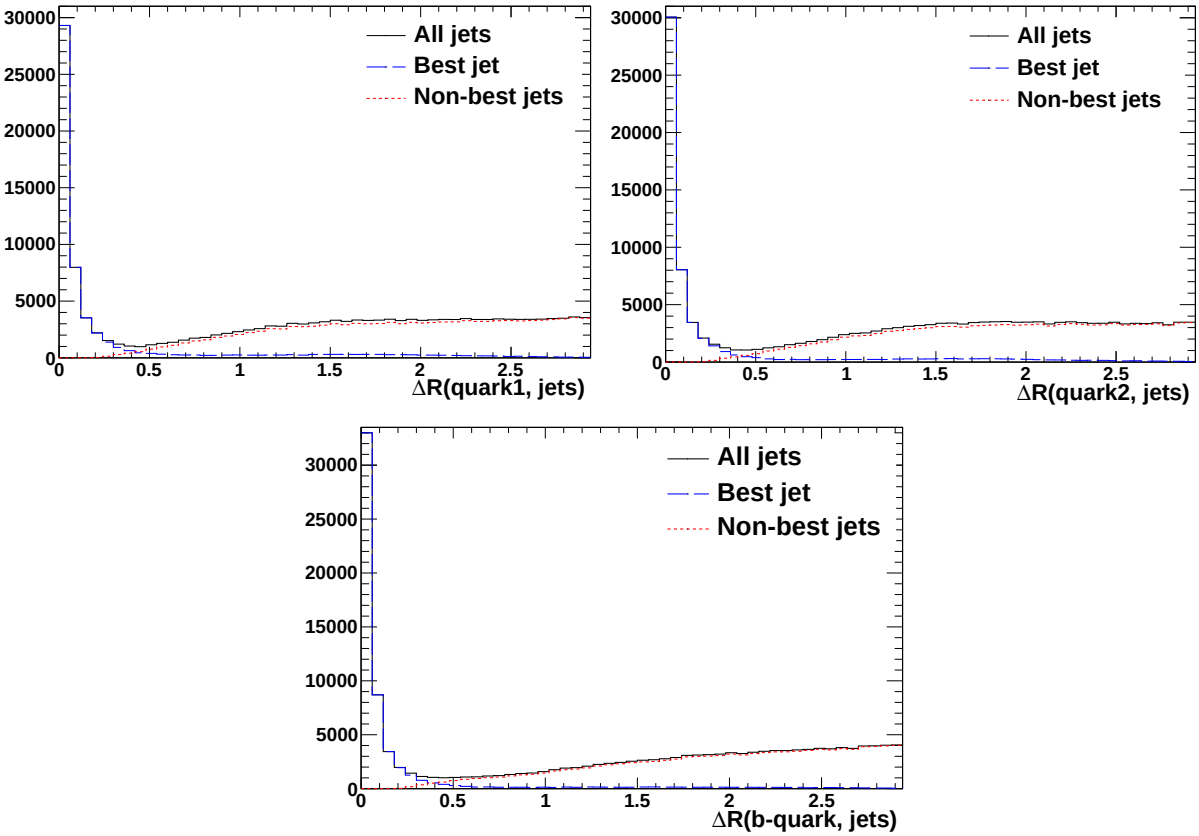


Figure B.1: Distribution of the distances, ΔR, of the two light quarks (top left and top right) and of the b quark (bottom) to all measured jets (solid black line), to the closest jet (dashed blue line) and to all but the closest jet (dotted red line). A clear separation between the distributions for the closest and for all but the closest jet can be seen, motivating the chosen cut at $\Delta R = 0.4$ for the mapping between quarks and jets.

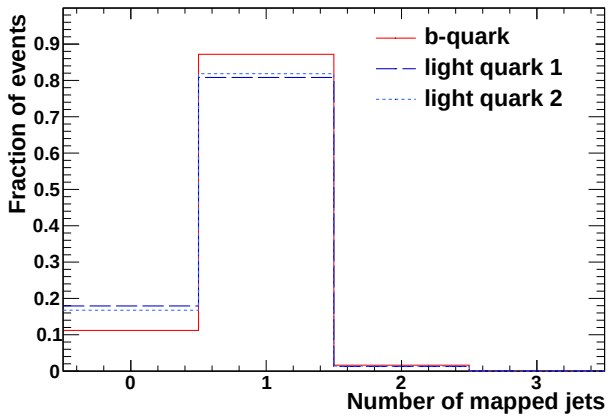


Figure B.2: Number of jets that pass the ΔR cut for each quark. For each of the quarks, this number is greater than one in between 1% and 2% of the events, giving rise to ambiguities that have to be resolved in order to obtain a unique matching.

mapped to more than one jet. All other kinds of ambiguities remain unresolved, and all quarks that are involved in such mappings remain unmatched.

Now we have established a well-defined procedure to obtain a unique matching of the measured jets to the quarks. The results that one obtains with this procedure are reported in Chapter 7.1.2.

B.2 Fit-truth matching

For the evaluation of the fit efficiency of the KLFitter a fit-truth matching was used, i.e. a procedure to match the model quarks of the best-fit permutation to the true quarks. This procedure is described in the following.

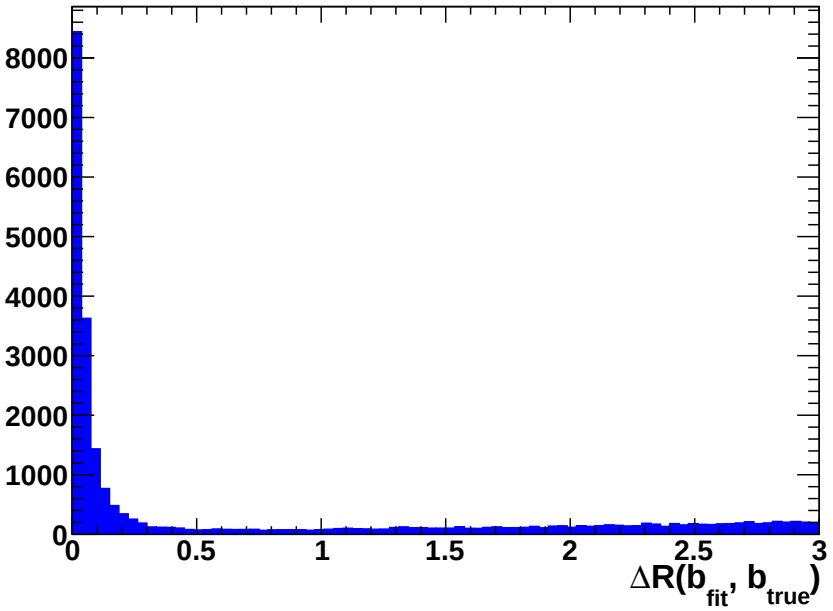


Figure B.3: Distribution of the distance, ΔR, between the b quark of the best-fit permutation and the true b quark.

Figure B.3 shows the distribution of the ΔR distance between the b quark of the best-fit permutation and the true b quark. Again a cut at $\Delta R = 0.4$ was used to consider a fit quark as matched to the true quark. As the b quark is uniquely determined in both the best-fit and true permutation, no ambiguities have to be resolved here.

For the light quarks the situation is a bit more complicated. They are indistinguishable, i.e. it is not clear a priori which of the best-fit quarks has to be compared to which of the true quarks. Two different assignments are possible, which are always referred to as the two possible (quark) *combinations* in the following. For each combination there are two ΔR values, one for the first and one for the second light quark. As a first approach, the two possible combinations are now classified by the larger of the two ΔR values. For each combination classified in that manner, the resulting two-dimensional distributions of the two ΔR values versus each other are shown in figure B.4. It is clearly visible that for the combination with the higher maximum ΔR there are no entries that are below 0.4 on both axes. This means that there are no events in which both fit quarks can be matched to the true quarks for both combinations. This is taken as the first criterion to pick one quark combination as the correct one: if for one combination both quarks can be matched, this combination is picked. The remaining events have at most one matched quark for any combination. For those events that have no matched quark at all, choosing a combination

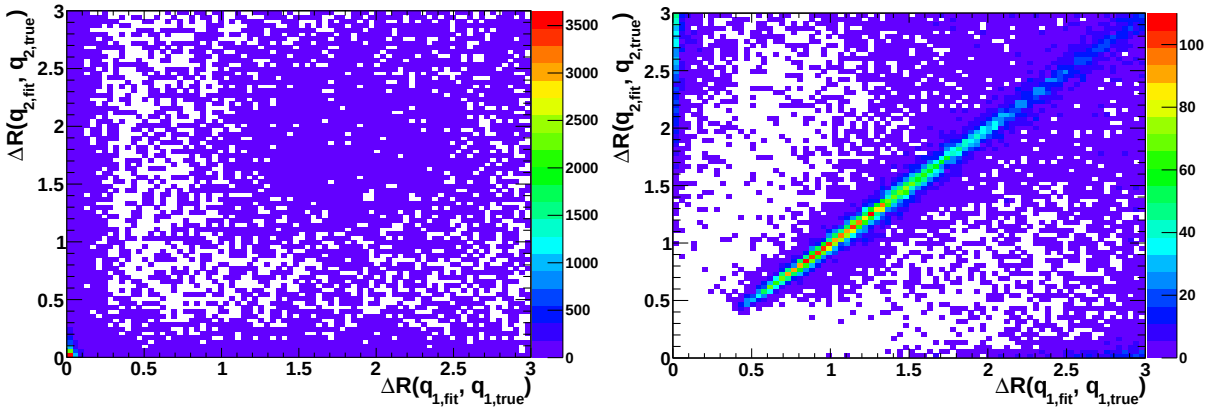


Figure B.4: Distribution of the distances, ΔR, between the light quarks of the best-fit permutation and the true light quarks, for the combination for which the larger of the two ΔR values is smaller (a) and for the combination for which the larger of the two ΔR values is larger (b) than for the other combination.

is irrelevant and can be done just randomly. For those events where exactly one true quark has a match by at least one of the fit quarks, the two-dimensional distribution of the ΔR distances of this true quark to both fit quarks versus each other is shown in figure B.5. There are only very few events for which both fit quarks match the true quark. For these events, that quark combination is chosen that has the smaller ΔR for the match. For all other events only one combination has a match and this combination can be picked as the obvious choice.

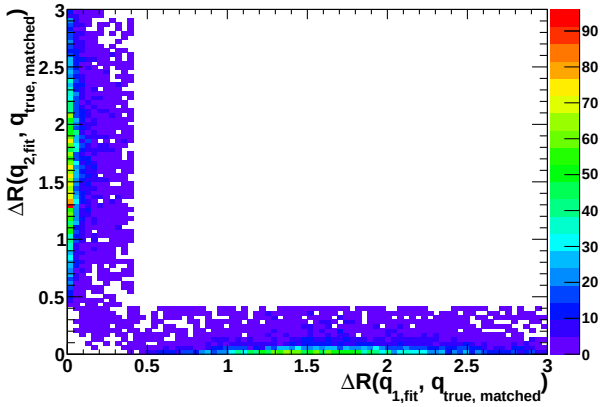
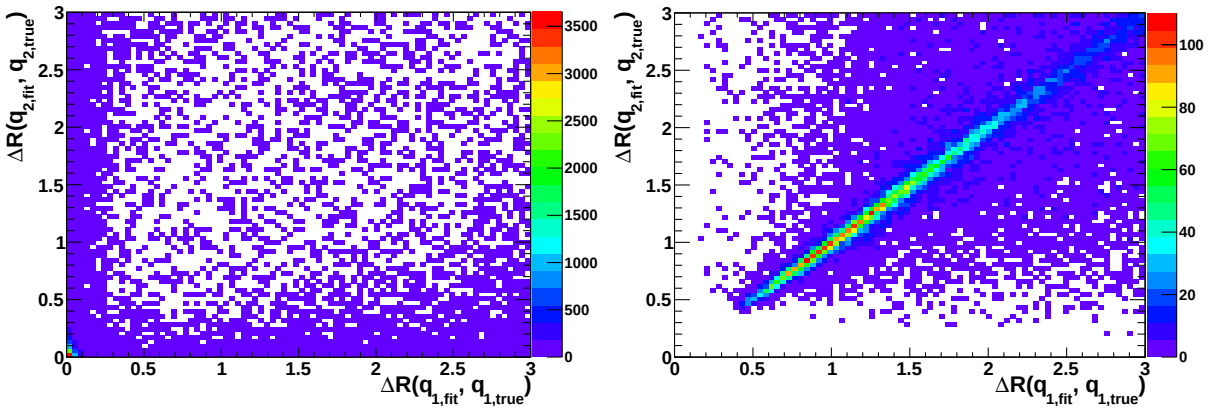


Figure B.5: Two-dimensional distribution of the distances, ΔR, of both fit quarks to the true quark versus each other for events where exactly one true quark has a match by at least one of the fit quarks.

The resulting two-dimensional ΔR distributions for both quarks versus each other for both the picked and the discarded combination is shown in figure B.6. It can be seen nicely that the discarded combinations contain only very few matches, while most of the picked combinations contain at least one match. The individual ΔR distribution for each quark of the picked combination is shown is figure B.7.

Now we also have established a well-defined procedure to obtain a unique matching of the best-fit quarks to the true quarks.



(a) (b)

Figure B.6: Distribution of the distances, ΔR, between the light quarks of the best-fit permutation and the true light quarks, for picked the combination (a) and for the discarded combination (b).

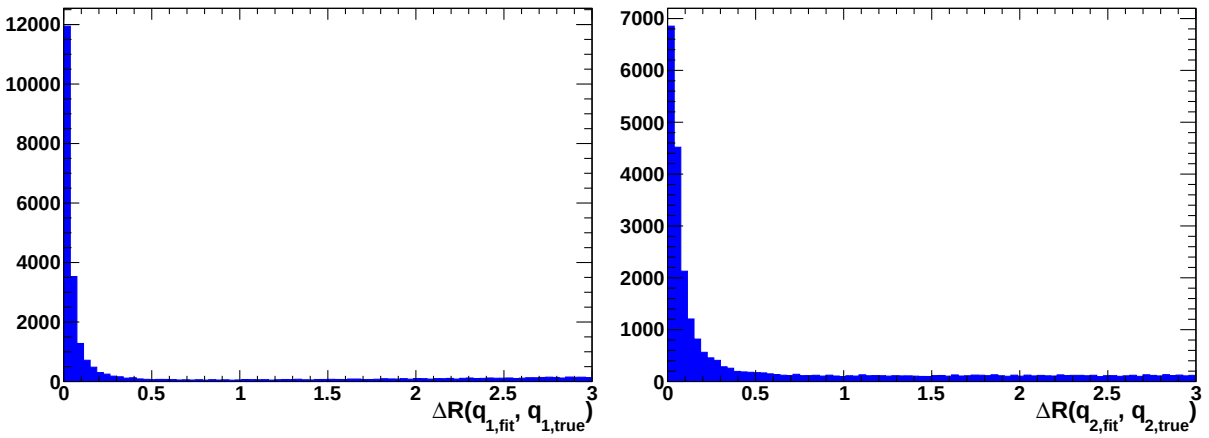


Figure B.7: Distribution of the ΔR distance of both light quarks to the respective assigned true quark according to the picked combination.

Bibliography

[1] S.L. Glashow, 'Partial Symmetries of Weak Interactions', *Nucl.Phys.* 22 (4 1961) 579–588, DOI: 10.1016/0029-5582(61)90469-2.

[2] Steven Weinberg, 'A Model of Leptons', *Phys.Rev.Lett.* 19 (1967) 1264–1266, DOI: 10.1103/PhysRevLett.19.1264.

[3] Abdus Salam, 'Weak and Electromagnetic Interactions', *Elementary particle theory. Relativistic groups and analyticity. Proceedings of the Eighth Nobel Symposium*, ed. by Nils Svartholm, Stockholm, 1968 367–377.

[4] S.L. Glashow, J. Iliopoulos and L. Maiani, 'Weak Interactions with Lepton-Hadron Symmetry', *Phys.Rev.* D2 (1970) 1285–1292, DOI: 10.1103/PhysRevD.2.1285.

[5] H.David Politzer, 'Reliable Perturbative Results for Strong Interactions?', *Phys.Rev.Lett.* 30 (1973) 1346–1349, DOI: 10.1103/PhysRevLett.30.1346.

[6] H.David Politzer, 'Asymptotic Freedom: An Approach to Strong Interactions', *Phys.Rept.* 14 (4 1974) 129–180, DOI: 10.1016/0370-1573(74)90014-3.

[7] D.J. Gross and Frank Wilczek, 'Asymptotically Free Gauge Theories. 1', *Phys.Rev.* D8 (1973) 3633–3652, DOI: 10.1103/PhysRevD.8.3633.

[8] Steven Weinberg, 'The Making of the standard model', *Eur.Phys.J.* C34 (2004) 5–13, DOI: 10.1140/epjc/s2004-01761-1, arXiv:hep-ph/0401010 [hep-ph].

[9] Gerard 't Hooft, 'Renormalizable Lagrangians for Massive Yang-Mills Fields', *Nucl.Phys. B* 35 (1 1971) 167–188, DOI: 10.1016/0550-3213(71)90139-8.

[10] Gerard 't Hooft and M.J.G. Veltman, 'Regularization and Renormalization of Gauge Fields', *Nucl.Phys. B* 44 (1 1972) 189–213, DOI: 10.1016/0550-3213(72)90279-9.

[11] Gerard 't Hooft and M.J.G. Veltman, 'Combinatorics of gauge fields', *Nucl.Phys.* B50 (1972) 318–353, DOI: 10.1016/S0550-3213(72)80021-X.

[12] Makoto Kobayashi and Toshihide Maskawa, 'CP Violation in the Renormalizable Theory of Weak Interaction', *Prog.Theor.Phys.* 49 (1973) 652–657, DOI: 10.1143/PTP.49.652.

[13] Particle Data Group, K. Nakamura et al., 'Review of particle physics', *J.Phys.G* 37 (2010) 075021, DOI: 10.1088/0954-3899/37/7A/075021.

[14] Stefano Frixione and Bryan R. Webber, 'Matching NLO QCD computations and parton shower simulations', *JHEP* 06 (2002) 029, DOI: 10.1088/1126-6708/2002/06/029, arXiv:hep-ph/0204244 [hep-ph].

[15] Borut Paul Kersevan and Elzbieta Richter-Was, 'The Monte Carlo event generator AcerMC version 1.0 with interfaces to PYTHIA 6.2 and HERWIG 6.3', *Comput.Phys.Commun.* 149 (3 2003) 142–194, DOI: 10.1016/S0010-4655(02)00592-1, arXiv:hep-ph/0201302 [hep-ph].

[16] Gregory Soyez, 'The SISCone and anti-k_t jet algorithms' (2008), arXiv:0807.0021 [hep-ph].

[17] Oliver Sim Brüning et al., 'LHC Design Report. 1. The LHC Main Ring', CERN-2004-003-V-1, CERN-2004-003, CERN, 2004; Oliver Sim Brüning et al., 'LHC Design Report. 2. The LHC infrastructure and general services', CERN-2004-003-V-2, CERN-2004-003, CERN, 2004; Michael Benedikt et al., 'LHC Design Report. 3. The LHC injector chain', CERN-2004-003-V-3, CERN-2004-003, CERN, 2004.

[18] (ed.) Evans Lyndon and (ed.) Bryant Philip, 'LHC Machine', *JINST* 3 (2008) S08001, DOI: 10.1088/1748-0221/3/08/S08001.

[19] ATLAS Collaboration, G. Aad et al., 'ATLAS detector and physics performance: Technical Design Report, 1', ATLAS-TDR-014, CERN-LHCC-99-014, CERN, 1999.

[20] 'Some LHC milestones.... Quelques dates clés du LHC...', *CERN Bulletin* (38 2008), Cover article, BUL-NA-2008-132.

[21] LHCb Collaboration, A.Augusto Alves et al., 'The LHCb Detector at the LHC', *JINST* 3 (2008) S08005, DOI: 10.1088/1748-0221/3/08/S08005.

[22] ALICE Collaboration, K. Aamodt et al., 'The ALICE experiment at the CERN LHC', *JINST* 3 (2008) S08002, DOI: 10.1088/1748-0221/3/08/S08002.

[23] ATLAS Collaboration, G. Aad et al., 'The ATLAS Experiment at the CERN Large Hadron Collider', *JINST* 3 (2008) S08003, DOI: 10.1088/1748-0221/3/08/S08003.

[24] CMS Collaboration, R. Adolphi et al., 'The CMS experiment at the CERN LHC', *JINST* 3 (2008) S08004, DOI: 10.1088/1748-0221/3/08/S08004.

[25] TOTEM Collaboration, G. Anelli et al., 'The TOTEM experiment at the CERN Large Hadron Collider', *JINST* 3 (2008) S08007, DOI: 10.1088/1748-0221/3/08/S08007.

[26] LHCf Collaboration, O. Adriani et al., 'The LHCf detector at the CERN Large Hadron Collider', *JINST* 3 (2008) S08006, DOI: 10.1088/1748-0221/3/08/S08006.

[27] James Pinfold et al., 'Technical Design Report of the MoEDAL Experiment', CERN-LHCC-2009-006. MoEDAL-TDR-001, CERN, 2009.

[28] Joseph R. Incandela et al., 'Status and Prospects of Top-Quark Physics', *Prog.Part.Nucl.Phys.* 63 (2 2009) 239–292, DOI: 10.1016/j.ppnp.2009.08.001, arXiv:0904.2499 [hep-ex].

[29] (ed.) Brock Ian C. and (ed.) Schorner-Sadenius Thomas, *Physics at the Terascale*, 2011, ISBN: 3527410015.

[30] F. Halzen and Alan D. Martin, *Quarks and Leptons: An Introductory Course in Modern Particle Physics*, 1984, ISBN: 9780471887416.

[31] B. Pietrzyk, 'LEP asymmetries and fits of the standard model' (1994), arXiv:hep-ex/9406001 [hep-ex].

[32] CDF Collaboration, F. Abe et al., 'Observation of top quark production in $\bar{p}p$ collisions', *Phys.Rev.Lett.* 74 (1995) 2626–2631, DOI: 10.1103/PhysRevLett.74.2626, arXiv:hep-ex/9503002 [hep-ex]; D0 Collaboration, S. Abachi et al., 'Observation of the top quark', *Phys.Rev.Lett.* 74 (1995) 2632–2637, DOI: 10.1103/PhysRevLett.74.2632, arXiv:hep-ex/9503003 [hep-ex].

[33] CDF Collaboration, T. Aaltonen et al., 'Observation of Electroweak Single Top Quark Production', *Phys.Rev.Lett.* 103 (2009) 092002, DOI: 10.1103/PhysRevLett.103.092002, arXiv:0903.0885 [hep-ex]; D0 Collaboration, V.M. Abazov et al., 'Observation of Single Top Quark Production', *Phys.Rev.Lett.* 103 (2009) 092001, DOI: 10.1103/PhysRevLett.103.092001, arXiv:0903.0850 [hep-ex].

[34] ATLAS Collaboration, G. Aad et al., 'Measurement of the top quark pair production cross-section based on a statistical combination of measurements of dilepton and single-lepton final states at $\sqrt{s} = 7$ TeV with the ATLAS detector', ATLAS-CONF-2011-108, CERN, 2011.

[35] M. Aliev et al., 'HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR', *Comput.Phys.Commun.* 182 (4 2011) 1034–1046, DOI: 10.1016/j.cpc.2010.12.040, arXiv:1007.1327 [hep-ph].

[36] ATLAS Collaboration, G. Aad et al., 'Observation of t Channel Single Top-Quark Production in pp Collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector', ATLAS-CONF-2011-088, CERN, 2011.

[37] A.D. Martin et al., 'Physical gluons and high-E_T jets', *Phys.Lett. B* 604 (1-2 2004) 61–68, DOI: 10.1016/j.physletb.2004.10.040, arXiv:hep-ph/0410230 [hep-ph].

[38] Nikolaos Kidonakis, 'Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production', *Phys.Rev.* D83 (2011) 091503, DOI: 10.1103/PhysRevD.83.091503, arXiv:1103.2792 [hep-ph]; Nikolaos Kidonakis, 'NNLL resummation for s-channel single top quark production', *Phys.Rev.* D81 (2010) 054028, DOI: 10.1103/PhysRevD.81.054028, arXiv:1001.5034 [hep-ph]; Nikolaos Kidonakis, 'Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-', *Phys.Rev.* D82 (2010) 054018, DOI: 10.1103/PhysRevD.82.054018, arXiv:1005.4451 [hep-ph].

[39] A.D. Martin et al., 'Parton distributions for the LHC', *Eur.Phys.J. C* 63.2 (2009) 189–285, DOI: 10.1140/epjc/s10052-009-1072-5, arXiv:0901.0002 [hep-ph].

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

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

[42] Tilman Plehn, 'Lectures on LHC Physics' (2009), arXiv:0910.4182 [hep-ph].

[43] T Cornelissen et al., 'Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)', ATL-SOFT-PUB-2007-007, CERN, 2007.

[44] ATLAS Collaboration, G. Aad et al.,
'Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics' (2009), arXiv:0901.0512 [hep-ex].

[45] ATLAS Collaboration, G. Aad et al.,
'Performance of the ATLAS Detector using First Collision Data', *JHEP* 09 (2010) 056, DOI: 10.1007/JHEP09(2010)056, arXiv:1005.5254 [hep-ex].

[46] R. Frühwirth, 'Application of Kalman filtering to track and vertex fitting', *Nucl.Instrum.Meth. A* 262 (2-3 1987) 444–450, DOI: 10.1016/0168-9002(87)90887-4.

[47] ATLAS Collaboration, G. Aad et al.,
'Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV', CERN-PH-EP-2011-191, Submitted to European Physical Journal C: CERN, 2011, arXiv:1112.6426.

[48] ATLAS Collaboration, G. Aad et al., 'Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data', CERN-PH-EP-2011-117, Submitted to Eur. Phys. J. C: CERN, 2011, arXiv:1110.3174.

[49] ATLAS Collaboration, G. Aad et al.,
'Expected electron performance in the ATLAS experiment', ATL-PHYS-PUB-2011-006, CERN, 2011.

[50] ATLAS Collaboration, G. Aad et al., 'Electron and photon reconstruction and identification in ATLAS: expected performance at high energy and results at 900 GeV', ATLAS-CONF-2010-005, CERN, 2010.

[51] Th Lagouri et al., 'A Muon Identification and Combined Reconstruction Procedure for the ATLAS Detector at the LHC at CERN', *IEEE Trans.Nucl.Sci.* 51 (6 2004) 3030–3033, DOI: 10.1109/TNS.2004.839102.

[52] ATLAS Collaboration, G. Aad et al.,
'Studies of the performance of the ATLAS detector using cosmic-ray muons', *Eur.Phys.J. C* 71.3 (2011) 1593, DOI: 10.1140/epjc/s10052-011-1593-6, arXiv:1011.6665 [physics.ins-det].

[53] ATLAS Collaboration, G. Aad et al., 'Determination of the muon reconstruction efficiency in ATLAS at the Z resonance in proton-proton collisons at $\sqrt{s} = 7$ TeV', ATLAS-CONF-2011-008, CERN, 2011.

[54] ATLAS Collaboration, G. Aad et al.,
'Jet energy resolution and selection efficiency relative to track jets from in-situ techniques with the ATLAS Detector Using Proton-Proton Collisions at a Center of Mass Energy $\sqrt{s} = 7$ TeV', ATLAS-CONF-2010-054, CERN, 2010.

[55] ATLAS Collaboration, G. Aad et al., 'Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data', ATLAS-CONF-2011-102, CERN, 2011.

[56] Giacinto Piacquadio and Christian Weiser,
'A new inclusive secondary vertex algorithm for b-jet tagging in ATLAS', *J.Phys.Conf.Ser.* 119 (2008) 032032, DOI: 10.1088/1742-6596/119/3/032032.

[57] ATLAS Collaboration, G. Aad et al., 'Performance of Missing Transverse Momentum Reconstruction in Proton-Proton Collisions at 7 TeV with ATLAS', *Eur. Phys. J. C* 72 (2011) 1844, DOI: 10.1140/epjc/s10052-011-1844-6, arXiv:1108.5602 [hep-ex].

[58] ATLAS Collaboration, G. Aad et al., 'Performance of the Missing Transverse Energy Reconstruction and Calibration in Proton-Proton Collisions at a Center-of-Mass Energy of 7 TeV with the ATLAS Detector', ATLAS-CONF-2010-057, CERN, 2010.

[59] S Ask et al., 'Report from the Luminosity Task Force', ATL-GEN-PUB-2006-002. ATL-COM-GEN-2006-003. CERN-ATL-COM-GEN-2006-003, CERN, 2006.

[60] ATLAS Collaboration, G. Aad et al., 'Luminosity Determination in pp Collisions at $\sqrt{s} = 7$ TeV using the ATLAS Detector at the LHC', *Eur.Phys.J. C* 71.4 (2011) 1630, DOI: 10.1140/epjc/s10052-011-1630-5, arXiv:1101.2185 [hep-ex].

[61] ATLAS Collaboration, G. Aad et al., 'Luminosity Determination in pp Collisions at $\sqrt{s} = 7$ TeV using the ATLAS Detector in 2011', ATLAS-CONF-2011-116, CERN, 2011.

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

[63] A. Sherstnev and R.S. Thorne, 'Parton Distributions for LO Generators', *Eur.Phys.J. C* 55.4 (2008) 553–575, DOI: 10.1140/epjc/s10052-008-0610-x, arXiv:0711.2473 [hep-ph].

[64] Pavel M. Nadolsky et al., 'Implications of CTEQ global analysis for collider observables', *Phys.Rev.D* 78 (1 2008) 013004, DOI: 10.1103/PhysRevD.78.013004, arXiv:0802.0007 [hep-ph].

[65] Michelangelo L. Mangano et al., 'ALPGEN, a generator for hard multiparton processes in hadronic collisions', *JHEP* 07 (2003) 001, DOI: 10.1088/1126-6708/2003/07/001, arXiv:hep-ph/0206293 [hep-ph].

[66] J. Pumplin et al., 'New generation of parton distributions with uncertainties from global QCD analysis', *JHEP* 07 (2002) 012, DOI: 10.1088/1126-6708/2002/07/012, arXiv:hep-ph/0201195 [hep-ph].

[67] J. M. Butterworth, Jeffrey R. Forshaw and M. H. Seymour, 'Multiparton interactions in photoproduction at HERA', *Z. Phys. C* 72.4 (1996) 637–646, DOI: 10.1007/s002880050286, arXiv:hep-ph/9601371.

[68] K Becker et al., 'Mis-identified lepton backgrounds in top quark pair production studies for EPS 2011 analyses', internal report ATL-COM-PHYS-2011-768, CERN, 2011.

[69] J Erdmann et al., 'Kinematic fitting of ttbar-events using a likelihood approach: The KLFitter package', internal report ATL-COM-PHYS-2009-551, CERN, 2009.

[70] Allen C. Caldwell, Daniel Kollar and Kevin Kroninger, 'BAT: The Bayesian analysis toolkit', *Computer Physics Communications* 180 (11 2009) 2197–2209, DOI: 10.1016/j.cpc.2009.06.026, arXiv:0808.2552 [physics.data-an]; Allen C. Caldwell, Daniel Kollar and Kevin Kroninger, 'BAT: The Bayesian analysis toolkit', *J.Phys.Conf.Ser.* 219 (2010) 032013, DOI: 10.1088/1742-6596/219/3/032013.

[71] M. Feindt and U. Kerzel, 'The NeuroBayes neural network package', *Nucl.Instrum.Meth. A* 559 (1 2006) 190–194, DOI: 10.1016/j.nima.2005.11.166.

[72] J. Stillings, 'Private communication', 2011.

List of Figures

List of Tables

Danksagung

Ich möchte mich sehr herzlich bei allen bedanken, die durch ihre Hilfe und Unterstützung zum Entstehen und Gelingen dieser Doktorarbeit beigetragen haben.

An erster Stelle möchte ich meinem Doktorvater, Prof. Ian C. Brock, danken: dafür, dass er mir die Möglichkeit gegeben hat in seiner Gruppe zu promovieren, als ich mich mit zweijähriger Verzögerung schließlich doch noch dazu entschieden hatte; für all die Geduld, die er auch sonst immer hatte, wann immer etwas ein wenig länger dauerte; für die stetige Unterstützung; und für all den Freiraum, den ich bei meiner wissenschaftlichen Arbeit stets hatte. Meinen weiteren Prüfern, Herrn Prof. Klaus Desch, Herrn Prof. Carsten Urbach und Herrn Prof. Heiko Röglin, danke ich, dass sie sich die Zeit genommen haben, sich als Mitglieder der Prüfungskommission mit meiner Doktorarbeit zu befassen.

All meinen aktuellen und ehemaligen Kollegen – insbesondere denen, die im Laufe der Zeit zahlreiche Büros mit mir geteilt haben – danke ich für die stets angenehme Arbeitsatmosphäre sowie für zahlreiche hilfreiche, fachliche wie auch nicht-fachliche, Diskussionen. Namentlich erwähnen möchte ich hier stellvertretend Detlef Bartsch, Verena Schönberg, Markus Jüngst, Robert Zimmermann, Jan Stillings, Sebastian Mergelmeyer, Steffen Schaepe, Ozan Arzlan und Pienpen Seema. Besonderer Dank gebührt Jan Stillings und Sebastian Mergelmeyer fürs Korrekturlesen der Doktorarbeit.

Meinen beiden Mitbewohnern während meiner Zeit am CERN, Carolin Zendler und Robindra Prabhu, danke ich für die schöne WG-Zeit. Weiterhin danke ich den (damaligen) Mitgliedern der "Reculet Hikers' Society" für die zahlreichen schönen Kochabende und sonstigen gemeinsamen Aktivitäten, stellvertretend namentlich hervorgehoben seien hier Detlef, Nico und Holger und Kristof.

Schließlich gilt mein ganz besonderer Dank meiner Frau Rebecca, die mich während der ganzen Zeit immer unterstützt und mir einen sicheren Rückhalt gegeben hat, auch wenn sie es dabei besonders während der "heißen", stressigen Phase oft nicht leicht mit mir hatte. Vielen Dank! Ohne Dich wäre das alles viel schwieriger gewesen.