Study of the Properties of Additional Jets in Top Pair Events with the ATLAS Detector

Diplomarbeit
zur Erlangung des wissenschaftlichen Grades
Diplom-Physiker

vorgelegt von

Felix Thomas
geboren am 01.02.1986 in Riesa

CERN-THESIS-2011-050
//2011

TECHNISCHE UNIVERSITÄT DRESDEN

INSTITUT FÜR KERN- UND TEILCHENPHYSIK
FACHRICHTUNG PHYSIK
FAKULTÄT FÜR MATHEMATIK UND NATURWISSENSCHAFTEN

2011

Eingereicht am 30.06.2011

1. Gutachter: Prof. Dr. Michael Kobel
2. Gutachter: Dr. Ulrich Husemann

Abstract

After the successful rediscovery of the top-quarks at the LHC whose properties have already been described satisfactorily by Monte Carlo simulations, precision measurements are now the next step. The aim of this thesis is to study the properties of the additional jets to the top pair final state and to compare the predictions from Monte Carlo simulations with real data. Various sensitive kinematical distributions of these jets have been studied using a kinematical fit for the final event selection. The influence of different top reconstruction algorithms on the variables have been studied and the systematical uncertainties derived.

Kurzdarstellung

Nach der Wiederentdeckung des Top-Quarks am LHC beginnt nun die Zeit der Präzisionsmessungen im Bereich der Top-Physik. Die dabei verfolgten Ziele sind zum einen die Vermessung der Eigenschaften des Top-Quarks mit möglichst hoher Genauigkeit sowie die Detektorkalibration für die Prozesse der Top-Physik als Standardkerze dienen sollen. Das Ziel dieser Arbeit ist die Untersuchung der Eigenschaften zusätzlicher Jets, die neben den Objekten des Top-Paar Endzustandes im Event auftreten und deren Vergleich in Daten und Monte Carlo Simulationen. Die schnitt-basierte Eventselektion wird durch die Rekonstruktion des Top-Paares mittels einen kinematischen Fits ergänzt um die kinematischen Eigenschaften der zusätzlichen Jets zu bestimmen. Der Einfluss verschiedener Algorithmen zur Top-Paar Rekonstruktion wurde untersucht und deren systematische Fehler ermittelt.

Contents

Contents

1 Introduction

At the time of writing the Large Hadron Collider (LHC) [1] has been collecting data for about two years. After a difficult start in September 2008 due to technical problems concerning the magnets of the LHC it resumed operation in November 2009 [2]. After a steady increase of the center-of-mass energy ($\sqrt{s}$) it now operates at a $\sqrt{s}$ of 7 TeV which is half the design value. Still this is the highest ever achieved $\sqrt{s}$ of any collider every built.
In the year 2010 enough data was accumulated for early physics studies. In 2011 the same amount of data that was taken in about eight months in 2010 can be recorded in less than a week. [3] This marks the transition to high precision measurements with new and exciting physics results.
A field of utmost importance is top physics. The top quark discovery in 1995 at the Tevatron [4, 5] marked the completion of the third generation of quarks.
The top quark has a variety of unique properties and their precise measurement is an ongoing effort. Not only is the top quark the heaviest quark of all (about 30 times heavier than the second heaviest quark, the bottom quark) it is also the one with the shortest lifetime. In fact its lifetime is shorter than the time it takes for it to hadronize, meaning that there are no bound states of top quarks which sets it apart from all other quarks. This makes it an especially interesting object for QCD studies. The high $\sqrt{s}$[1] and luminosity reached at the LHC an abundant amount of top quarks will be produced at the LHC making it a real "top factory". Precision measurements in the field of top physics will be limited by systematics not by statistics. Therefore processes involving top quarks will be used for detector calibration and the improvement of Monte Carlo (MC) simulations.
But top physics is also important for the discovery of new physics. Events involving top quarks exhibit rich final states with a plethora of physical objects. These might mimic events involving new physics e. g. that of hypothetical particles with masses approximately the same as the top mass. For some hypotheses involving Higgs particles this is exactly the case making top pair events an important background for such searches [6].
Therefore top physics is a real workhorse and major key for physics at the LHC and a deeper understanding of the physics involving top quarks is beneficial to a multitude of studies.

1.1 Motivation

Many analyses at the LHC are performed via the creation of predictions using a MC Generator which incorporates various assumptions. These predictions from MC

[1] As can be seen in Figure 2.1 the production rate of events involving top quarks is directly proportional to the $\sqrt{s}$

simulations (e. g. a cross-section of a given physics process) are then compared to the measured data. Deviations between both might lead to the discovery of new particles or physical phenomena. This demands very sophisticated theories and MC Generators delivering predictions that are precise and well understood because if there are conflicts between predictions of theory and measured data their sources have to be identified in order to extract meaningful information.
The theoretical descriptions employed by the generators are not entirely stemming from first principles. Some generation steps like the evolution from quarks to jets are not describable with perturbative theories and have to be dealt with via phenomenological models. These models work with input parameters which have to be measured in the experiment.
One interesting set of variables are the properties of jets that do not originate from the hard scattering process but from other sources such as initial state radiation. The distributions of the kinematic variables such as momentum, pseudo-rapidity or azimuthal angle of these extra jets can be measured and then be compared to theoretical predictions.
Another field of interest are the flavor specific variables of the extra jets. Using these variables statements about the probability that a certain jet originates from a quark of a specific flavor can be made. These probabilities are input variables for the aforementioned phenomenological models. Therefore the determination of the flavor fractions of jets contributes to the improvement of MC simulations.

1.2 Outline

Chapter 2 is a general overview of the theory that is the current and best explanation of the nature of elementary particles, the Standard Model (SM). This will be followed by a short description of the physics involving top quarks at hadron colliders.
Chapter 3 describes the LHC and the ATLAS detector. How the detector response is translated into physical objects is explained in Chapter 4. Additional information on the techniques of b-tagging will be given as these are crucial to the analysis.
Following these general parts will be Chapter 5 which deals with the analysis algorithms used in this study. First the prescriptions that define the physical objects used for analysis will be given. Then the event selection and top quark reconstruction strategies will be illustrated followed by a deeper analysis of the used top reconstruction algorithm. This will be accompanied by a possible way of correcting the shortcomings of said top reconstruction algorithm.
In Chapter 6 the results of the reconstruction and correction endeavors will be presented followed by Chapter 7, the conclusion.
Additional information concerning the analysis may be found in the appendix.

2 Theoretical Basics - The Standard Model of Particle Physics

The set of theories describing the interactions occurring in particle physics is called the Standard Model (SM). One cannot pinpoint one physicist who has done the main work of conceiving the theory. Overall about 50 physicists largely contributed to this framework with many of them receiving nobel prizes for their work[1]. The SM consists of the electroweak theory describing the weak and the electromagnetic interaction and Quantum Chromodynamics (QCD) describing the strong interaction.
All of the interactions in the SM are gauge theories and are described by quantum field theory. In a gauge theory the Lagrangian describing the physics is invariant under a specific set of continuous local transformations. An example for such a theory is the electroweak theory which will be discussed in Section 2.3.1.
It is remarkable that the SM although conceived mainly in the 60s and 70s of the last century has passed every experimental test done to this date with astonishing precision [10]. Although being capable of describing literally any physical process in particle physics, it fails to be a theory of everything.
One major drawback is the not yet successful implementation of gravitation into the context of gauge theories. Although being of negligible impact in particle physics, gravity should be incorporated into a theory that attempts to describe every physical process known to man.
Another possible drawback is that the SM does not offer a candidate for dark matter. Dark matter has not been observed directly but there is evidence for its existence [11]. Therefore dark matter may necessitate new physics beyond the SM.

2.1 Particle Content of the Standard Model

The SM describes nature with a set of elementary particles, which are defined as objects without a permanent substructure. In the SM there exist three generations of fermions (particles with spin 1/2), each consisting of a pair of quarks and a pair of leptons.
Some of their properties are listed in Table 2.1. Further information may be found in the recent issue of the Review of Particle Physics provided by the Particle Data Group [12]. For each of the fermions in Table 2.1 there is an antiparticle with same

[1]Some key papers can be found in [7–9]

Generation	1	2	3	Q	I_3^W	C	Interaction
Quarks	u (up)	c (charm)	t (top)	$+2/3$	$+1/2$	r,g,b	weak, strong, electromagn.
	d (down)	s (strange)	b (bottom)	$-1/3$	$-1/2$	r,g,b	
Leptons	e^- (electron)	μ^- (muon)	τ^- (tauon)	-1	$+1/2$	-	weak, electromagn.
	ν_1 (neutrinos)	ν_2	ν_3	0	$-1/2$	-	weak

Table 2.1: The fermions of the Standard Model. Each quark carries an electromagnetic charge Q, a weak isospin I_3^W and a color charge C. These charges belong to the electromagnetic, weak and strong interaction respectively. In contrast the charged leptons are missing a color charge and the neutrinos have only a weak isospin.

mass but oppositely signed charges, e. g. the antiparticle of the electron is the positron which has an electrical charge $Q = +1$[2].
Apart from their mass they have a set of quantum numbers governing their interactions and defining their place in the particle zoo of the SM. The nature of a quantum number is determined by the theoretical description of its associated interaction. For instance while the electrical charge Q of an elementary particle is an integer value its color charge C (the charge associated to the strong interaction) is a three-dimensional vector. Interestingly some of the particles have the same quantum numbers concerning the forces that affect them but different masses providing means of classification e. g. the three generations.
There are three generations of quarks and leptons. Only the elementary particles of the first generation are stable. Inside the generations quarks as well as leptons form pairs consisting of an up-type and a down-type. Through the weak interaction, discussed in Section 2.3.1 the particles inside the pair can be transformed into each other.
Because of the coupling of the strong interaction being close to 1 at low energies, quarks can not be observed as isolated elementary particles but form composite particles called hadrons. Mesons are hadrons formed by a quark and an antiquark (e. g. pions, kaons) with integer spin while baryons are hadrons consisting of three quarks (e. g. neutrons, protons) exhibiting spins of $2n + 1/2$. A peculiar property of quarks is that they are not observable as isolated elementary particles (in contrast to leptons) but that they can only be found as constituents of hadrons. This is called confinement

[2] There are other quantum numbers beside the electrical charge e. g. the third component of the weak isospin. All additive quantum numbers of an antiparticle are oppositely signed to the ones of the particle.

Interaction	strong	el.-mag.	weak	gravitation
physical phenomena	nuclear binding	electricity, magnetism, optics	radioactive decay	curvature of space-time continuum
rel. strength	1	10^{-2}	10^{-5}	10^{-38}
range	10^{-15} m	∞	10^{-18} m	∞
force carrier (mass) exp. (mass) theor.	gluons $<$ few MeV 0	photons $< 1 \times 10^{-18}$ eV 0	W and Z bosons ~ 100 GeV	gravitons? - 0
couples to	quarks	quarks, el. charged leptons	quarks, leptons	all massive particles
effect between ident. particles	repulsive	repulsive	repulsive	attractive

Table 2.2: The four fundamental forces of nature

and it bears some troubling implications for reconstruction algorithms which will be discussed Section 5.4.2.

2.2 Forces

The SM of particle physics explains three of the four interactions known to man namely the electromagnetic interaction described by Quantum Electrodynamics (QED) [13–15], the weak interaction described by Quantum Flavordynamics (QFD) and the strong interaction described by QCD [16–18].
The forces in Table 2.2 are all mediated via elementary particles called gauge bosons. In contrast to fermions these particles have a integer spin.

2.3 Theoretical Description

The starting point for the theoretical description is the Lagrangian of a free fermionic field

$$\mathcal{L} = i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi. \tag{2.1}$$

By applying the Euler-Lagrange equations one obtains the Dirac equation from equation 2.1.

As mentioned before the theories describing the three interactions between particles are gauge theories. These theories are quantum field theories that include the constraint of local gauge invariance meaning that the Lagrangian representing the physics is invariant under a continuous location dependent transformation. These transformations are governed by group theory. Transformations are classified by the properties of the generators that describe them.
For example the U(1) gauge symmetry for a fermion field is defined as

$$\psi(x_\mu, p^\mu) \to \psi'(x_\mu, p^\mu) = e^{i\theta(x_\mu)\hat{Q}}\psi(x_\mu, p^\mu) \tag{2.2}$$

$$\bar{\psi}(x_\mu, p^\mu) \to \bar{\psi}'(x_\mu, p^\mu) = e^{-i\theta(x_\mu)\hat{Q}}\bar{\psi}(x_\mu, p^\mu). \tag{2.3}$$

When applying the transformation to the Lagrangian it can be seen that it is not invariant under this transformation. Therefore it is necessary to introduce an additional term to the Lagrangian

$$\mathcal{L} = i\bar{\psi}\gamma^\mu(\partial_\mu + ie\hat{Q}A_\mu)\psi - m\bar{\psi}\psi, \tag{2.4}$$

where the field A_μ transforms as

$$A'_\mu = A_\mu - \frac{1}{e}\partial_\mu\theta(x_\mu). \tag{2.5}$$

Now the Lagrangian is invariant under the U(1) gauge transformation thus satisfying the requirement of the gauge theory. The newly introduced interaction term describes the interaction between the gauge field A_μ and the fermionic field

$$\mathcal{L}_{\text{int}} = -e\bar{\psi}\gamma^\mu Q\psi A_\mu. \tag{2.6}$$

This expression corresponds to the interaction of photons with charged particles.

2.3.1 The Electroweak Theory

The electroweak theory is a prime example for a gauge theory, an integral part of the SM describing the decays of unstable particles and one of the great successes of theoretical physics in the 20th century.
The Lagrangian of the theory which is invariant under the group SU(2) × U(1) may be written as follows:

$$\begin{aligned} L =& \bar{F}_L\gamma^\mu\left(i\partial_\mu - g\frac{1}{2}\tau_a \cdot W^a_\mu - g'\frac{Y}{2}B_\mu\right)F_L + \bar{F}_R\gamma^\mu\left(i\partial_\mu - g'\frac{Y}{2}B_\mu\right)F_R \\ &- \frac{1}{4}W^a_{\mu\nu}W^{\mu\nu}_a - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \end{aligned} \tag{2.7}$$

Here F_L denote left-handed fermion fields while F_R denote right-handed fermion fields. The left-handed fermionic fields are doublets under SU(2)

$$\begin{pmatrix} u_L \\ d_L \end{pmatrix}, \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \begin{pmatrix} t_L \\ b_L \end{pmatrix}, \begin{pmatrix} \nu_{e,L} \\ e^-_L \end{pmatrix}, \begin{pmatrix} \nu_{\mu,L} \\ \mu^-_L \end{pmatrix}, \begin{pmatrix} \nu_{\tau,L} \\ \tau^-_L \end{pmatrix},$$

while the right-handed fermionic fields are singlets:

$$u_R,\ d_R,\ c_R,\ s_R,\ t_R,\ b_R,\ e_R,\ \mu_R,\ \tau_R,\ \nu_{e,R},\ \nu_{\mu,R},\ \nu_{\tau,R}.$$

The fields introduced by the gauge symmetry are the three W_a and the B with their couplings g and g'. One representation of the generators τ_a of the SU(2) are the Pauli matrices

$$\tau_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \tau_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \tau_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{2.8}$$

while the generator of the U(1) part is Y.
With this in mind the interaction terms for the left-handed fermionic fields may be written as:

$$\bar{F}_L \gamma_\mu \frac{1}{2} \begin{pmatrix} gW_3^\mu + g'YB^\mu & g(W_1^\mu - iW_2^\mu) \\ g(W_1^\mu + iW_2^\mu) & -gW_3^\mu + g'YB^\mu \end{pmatrix} F_L. \tag{2.9}$$

The diagonal terms describe the interaction between left-handed fermionic fields of the same flavor while the off-diagonal terms describe the interactions between left-handed fermionic fields with different flavors. The gauge bosons that are realized in nature have to be linear combinations of the fields W_a and B. By demanding experimentally known facts like the parity conservation of the electromagnetic force[3] and that the photon only couples to charged particles the relationships between the photon field A and the W boson fields W^-, W^+ and Z_0 as well as between the couplings g, g' and e and g_w are

$$W_\mu^- = W_\mu^1 + iW_\mu^2, \tag{2.10}$$
$$W_\mu^+ = W_\mu^1 - iW_\mu^2, \tag{2.11}$$
$$A_\mu = \cos(\theta_W) B_\mu + \sin(\theta_W) W_\mu^3, \tag{2.12}$$
$$Z_\mu = -\sin(\theta_W) B_\mu + \cos(\theta_W) W_\mu^3, \tag{2.13}$$
$$g_w = g, \tag{2.14}$$
$$e = g\sin(\theta_W), \tag{2.15}$$

with θ_W being the Weinberg angle defined as:

$$\tan(\theta_W) = \frac{g'}{g}. \tag{2.16}$$

With these relationships it is now possible to calculate observables of processes described by the electroweak theory and test them experimentally.

[3]The parity conserving nature of the electromagnetic interaction is reflected by the fact that the photon couples to both left-handed and right-handed particles with the same coupling.

2.3.2 CKM Matrix

In the short description of the electroweak theory one important fact has been neglected. While the interactions take place between the flavor eigenstates of the fields their propagation is governed by their mass eigenstates. These two are linear combinations of each other and their transformation is described by the CKM matrix[4] [19,20]

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{2.17}$$

Here the primed quarks are the flavor eigenstates while the unprimed ones are the mass eigenstates. The magnitudes of the matrix elements are (errors were neglected for readability) [12]

$$\begin{pmatrix} 0.97428 & 0.2253 & 0.00347 \\ 0.2252 & 0.97345 & 0.0410 \\ 0.00862 & 0.0403 & 0.999152 \end{pmatrix}. \tag{2.18}$$

The implications of the concrete values of the CKM matrix directly affect the analysis. The matrix element V_{tb} is almost one and the elements V_{ts} and V_{td} are close to zero. This leads to the fact that the top quark almost exclusively decays to a bottom quark and a W boson making tagging techniques which identify bottom quarks a viable ingredient for an analysis.
In addition the small sizes of V_{ub} and V_{cb} suppress the decay of the bottom quark to the lighter charm and up quark. The decay into a top quark is kinematically forbidden as the top quark is much heavier than the bottom quark. Therefore hadrons that have bottom quarks as constituents exhibit long lifetimes. This leads to measurable decay lengths which are one possibility of identifying jets stemming from bottom quarks.

2.4 Top Physics

After having already mentioned the implications that arise for the top quark decay from the CKM matrix it is time to discuss the properties of this remarkable particle further.
As mentioned before the top quark was observed directly at the Tevatron in 1995 [4,5]. With the discovery of the bottom quark in 1977 [21] there was compelling evidence that an up-type quark for the third generation has to exist. By exploiting the precise measurement of the lifetime of the muon and the calculation of the muon decay process to one-loop order, theorists were able to estimate the top quark mass to be around 160 GeV. Eventually the top quark was found to be in exactly the mass range the theorists predicted.
The mass of the top quark as found in the Review of Particle Physics [12] is

$$m_{\text{top}} = 172.0 \pm 0.9 \pm 1.3 \text{ GeV}^5 \tag{2.19}$$

[4] CKM stands for Cabibbo, Kobayashi and Maskawa.
[5] The values are the measured mass and the statistical and systematical error on the mass mearuement.

which is comparable to the weight of a gold atom. This high mass has consequences for both production and decay processes of the top.

Top Quark Decay

The top decays almost exclusively via $t \to Wb$. Its decay width according to the SM is given as [22][6]

$$\Gamma_\mathrm{t} = \frac{G_\mathrm{F} m_\mathrm{t}^3}{8\pi\sqrt{2}} \left(1 - \frac{M_\mathrm{W}^2}{m_\mathrm{t}^2}\right)^2 \left(1 + 2\frac{M_\mathrm{W}^2}{m_\mathrm{t}^2}\right) \left[1 - \frac{2\alpha_s}{3\pi}\left(\frac{2\pi^2}{3} - \frac{5}{2}\right)\right]. \tag{2.20}$$

An apparent feature is that with increasing top quark mass its decay width becomes bigger. This leads to a smaller lifetime. The full width of 1.3 GeV translates to a mean lifetime of about $5 \cdot 10^{-25}$ s. This is one to two magnitudes shorter than the timescale of hadronization. Therefore the top does not exhibit bound states as the other quarks do.
Furthermore this short mean lifetime preserves the spin polarization of the top which affects the subsequent decay. Through precise analysis of the decay products and their kinematics, conclusions about the original polarization of the top can be made. The top exhibits the possibilities for such measurements especially well.
The experimental identification of top quarks is done through their decay products. The plethora of possible channels is introduced by the W boson decay whose dominant channels are [12]:

$$W \to e\nu_e (10.75\,\% \pm 0.13\,\%)$$
$$W \to \mu\nu_\mu (10.57\,\% \pm 0.15\,\%)$$
$$W \to \tau\nu_\tau (11.25\,\% \pm 0.20\,\%)$$
$$W \to \text{hadrons}(67.6\,\% \pm 0.27\,\%)$$

For top pairs whose first decay step is $t\bar{t} \to W^+W^-b\bar{b}$ there are three possible groups of channels according to the way the W bosons decay:

dileptonic In the dileptonic channel both W bosons decay to leptons. This channel provides a very clean signal due to the two leptons with high transverse momentum that may be used for triggering thus suppressing QCD background. Due to the two neutrinos present in the event the reconstruction of the top pair is very difficult and error prone. The branching ratio of the dileptonic channel is about 1/9.

semileptonic Here one W boson decays to leptons and the other decays to quarks. This channel offers good QCD suppression due to the high p_T lepton. Furthermore it exhibits only one neutrino which might be reconstructed from the missing transverse energy thus making a top pair reconstruction possible. Another advantage is the relatively high branching ratio of about 4/9.

[6] Negligible terms were dropped. G_F denotes the Fermi coupling constant.

$\sqrt{s}$ in TeV	1.96	7	14
σ in pb	$7.01^{+0.82}_{-1.03}$	$164.57^{+11.45}_{-15.78}$	854^{+129}_{-129}
source	[25]	[26]	[25]

Table 2.3: Cross-sections of top pair production at various $\sqrt{s}$.

fully hadronic This channel is characterized by both W bosons decaying to quarks. The lack of leptons makes it hard to distinguish this channel from QCD background resulting in the worst signal to background ratio of all three channels[7]. In principle the reconstruction of a top pair is possible but the isolation of the correct set of jets stemming from the top pair is hard due to a multitude of possible combinations. The channel has a branching ratio of about 4/9.

In this analysis the semileptonic channel was chosen because it presents the best compromise between very good signal to background ratio and top pair reconstruction. Apart from this the relatively high branching ratio is a nice feature. Its characteristics will be reviewed at the beginning of the analysis chapter.

Top Quark Production

Prior to its decay the top quark has to come into existence. Therefore understanding top quark production is a key to understanding top quark production rate at the LHC. Figure 2.1 [23] shows the cross-sections for different physics processes at the $\sqrt{s}$ of the Tevatron, the LHC at design-$\sqrt{s}$ and at its current $\sqrt{s}$ of 7 TeV (black dotted line).
It is noteworthy that the difference in the top production cross-section at the Tevatron and the LHC at design $\sqrt{s}$ spans two orders of magnitude.
There are two production channels of interest for top quark events, the top pair production and the single top production. As this thesis deals with top pairs only their production will be described in detail. It should be noted though that single top production provides a direct access to the measurement of the V_{tb} element of the CKM matrix making tests of the unitarity of the matrix possible. Further information and recent results can be found in [24].
The cross-sections of the top pair production at the various energies are given in Table 2.3. Though results at Next-to-Next-to-Leading order (NNLO) are known, all cited cross-sections are at NLO for consistency reasons. All calculations such as weighting of samples were done with the results at NLO.
Top pair production happens mainly through QCD processes. Unfortunately a purely perturbative approach is not possible when describing the production processes. Figure 2.2 depicts the whole process that lies behind a top pair event. Relevant for the discussion of the top pair production is the left side before the top quarks decay.

[7]This holds true for hadron colliders. At a lepton collider the background from QCD would be more manageable

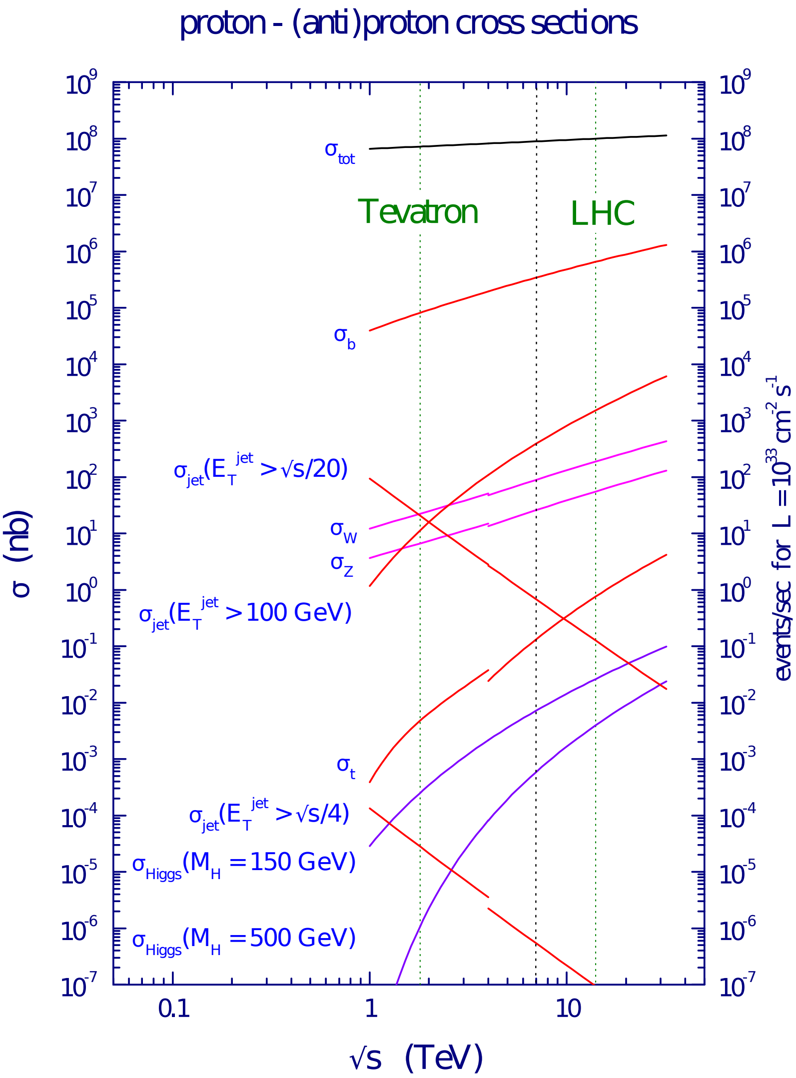

Figure 2.1: Cross-sections of physics processes at hadron colliders. [23]

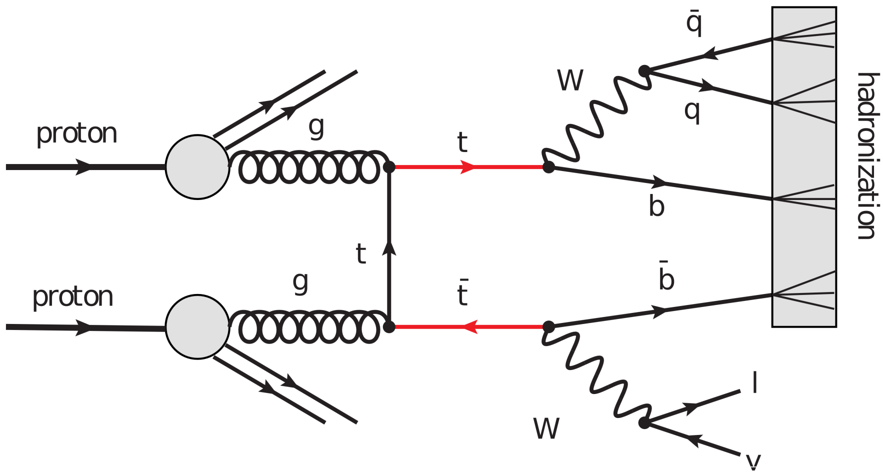

Figure 2.2: Top pair production at the LHC

The nature of protons is governed by their composite structure. Besides the three valence quarks (two up and one down quark) there exists a multitude of volatile quark antiquark pairs called sea-quarks and gluons. The momentum of the proton is the vectorial sum of the momenta of all these elementary particles. When the proton is accelerated to high energies all of its constituents travel in the approximately same direction (momenta transverse to the proton's direction of flight are considered negligible). Therefore a proton might be regarded as a beam of elementary particles with a wide spectrum of energies.

The collision of two protons is determined by the interactions of their partons (quarks and gluons). Such an interaction can be thought of as a hard scattering process of two partons from the protons. The momenta of the elementary particles inside a proton are fractions of its overall momentum, expressed as momentum fractions x.

The calculation of the top pair production cross-section is factorized into two components: the probability to find partons with momentum fraction x inside the respective protons and the cross-section of the actual interaction between the partons. The first part is expressed via parton distribution functions$f_i(x_i, \mu_{\mathrm{F}})$ where μ_{F} denotes the factorization scale. By studying deep inelastic scattering processes, these functions have been derived which give the probability to observe an elementary particle of the flavor i with a momentum fraction x_i. Figure 2.3 shows the parton distribution functions (PDF) for the various flavors found in a proton.

The second part is calculated using perturbative QCD. Due to the high momentum transfer and the therefore low coupling constant α_s a perturbative approach is possible. Depending on the physics process in question the cross-sections for the hard scattering processes are known to NLO or even higher orders.

The overall cross-section is then expressed via the formula [30]

$$\sigma^{t\bar{t}}(\sqrt{s}, m_{\mathrm{t}}) = \sum_{i,j=q,\bar{q},g} \int \mathrm{d}x_i \mathrm{d}x_j f_i(x_i, \mu_{\mathrm{F}}^2) f_j(x_j, \mu_{\mathrm{F}}^2) \cdot \hat{\sigma}^{ij\to t\bar{t}}(\rho, m_{\mathrm{t}}^2, x_i, x_j, \alpha_s(\mu_{\mathrm{R}}^2), \mu_{\mathrm{R}}^2). \tag{2.21}$$

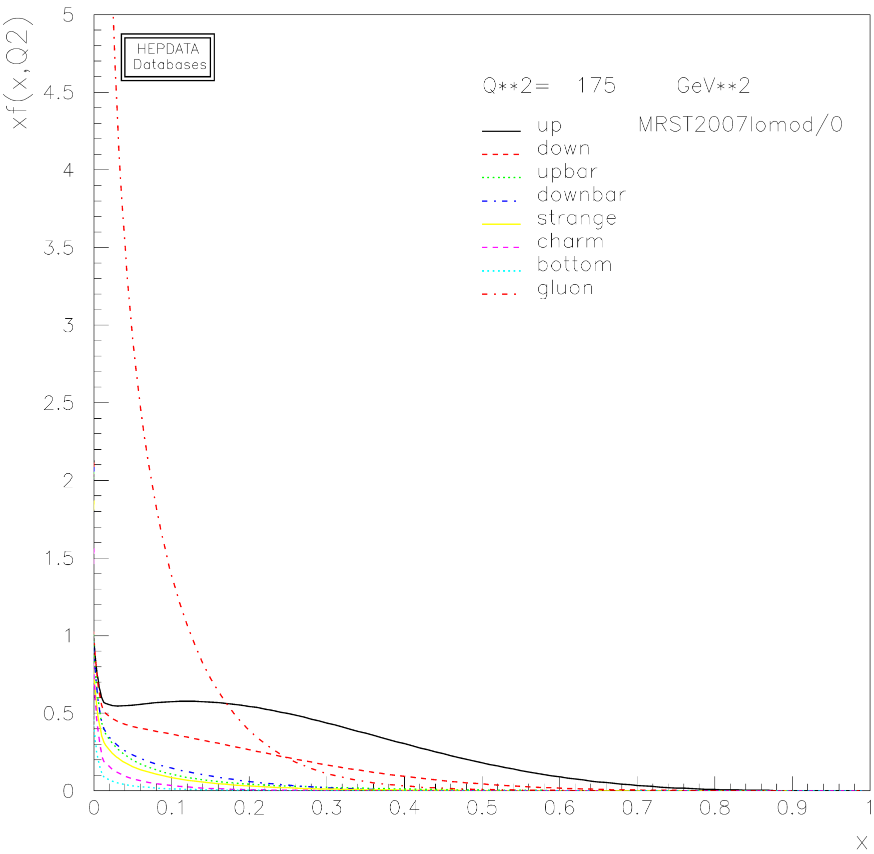

Figure 2.3: Parton Distribution Function (PDF) of the proton that was used in the generation of the signal sample of this analysis (AcerMC [27]) [28, 29].

where $\rho = 4 * m_{\mathrm{t}}^2/\sqrt{\hat{s}}$, $\hat{s} = x_i x_j \cdot s$ and μ_{R} is the renormalization scale. Examining the formula one can see that the production rate of a given physics process is determined both by the actual cross-section of the hard scattering and the probability that two partons with a sufficiently high momentum fraction can be found.
This may be illustrated as follows: The center of mass energy squared for the top pair production has to be at least $\hat{s} = x_i x_j \cdot s \geq 4m_t^2$. When assuming that both momentum fractions are roughly the same the formula

$$x \approx \frac{2m_{\mathrm{t}}}{\sqrt{s}} \tag{2.22}$$

defining the minimal momentum fraction that the particles have to possess for top pair production is obtained. The $\sqrt{s}$ of the Tevatron gives $x = 0.18$ while for the LHC the results are $x = 0.025$ for $\sqrt{s} = 14$ TeV and $x = 0.045$ for $\sqrt{s} = 7$ TeV.
This has two consequences. Overall the number of partons in the proton that lead to a valid initial state for the hard scattering process increases when x gets smaller. This by itself leads to higher overall cross-section.
What is even more noteworthy is that the gluon PDF rises dramatically with decreasing x. This shifts the ratio between the production channels $qq \to t\bar{t}$ and $gg \to t\bar{t}$. At the Tevatron top pair production via quark interaction is favored whereas the production processes at the LHC are clearly dominated by gluon-gluon fusion.

3 The LHC and the ATLAS Experiment

The LHC is a proton-proton collider at CERN near Geneva, Switzerland. It is a cyclic particle accelerator situated in a tunnel with a circumference of about 27 kilometers which lies on average 100 meters below the ground.
The reasons for building the LHC were manifold: discovery of the Higgs boson and measurement of its properties, studies concerning a possible substructure of quarks and leptons, finding evidence for the existence of supersymmetric particles, top quark studies with larger samples than produced by the Tevatron and the generation and examination of quark-gluon plasma to name just a few.
This ambitious program demanded an accelerator which is able to deliver extremely high beam energies. The $\sqrt{s}$ of the LHC was designed to be $\sqrt{s} = 14$ TeV, one magnitude higher than the Tevatron ($\sqrt{s} = 1.96$ TeV).
The construction works on the tunnel began in 2000 after the shutdown of LEP, the predecessor of the LHC, and lasted until 2008. On September, 10th, 2008 the first protons bunches (packets of protons) completed a full cycle in the LHC. A technical incident occurred on September, 19th, 2008 forcing the stop of all activities.
A year of repairs and modifications followed and the LHC brought protons to collision again on November 23rd, 2009. After initial tests, the new $\sqrt{s}$ was set to $\sqrt{s} = 7$ TeV in order to prevent further technical failure of this magnitude.
As of now, the LHC is operating reliably and with ever increasing instantaneous luminosity.
The accelerator provides beams for six experiments:

- ATLAS (A Toroidal LHC ApparatuS) [31]: ATLAS is a general purpose detector suitable for measuring a wide range of physics signals. Its components and performance will be described below since it is the experiment for which this thesis is written.

- CMS (Compact Muon Solenoid) [32]: CMS is the second general purpose detector and somewhat of a counterpart of ATLAS. Both detectors are supposed to measure the same physics contents but do this with slightly different technologies.

- LHCb (Large Hadron Collider beauty) [33]: As the other experiments, LHCb is supposed to find new particles but also to give answers to the question of the matter/antimatter imbalance in nature. The focus lies on b-quarks and their decays.

- ALICE (A Large Ion Collider Experiment) [34]: In heavy ion mode lead ions instead of protons are brought to collision. This is done in order to study a

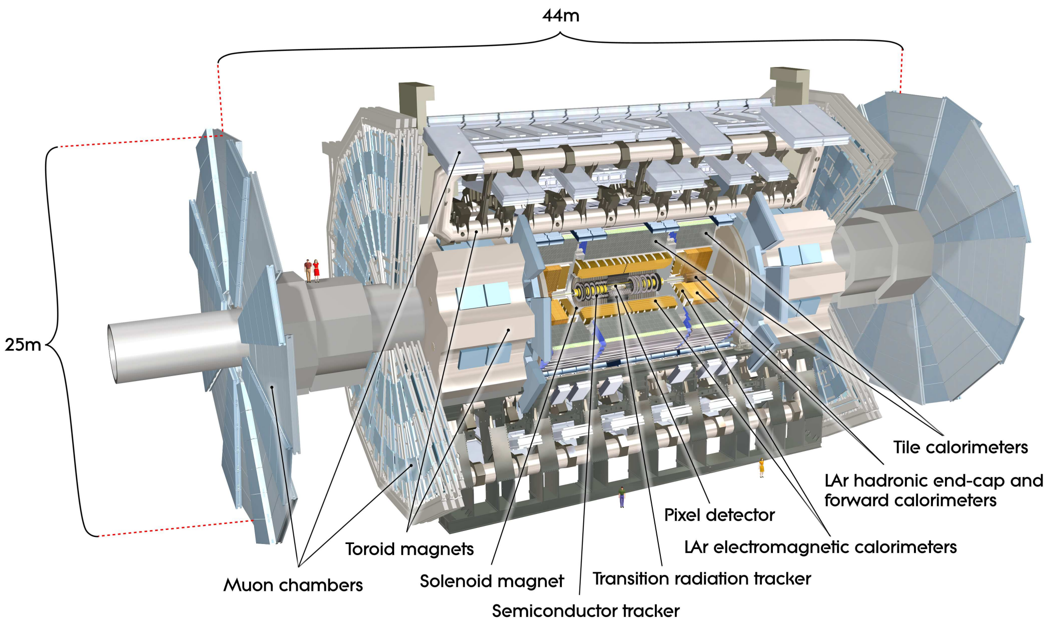

Figure 3.1: Cutaway view of the ATLAS detector

unique physical state, the quark-gluon plasma which is believed to represent the conditions in the early universe. Since lead ions are composite objects the challenge for ALICE lies in the reconstruction of events with a very high amount of objects.

- TOTEM (TOTal Elastic and diffractive cross section Measurement) [35]: TOTEM is dedicated to the measurement of the total cross-section of proton-proton interactions and the examination of the proton structure. It is located near the CMS detector.

- LHCf (Large Hadron Collider forward) [36]: LHCf studies the particles which travel closely to the beam line (the so-called forward direction). It is the smallest of the six experiments and is situated near ATLAS at Interaction Point 1 (IP1).

3.1 The ATLAS Detector

The ATLAS detector depicted in Figure 3.1 measures 44 meters in length and 25 meters in width as well as in height and the total weight of the detector is 7000 tons. It is situated at Interaction Point 1 and as a general purpose detector its task lies in the detection of a large variety of physics signals. This requires a list of favorable properties the detector should have:

- detector hermicity: the whole solid angle has to be equipped with detecting material

- precise tracking: important for the reconstruction of charged particles, measurements of momentum and advanced techniques such as b-tagging

- no/low leakage: particles (except neutrinos) should not be able to leave the detector but deposit all of their energy in it
- fine granularity: low uncertainties on the direction of flight of particles are important for reconstruction and distinction of objects
- high resolution: the energy information of objects should be as precise as possible (e. g. for the reconstruction of objects from their decay products)

How these properties are realized in ATLAS will be described in the following sections concerning the sub-detector systems. But first of all a short introduction to the coordinate system of ATLAS is necessary.

3.1.1 The Coordinate System of ATLAS

The coordinate system is defined as such:

- the origin $(0, 0, 0)$ is located at the nominal interaction point,
- the x-axis points to the middle of the accelerator ring,
- the y-axis points upwards,
- the z-axis completes the right handed coordinate system (points into the beam direction).

This Cartesian coordinate system is used for tracking information and location of vertices and the likes. When analyzing the direction of flight of objects such as leptons or jets the two dimensional system is used:

- the azimuthal angle ϕ. It is defined as

$$\phi = \arctan \frac{p_y}{p_x}. \tag{3.1}$$

- the polar angle θ, which is defined as

$$\theta = \arctan \frac{p_\mathrm{T}}{p_z}. \tag{3.2}$$

Here the subscripts x, y and z denote the projections of the momentum vector on the respective axis and T stands for transverse. The transverse momentum itself is calculated using the formula

$$p_\mathrm{T} = \sqrt{p_x^2 + p_y^2}. \tag{3.3}$$

In practice, θ is replaced by the pseudo-rapidity η. This quantity is defined as

$$\eta = -\log \tan \frac{\theta}{2}. \tag{3.4}$$

If one wants to calculate a distance between two objects in the η-ϕ-plane, the formula

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{3.5}$$

can be applied.
With these formulas in mind a more precise evaluation of the sub-detector systems is possible.

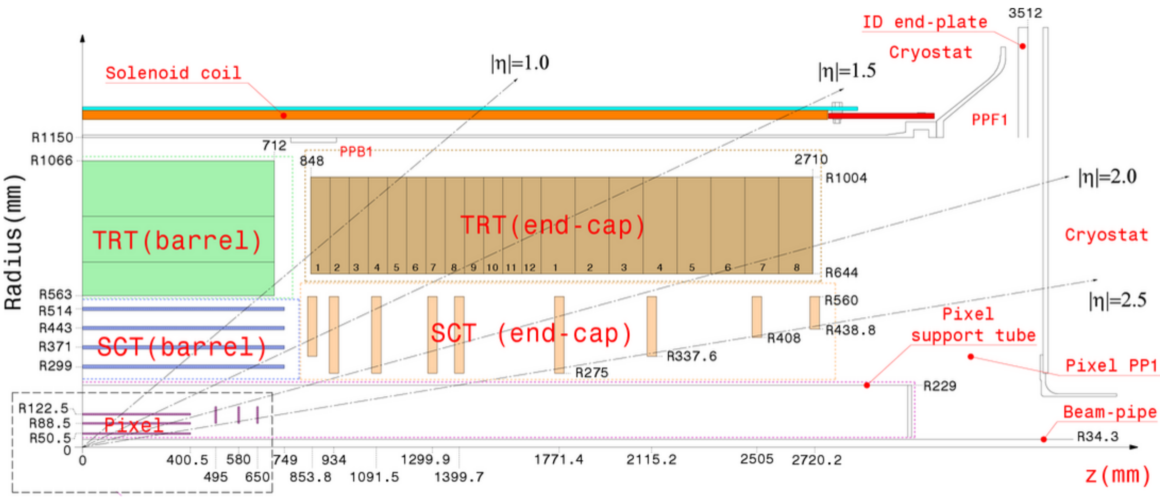

Figure 3.2: Geometry of the Inner Detector. The sizes of the different components and their positions are given with respect to the z-axis and as radii with the beam as center.

3.1.2 The Inner Detector

As shown in Figure 3.1 the Inner Detector (ID) is the first sub-detector system a particle coming from the collision point passes. Its main purpose is the precise localization of tracks and vertices (points where tracks fork). A solenoid magnet provides a magnetic field of 2 T where the ID is located causing charged particles to exhibit tracks which are bent. The curvature of these bent tracks is used for calculating the momentum of the charged particles.
All sub-detectors that make up the ID have two regions; a barrel region where the detector modules are positioned on the lateral surface of a thought cylinder around the beam axis and an end-cap region where they are arranged on circular rings perpendicular to the beam axis. The regions exhibit different arrangements of the detecting elements and different performances regarding the spatial resolution.

The Pixel Detector

Figure 3.2 depicts the innermost part of the ID, the pixel detector. It spans the whole azimuthal angle ϕ and the pseudo-rapidity $|\eta| < 2.5$. The pixel detector is a semiconductor detector and consists of about 140 million pixels with a size of $50\,\mu m \times 400\,\mu m$. This leads to a precision of about $14\,\mu m$ when measuring tracks.

The Semi Conductor Tracker

The next part of the ID is the Semi Conductor Tracker (SCT). It spans the same region in the η-ϕ-plane as the pixel detector. The four layers in the barrel region contain strip sets (stereo-strips) made of two silicon strips which run at an angle of 40 mrad to each other. In the barrel region, these stereo-strips run parallel to the beam axis and in the end cap region they are arranged radially. The mean pitch of the individual strips is about $80\,\mu m$ and they are several centimeters long. The

SCT consists of 4088 detector modules with 768 readout strips per module. When measuring positions the SCT exhibits a precision of 17 µm(R-ϕ) and 580 µm(z) in the barrel region and 17 µm(R-ϕ) and 580 µm(R) in the end-cap region.

The Transition Radiation Tracker

The outermost part of the ID is the Transition Radiation Tracker (TRT). The covered region is the whole azimuthal angle ϕ and the pseudo-rapidity $|\eta| < 2.0$. It consists of 4 mm wide tubes with a wire in the center. The tubes are filled with a gas mixture (70% Xe, 27% CO_2 and 3% O_2). In the barrel region the tubes run parallel to the beam axis and radially in the end-cap region. The length of the tubes (144 cm in the barrel region, 37 cm in the end-cap region) does not permit a precise measurement of the z or R coordinates, only ϕ is measured. This is made up for by a high number of track hits of typically 30 hits per track. Overall the accuracy of the ϕ-measurement is 130 µm.
When a particle passes through the tube it ionizes the gas leading to a measurable current between the wall of the tube and the wire in the center. In addition highly relativistic particles emit measurable amounts of photons when traversing radiators located between the tubes. These radiators are made off foils of differing electrical permittivity. When a particle crosses the interfaces of the two foils it emits a certain amount of photon depending on its velocity. The photons that make up this transition radiation are absorbed by the xenon in the gas mixture and produce a significantly larger signal in the tubes.
As electrons exhibit on average higher velocities than other (heavier) particles they will also cause more transition radiation. This provides an additional information for identifying electrons.

3.1.3 The Calorimeters

Following the Inner Detector are the calorimeters of the ATLAS detector, namely the Electromagnetic Calorimeter and the Hadronic Calorimeter. Their purpose lies in the measurement of the energy of the particles coming from the collision. Additionally they have to supply information about the direction of flight of the particles and their granularity should be fine enough to separate objects that are close to each other. This generates a list of desirable properties:

- rapidity coverage: A coverage in η from -5 to 5 is desirable in order to detect particles with a small angle to the beam.
- thickness: The calorimeters have to be thick enough to stop the passing particles in order to prevent loss of energy information.
- high reconstruction efficiency: Objects have to be reliably reconstructed in a large energy range from some GeV to several TeV.
- fine granularity: In order to achieve a separation of objects in crowded events, the calorimeters have to provide precise spatial information.

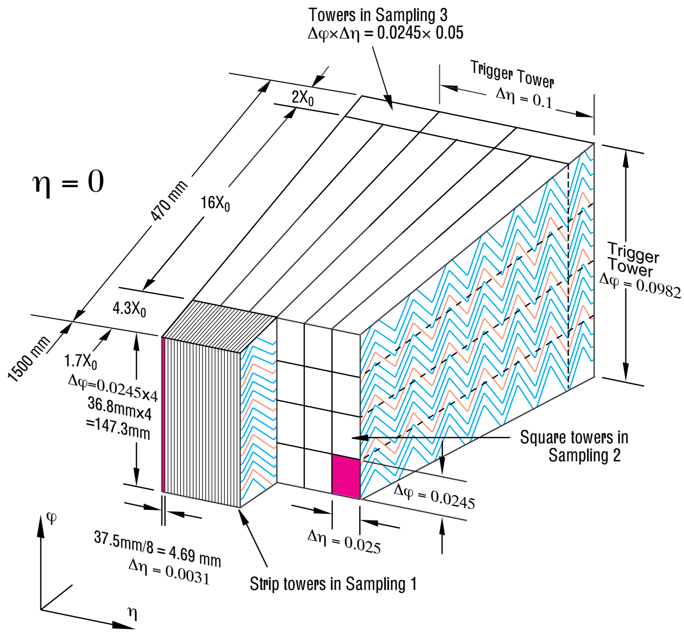

Figure 3.3: Schematics of the Electromagnetic Calorimeter in the barrel region

- particle identification: The response from the calorimeter has to be unique for different particle types in order to identify them.
- fast response: With higher instantaneous luminosity there will be several collisions per bunch crossing and therefore the detector response has to be fast.

The Electromagnetic Calorimeter

The Electromagnetic Calorimeter is primarily used for measuring the energy of photons and electrons. Since photons have no charge the Electromagnetic Calorimeter has to provide the direction of flight information for these particles. In the case of electrons the calorimeter and the Inner Detector provide more precise information on the direction of flight measurement.
It is divided into a barrel part which spans the $|\eta| < 1.475$ region end-cap region going from $1.375 < |\eta| < 2.5$. The basic setup is shown in Figure 3.3.
The EM Calorimeter utilizes lead as a showering material and liquid argon as active material. The two materials are sandwiched in an accordion style to ensure complete ϕ-symmetry without cracks. The thickness of the lead absorbers varies with pseudorapidity in order to optimize energy resolution.
As an electron or photon traverses a lead sheet it loses energy due to a showering process where electrons emit photons and photons split into electron-positron pairs. The particles in the shower cause ionization in the liquid argon volumes and the produced ions are collected by electrodes. This signal is then translated into an energy information. A total thickness of about 24 X_0[1] in the barrel region and 26

[1] X_0 denotes radiation length. It is the mean distance after which an electron loses $1/e$ of its energy due to bremsstrahlung or 7/9 of the mean free path for pair production by a photon.

X_0 in the end-caps ensures that most of the electrons and photons will deposit their whole energy in the electromagnetic calorimeter. In the region of $|\eta| < 2.5$ (barrel region and outer end-cap wheel) the Electromagnetic Calorimeter is segmented in three samplings as shown in Figure 3.3. The first sampling acts as a "pre-shower" detector that provides particle identification information. It consists of thin strips with a pitch of about 4 mm in η enabling precise measurements in this direction. It is followed by the second sampling which is segmented into square towers. The single towers have a base area of $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ which corresponds to $16\,\mathrm{cm}^2$ at $\eta = 0$. The third sampling is built like the second sampling but the single towers have a base area of $\Delta\eta \times \Delta\phi = 0.05 \times 0.025$ resulting in a coarser granularity than that of the second sampling.
For the region of $|\eta| > 2.5$ (the inner wheel of the end cap region) the Electromagnetic Calorimeter is arranged in only two samplings with a coarser granularity than in the barrel region. Therefore precision physics are mainly focussing on the region with $|\eta| < 2.5$ but the end cap calorimeter performance is sufficient for jet reconstruction and missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) measurements.

The Hadronic Calorimeter

The Hadronic Calorimeter consists of the Tile Calorimeter ($|\eta| < 1.7$) and the Hadronic Liquid Argon Calorimeter ($1.5 < |\eta| < 3.2$).
The setup of the Tile Calorimeter is similar to that of the Electromagnetic Calorimeter. Here the showering material is steel whereas the active material is a plastic scintillator. The particles from the shower cause the emission of light in the scintillator which is then transported by fibers to photomultipliers. There the signal is extracted and provides the energy information.
In contrast to the Electromagnetic Calorimeter the Tile Calorimeter does not have an accordion structure. The 3 mm thick steel tiles and the scintillator tiles form stacks in the z-direction. These stacks are then put on top of each other to form towers that stand perpendicular to the beam axis. As the petals of a flower a set of these towers form a ring around the beam axis.
The granularity of the tile calorimeter is $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ for the first two samplings and $\Delta\eta \times \Delta\phi = 0.2 \times 0.1$ for the third. This is significantly coarser than the electromagnetic calorimeter but expected to be sufficient for separating jets coming from the decays of W bosons with high transverse momentum.
The determination of the top quarks and the calculation of their properties use data from four jets, therefore the energy resolution of the hadronic calorimeter is an important quantity. The design goal was an uncertainty of[2]

$$\frac{\Delta E}{E} = \frac{50\%}{\sqrt{E}} \oplus 3\% \quad \text{for} \quad |\eta| < 3. \tag{3.6}$$

In addition the jets should deposit all of their energy in the calorimeter which is ensured by a thickness of ten hadronic interaction lengths.

[2]The *oplus* denotes that the individual values are summed quadratically.

3.1.4 The Muon Spectrometer

The Muon Spectrometer is used for measuring the direction of flight of muons and their momenta. It covers the region of $|\eta| < 2.7$ and employs four different types of chambers: Monitored Drift Chambers (MDT), Cathode-Strip Chambers (CSC), Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC). This multitude of technologies is made necessary by differing demands in the respective regions.

3.1.5 The Trigger System

When the LHC reaches its design luminosity there will be bunch crossings every 25 ns which corresponds to a rate of 40 MHz. Only a few of the collisions that occur are interesting physics processes which have to be separated from the background. Additionally it is technically not possible to record all events at this rate. Therefore a system is needed to reduce the event rate to a manageable size: the trigger system. It consists of three levels: Level-1 (L1), Level-2 (L2), and the event filter (EF). The L1 trigger is a hardware trigger that evaluates the response of the calorimeters and the muon spectrometer and is synchronized with the bunch crossing frequency of the LHC. All detector data are kept in pipelines for the timespan of the L1 Trigger decision making process (about 2.5 μs). If the L1 trigger fires regions of interest (RoI) associated with the objects that caused the positive trigger decision are formed and the readout of the detector data to a distributed buffer system is initiated.
The L2 trigger takes the RoIs provided by the L1 trigger and evaluates the data from the full detector inside these regions. This trigger is a computing cluster and the algorithms that steer the decision of the L2 trigger are software based. The L2 trigger further reduces the rate from 75 kHz coming from the L1 trigger to 3.5 kHz.
When the L2 trigger fires the whole detector readout is used and a fast event reconstruction takes place. These events are evaluated by the EF. As the L2 trigger the EF is based on software that is executed on a computing cluster. In the case that the EF accepts the event it is stored for further analysis with a rate of about 200 Hz.

3.1.6 Data Quality

As every complex machinery ATLAS sometimes experiences technical difficulties. The dysfunction of a certain detector element may spoil the data and introduce artifacts in the analysis. To avoid this the detector systems and their sub-detectors are permanently monitored.
Whether a given block of data meets the quality demands depends on a set of factors. First of all it has to be decided whether the availability of a certain sub-detector is crucial to the analysis e. g. an analysis that only concentrates on the central η region might not be concerned when problems in the end caps arise.
After defining the set of sub-detectors crucial to the analysis these sub-detectors have to be checked. This is done with a set of flags. As with traffic lights, green represents for nominal function, yellow is a transient state where a decision on the data quality is pending and red means that the given sub-detector is not functioning as intended. The flags are organized tree-like manner, branches describe the different detectors and their sub-branches describe the sub-detectors they are made of.

The blocks of reprocessed data that meet the quality demands are charted in so-called good run lists. When analyzing data these good run lists determine whether a given event is accepted from the data quality point of view.

4 Reconstruction of Physics Objects

Detecting particles is a problem of translation. When a particle passes through the detector it causes is a collection of electric signals. The responses that it may cause in the Inner Detector and the Muon Spectrometer (called hits) are reconstructed to tracks and the signals from the calorimeter are translated to energy deposits. These tracks and energy deposits are the input for reconstruction algorithms which return physical objects with certain properties.
The process of measuring is not an one-to-one projection. Particles with different properties may cause a similar detector response. A reconstruction algorithm for a certain particle type should only accept signals caused by real particles of that type. Furthermore the properties of the reconstructed physical objects should match those of the particle causing the signal.
In reality reconstruction algorithms are not able to reconstruct all of the particles (corresponding to an efficiency smaller one) and they might use detector responses from other particle types and translate them into physical objects (mis-identification). The art of inventing a reconstruction algorithm lies in maximizing its efficiency while keeping the mis-identification rate low. Additionally robustness and simplicity are desirable.

4.1 Jet Reconstruction Algorithms

A typical name for a jet container is "EMJESAntiKt4Topo". The name denotes which clustering algorithm is used (either Tower or Topo [37]), how the jets are reconstructed from clusters (Cone or Kt [38,39]) and which energy calibration scheme is applied (H1, EMJES, GCW etc. [40–42]).

4.1.1 EMJESAntikt4Topo jets

The mentioned jet container (a collection of physical objects sharing a common definition) is the one used in this analysis and the specific choices for the individual reconstruction steps defining it will be discussed in more detail. Information on the alternative choices may be found in the listed references.

Topological Clustering

The clustering begins with the translation of the electric signal in a calorimeter cell into an energy information. Various factors that affect this translation such as issues

in the high voltage supply of the cell are taken into account. The resulting objects are called CaloCells which will be grouped to CaloClusters.
In the case of Topological Clustering [37], the algorithm searches for seed cells. These cells have a signal-to-noise ratio [1] that exceeds a high threshold t_{seed}. Each seed cell constitutes a "proto cluster". In the next step, the cells adjacent to the seed cells are evaluated, starting with the neighbors of the seed cell with the highest signal-to-noise ratio. If a neighbor cell has not been used as a seed so far and its signal-to-noise ratio lies above a certain threshold t_{neighbor}, it is added to the proto cluster. Two proto clusters are merged if they share a neighbor cell satisfying the t_{neighbor} threshold. If such a cell only satisfies a lower threshold t_{cell} it is added to the proto cluster which is the more significant neighbor of the cell. The original seed cell list is discarded and the cells added to proto clusters in this step form the new seed list. This is repeated until an empty seed list is returned.
In a last step all proto clusters satisfying a minimum energy requirement are saved as clusters and written into a list in descending order with regard to their energy.

The Anti-k_t Algorithm

The anti-k_t algorithm [39] belongs to a larger class of algorithms employing a metric for jet reconstruction. The general form of this metric is

$$\begin{aligned} d_{ij} &= \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2}, \\ d_{iB} &= k_{ti}^{2p}. \end{aligned} \tag{4.1}$$

where k_t stands for the transverse momentum of a particle, Δ_{ij}^2 is defined as

$$\Delta_{ij}^2 := (y_i - y_j)^2 + (\phi_i - \phi_j)^2 \tag{4.2}$$

and a factor R which determines how large a jet may become.[2] The factor p has a crucial influence on the behavior of the metric and for the anti-k_t algorithm it is set to $p = -1$.[3]
Similar to the clustering algorithm the reconstruction is performed in a loop. In a first step the distances d_{ij} between entities (particles or pseudo jets) and between the entities and the beam ("B") are calculated. The lowest distance is taken. If the lowest distance[4] is between an entity and the beam the entity is called a jet and is removed from the list of entities. If the lowest distance is between two entities they are merged (which produces pseudo jets). Then the first step is repeated until the list of entities is empty.
In this algorithm, entities with a high transverse momentum collect entities with a lower momentum in their vicinity (due to the small value of k_t^{-2}).[5] The factor Δ_{ij}^2/R^2

[1] Here the signal may be the cell energy while the noise is the expected root mean square (RMS) of electronics noise under given conditions.
[2] Here the rapidity y is used which is defined as $\frac{1}{2}\ln\frac{E+p_z}{E-p_z}$.
[3] Another choice is $p = 1$ (called k_t or Cambridge/Aachen).
[4] These are not distances in space but rather distances in the k_t-y-ϕ phase space.
[5] In comparison: for $p = 1$ the low momentum particles will cluster among themselves before high momentum particles are considered.

is of special interest. For the entity with the highest k_t this effectively sets the radius of the jet since d_{iB} will always be smaller than any d_{ij} for this entity if this factor is larger than 1. Therefore these entities will be labelled as jets and removed from the list of entities if there are no particles left in their vicinity.
One virtue of the anti-k_t algorithm is that it mostly produces cone shaped jets. There are some cases where two high momentum entities are separated by $R < \Delta_{ij} < 2R$. The resulting shapes of the jets will not necessarily be conical, since the jets will divide the entities in the overlap region among themselves.

Jet Calibration via Numerical Inversion

The last step before the jet reconstruction is completed is the energy calibration. This is necessary because the translation from detector signals to energy is not known for jets but for objects such as electrons. Therefore the energy of a jet is measured at the so-called "electromagnetic scale" and is then calibrated to the jet scale. The EM+JES calibration [41] is a straight forward and robust calibration scheme that is used for early data taking.
The calibration results in a response function $R(p_{\mathrm{T}}^{\mathrm{reco}}) = p_{\mathrm{T}}^{\mathrm{reco}}/p_{\mathrm{T}}^{\mathrm{true}}$ which maps the $p_{\mathrm{T}}^{\mathrm{reco}}$ to the real transverse momentum of a given jet via the formula

$$p_{\mathrm{T}}^{\mathrm{true}} = \frac{p_{\mathrm{T}}^{\mathrm{reco}}}{R(p_{\mathrm{T}}^{\mathrm{reco}})}. \tag{4.3}$$

The response function is derived from MC simulations using a technique called numerical inversion. Jets on detector-level[6] and on hadron-level[7] reconstructed by the same jet finding algorithm are taken and matched to each other requiring that the distance ΔR between a reconstructed jet and a hadron-level jet is less than 0.3.
Hadron-level jets in a small $p_{\mathrm{T}}^{\mathrm{truth}}$ slice are selected and the $p_{\mathrm{T}}^{\mathrm{reco}}$ of their matched detector-level jets are plotted providing gaussian distributions of $p_{\mathrm{T}}^{\mathrm{reco}}$ for the $p_{\mathrm{T}}^{\mathrm{truth}}$ slices. The average of these distributions is taken as an estimate for $p_{\mathrm{T}}^{\mathrm{reco}}$. By applying this procedure a mapping between $p_{\mathrm{T}}^{\mathrm{reco}}$ and $p_{\mathrm{T}}^{\mathrm{truth}}$ is obtained which is then used to calculate the response function.
In practice the response function is sliced in η as well as in jet p_{T}.

4.1.2 B-Tagging

Many interesting physics processes contain bottom quarks in their final state. Therefore techniques that are able to separate jets originating from bottom quarks from those coming from lighter quarks are an important ingredient for analyses of such processes. It is desired that these techniques provide tags for jets from bottom quarks with a high efficiency while rejecting jets from light quarks at the same time.
Jets that originate from a bottom quark have properties that make them distinguishable from jets stemming from lighter quarks or gluons[8]. When the bottom quark

[6] These are the jets reconstructed from the detector response.

[7] The constituents of these jets are the particles that come from the MC generator, thus omitting the simulation of the detector.

[8] In the following, jets originating from bottom quarks will be called bottom jets whereas jets from lighter quarks (up, down, strange) and gluons will be called light jets. If not mentioned otherwise

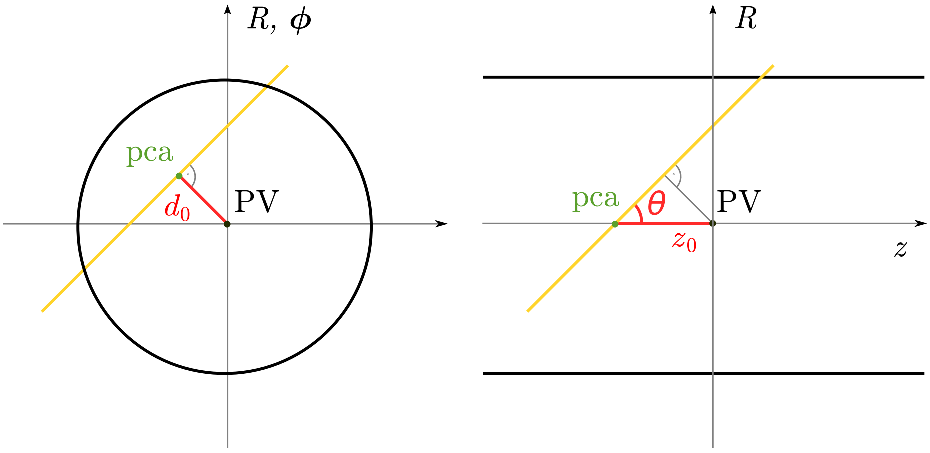

Figure 4.1: Definition of transverse impact parameter d_0 (left) and the longitudinal impact parameter z_0 (right). [44]

hadronizes to a b-hadron, the hadron keeps about 70 % of the original bottom quark momentum. These b-hadrons exhibit comparatively large masses. Decay products of these hadrons exhibit larger transverse momenta with respect to the jet axis than decay products inside light jets. Therefore the angle between the direction of flight of these decay products and the jet axis is larger. The most important property of the b-hadrons is their comparatively long lifetime lying in the order of pico-seconds. In conjunction with the time dilation resulting from the high velocity of the hadron in a jet with high transverse momentum this leads to flight path lengths in the order of millimeters. The tracks from the decay products of the b-hadron exhibit a displacement with respect to the primary vertex that is expressed via impact parameters. Alternatively the vertex formed by the tracks of the decay products of the b-hadron might be reconstructed and used for evaluation [43].

An appropriate spatial description is necessary for a correct implementation of these tagging strategies.

The transverse impact parameter d_0 is the smallest distance between the reconstructed primary vertex and a given track in the transverse plane (left side in Figure 4.1). This point on the track is called the point of closest approach (pca). Now the view is changed to the longitudinal plane (right side in Figure 4.1). A rotation of the detector around the beam axis is carried out until the pca lies on the z-axis. The measured distance along the z-axis between the pca and the primary vertex is called z_0 (the longitudinal impact parameter). Further information can be extracted by examining the angle between the track and the z-axis (Θ). The smallest distance between the track and the primary vertex in the longitudinal plane can be calculated as $z_0 \times \sin(\Theta)$.

Using these definitions the tagging algorithms compute weights for the each jet. In general these weights are higher for bottom jets than for light jets. The user may apply a cut on these so-called b-tagging weights thus separating the light jets from the bottom jets. If a jet is tagged as a bottom jet depends on its properties and the used b-tagging algorithm.

jets from charm quarks will count as light jets. In all other cases they will be called charm jets.

The b-tagging algorithms that will be described here use the secondary vertex reconstruction as well as measured impact parameters. At first the JetProb tagging algorithm will be discussed as it is the one used in this analysis for b-tagging purposes. Its calibration with 2010 data is described in [45]. The SV0 tagging algorithm which provides additional information about the jets will also be presented.
The IP3D + SV1 tagging algorithm is a more sophisticated technique for identifying bottom jets. It will be used once enough data has been accumulated to carry out its calibration.

The JetProb Tagging Algorithm

The JetProb algorithm was first used by the ALEPH collaboration and is described in [46]. It was also used in Tevatron analyses and is one of the b-tagging algorithms for early physics analyses at the LHC because its calibration is possible with a small amount of statistics.
The JetProb algorithm exploits the long lifetime of b-hadrons. It evaluates all tracks of the jet in question with respect to the probability that these tracks originate from the primary vertex. The combined probability of the tracks then gives a weight for the jet discriminating between bottom jets and light jets.
The inputs which are used for the track evaluation is the signed impact parameter significance d_0/σ_{d_0} which is shown in Figure 4.2 and a resolution function for prompt tracks $\mathcal{R}$ that has to be taken from data. The probability $\mathcal{P}$ for a given track i of a jet to be originating from the primary vertex is calculated as

$$\mathcal{P}_i = \int_{-\infty}^{-|d_0^i/\sigma_{d_0}^i|} \mathcal{R}(x)\mathrm{d}x. \tag{4.4}$$

The overall jet probability of a jet with N tracks is calculated via the formula

$$\mathcal{P}_{\mathrm{jet}} = \mathcal{P}_0 \sum_{j=0}^{N-1} \frac{(-\ln \mathcal{P}_0)^j}{j!} \tag{4.5}$$

where $\mathcal{P}_0$ is defined as

$$\mathcal{P}_0 = \prod_{i=1}^{N} \mathcal{P}_i' \quad \text{with} \begin{cases} \mathcal{P}_i' = \frac{\mathcal{P}_i}{2} & \text{if} \quad d_0^i > 0, \\ \mathcal{P}_i' = \left(1 - \frac{\mathcal{P}_i}{2}\right) & \text{if} \quad d_0^i < 0. \end{cases} . \tag{4.6}$$

The resulting probability for the jet may be interpreted as the compatibility of the jet with the hypothesis that it is a light jet. The smaller the value of the probability the more likely the jet stems from a b-hadron. Therefore one would define a maximum value and jets that exhibit a probability below this value are tagged as bottom jets.
All other b-tagging algorithms exhibit a behavior where a jet is called a bottom jet if it has a weight above a certain user-set threshold[9]. Additionally the algorithms have cut values that lie in the one digit area whereas the probability is a fractional value.

[9] These thresholds are determined by requiring a desired efficiency of finding bottom jets and that light jets are rejected with a certain rate

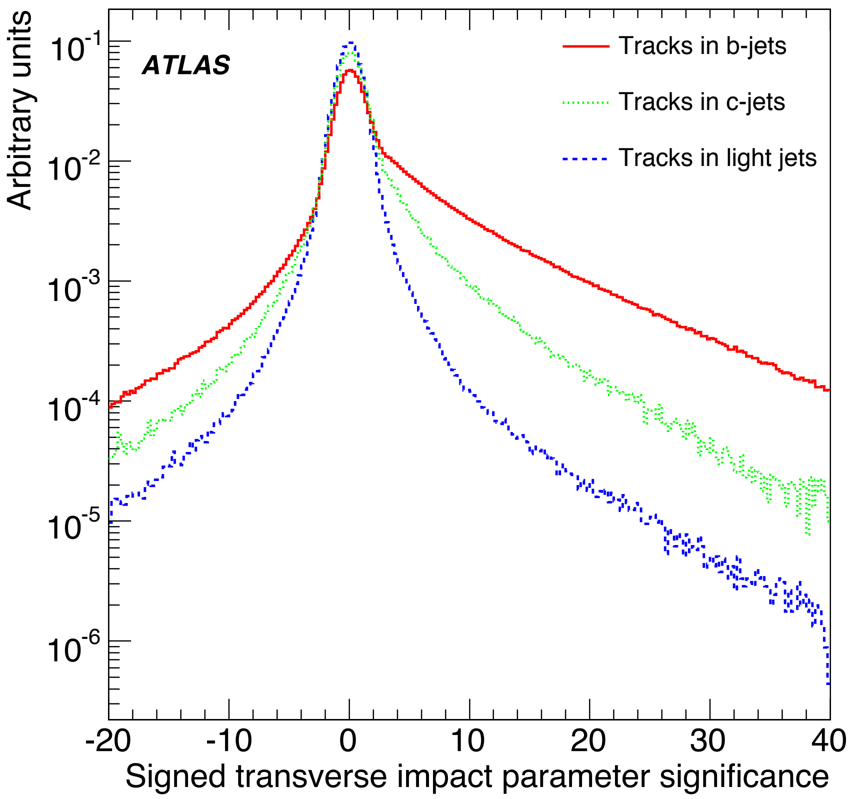

Figure 4.2: Signed impact parameter significance for tracks in jets of different flavors. While the distribution is symmetrical for track in light jets it is asymmetric for tracks from both charm and bottom jets where positive impact parameter significances are observed more often. [40]

p_{T}	> 0.5 GeV
d_0^{PrimVtx}	$< 2\,\mathrm{mm}$
$z_0^{\mathrm{PrimVtx}} \sin\theta$	$< 2\,\mathrm{mm}$
$\sigma(d_0^{\mathrm{PrimVtx}})$	$< 1\,\mathrm{mm}$
$\sigma(z_0^{\mathrm{PrimVtx}})$	$< 5\,\mathrm{mm}$
χ^2/ndof	< 3
Number of Pixel hits	≥ 2
Number of SCT hits	≥ 4
Number of Pixel + SCT hits	≥ 7

Table 4.1: Track requirements of SV0 tagging algorithm as defined in [47].

With the calculation of a weight according to

$$w_{\mathrm{JetProb}} = -\log_{10}\mathcal{P}_{\mathrm{jet}} \tag{4.7}$$

one obtains a quantity that has the desired properties. This modified weight defintion is the one actually used in analyses.

The SV0 Tagging Algorithm

The SV0 tagging algorithm explicitly reconstructs vertices from the decay products of the b-hadrons in a jet. These vertices are also called secondary vertices. It employs a cut on the decay length significance $L/\sigma(L)$ of such a reconstructed secondary vertex, exploiting the long lifetime of b-hadrons.

The input for this tagging algorithm is a list of tracks associated to the calorimeter jet that is evaluated. The tracks in this list satisfy the constraint that the distance ΔR between a given track and the jet axis is less than 0.4. If a track satisfies this constraint for multiple jets it is associated to the closest one.

In order to be used in the secondary vertex reconstruction these tracks have to pass certain constraints which are detailed in table 4.1.

Additionally the tracks have to have a three-dimensional impact parameter significance $dca/\sigma(dca) > 2.3$ with respect to the primary vertex.

After the track selection the SV0 algorithm begins reconstructing two-track vertices that exhibit a significant displacement from the primary vertex. For this, the sum of the impact parameters of the two tracks of a vertex candidate has to exceed 6.6. The fit of the two-track vertices has to yield a $\chi^2 < 4.5$. To make sure, that the vertex is not a valid primary vertex the χ^2 of the 3-dimensional distance between the primary vertex and the two-track vertex has to be larger than 6.25.

All two-track vertices with a mass corresponding to K_s^0 mesons, Λ^0 baryons or photon conversions and vertices that might stem from material interactions are removed.

In a last step all the two-track vertices are merged into one secondary vertex that is associated to the evaluated jet.

The IP3D + SV1 Tagging Algorithm

The IP3D + SV1 tagging algorithm [40] combines the output of two separate b-tagging algorithms. Both use a likelihood ratio approach that requires the distributions of discriminative variables for jets from light and bottom quarks respectively.
The IP3D tagging algorithm uses two-dimensional distributions of the transverse impact parameter versus the longitudinal impact parameter obtained for light jets and bottom jets. The tracks in a given jet are evaluated by measuring their impact parameters (z_0 and d_0). These are fed into the histograms yielding probabilities for the light jet and bottom jet hypothesis. Subsequently a jet weight is calculated according to the formula

$$w_{\text{track}} = \frac{b(z_0^{\text{track}}, d_0^{\text{track}})}{u(z_0^{\text{track}}, d_0^{\text{track}})} \tag{4.8}$$

where u is the value coming from the histogram with the distribution for light jets and b the histogram for bottom jets. The b-tagging weight of the jet is the sum of the logarithms of the individual track weights

$$w_{\text{jet}} = \sum_{i=1}^{N_{\text{Tracks}}} \ln W_i = \sum_{i=1}^{N_{\text{Tracks}}} \ln \frac{b(z_0^i, d_0^i)}{u(z_0^i, d_0^i)}. \tag{4.9}$$

The SV1 tagging algorithm reconstructs secondary vertices similar to the SV0 tagging algorithm. It uses a two-dimensional histogram of the invariant mass of all tracks associated to the secondary vertex versus the ratio between the sum of the energies of the tracks participating to the vertex and the sum of the energies of all tracks in the jet and a 1D histogram of the number of two-track vertices in the jet.
As in the IP3D tagging algorithm these histograms exist for the light jet hypothesis and the bottom jet hypothesis. The weight is again a likelihood ratio.
The individual weights of the two tagging algorithms are combined by simply summing them.

4.2 Electron Reconstruction Algorithms

The electron reconstruction is a two-step process. First energy clusters are formed and a search for an inner detector track which may be successfully extrapolated to the energy cluster takes place.[10] Following this step an identification scheme is applied providing a quality label for the electrons that may be used by the analyst.
The cluster finding uses a sliding window algorithm. Here a window of fixed size (3×5 cells) is moved across the η-ϕ plane evaluating energy information of the second sampling of the Electromagnetic Calorimeter. The clusters found in this step are the seeds for a window algorithm that uses larger clusters (3×7 cells in the η-ϕ plane in the barrel region and 5×5 cells in the end cap). Tracks from the Inner Detector are now matched to the barycenters of the energy clusters. If the matching is successful an electron candidate is found.

[10] This only applies to electrons in the region $|\eta| < 2.5$. For electrons outside this region there is no Inner Detector information available, these are called forward electrons.

The electron identification that follows the reconstruction defines several levels (such as loose, medium and tight) of electron quality. This is necessary since several other particles such as pions may cause similar signals in the detector.
The loose cuts demand certain qualities regarding the signal in the calorimeters. The middle and tight cuts include the loose cuts but put further constraints on the track that is associated to the energy cluster. Identification cuts are being constantly reviewed and optimized due to the the amount of gathered data and better understanding of the detectors.

4.3 Muon Reconstruction Algorithms

Muons are part of many interesting physics processes. Therefore they have to be reconstructed with high efficiency.
There exist two families of reconstruction algorithms for muons: STACO [48] and Muid [49]. The differences between these families lie in details, the general reconstruction schemes are the same.
Since there exist two sub-detectors (Inner Detector and Muon Spectrometer) which are able to record tracks from muons there is a variety of muon reconstruction schemes. There are standalone muons which take their information solely from the data of the Muon Spectrometer. So-called Combined Muons are reconstructed using information from both the Muon Spectrometer and the Inner Detector. Muons with low transverse momentum may be reconstructed using only Inner Detector information.
The muons used in this study are combined muons as reconstructed by Muid. The algorithm takes inner detector tracks and muon spectrometer tracks and combines them via a χ^2 minimization scheme. It incorporates detector response functions and accounts for possible scattering of the muon between the inner detector and the muon spectrometer to give realistic results.
For muons with an energy above 20 GeV the reconstruction efficiency is about 95 % − 98 % depending on the used algorithm and the muon kinematics [50].

4.4 Missing Transverse Energy ($E_\mathrm{T}^\mathrm{miss}$)

The Missing Transverse Energy $E_\mathrm{T}^\mathrm{miss}$ (abbreviated $\not{E}$ where convenient) is a vector that lies in the transverse plane. Its components $\not{E}_x$ and $\not{E}_y$ represent energy imbalances that exist in the respective direction. These imbalances are assumed to be caused by particles that escape detection, e. g. neutrinos. Therefore the $E_\mathrm{T}^\mathrm{miss}$ is an important variable for processes with such elusive particles in the final state.[11]
In general, the $E_\mathrm{T}^\mathrm{miss}$ is calculated from the information of the calorimeter cells and the muons. There are cells that are associated with physical objects and those that are not. This may be expressed as

$$\not{E}_{x,y} = \not{E}_{x,y}^{\mathrm{PhysObjs}} + \not{E}_{x,y}^{\mathrm{CellOut}}, \tag{4.10}$$

[11] Supposed the $E_\mathrm{T}^\mathrm{miss}$ in an event is caused by two neutrinos then the $E_\mathrm{T}^\mathrm{miss}$ is the vector sum of the neutrinos in the transverse plane. The resulting vector is not necessarily large.

where $\not{E}_{x,y} = -E_{x,y}$ with E being the energy of the calorimeter cell. The cell calibration differs with the physical objects the cell is associated with. This results in a variety of calibration schemes depending among other things on the chosen object reconstruction algorithms. In this study the "RefFinal" scheme was used.
In RefFinal the cells are associated to physical objects. If there are two objects sharing a cell, the cell is associated to the object coming first in this list: electrons, photons, hadronically decaying tau-leptons, jets and muons. Furthermore algorithms to avoid the double counting of cells are applied. The cells are weighted with respect to the object type they are associated with resulting in an estimate of $E_{\mathrm{T}}^{\mathrm{miss}}$. Further corrections for energy that might be deposited in the cryostat of the calorimeters are applied. The final formula used for the calculation of the $E_{\mathrm{T}}^{\mathrm{miss}}$ is

$$\not{E}_{x,y}^{\mathrm{RefFinal}} = \not{E}_{x,y}^{e} + \not{E}_{x,y}^{\gamma} + \not{E}_{x,y}^{\tau} + \not{E}_{x,y}^{\mathrm{Jet}} + \not{E}_{x,y}^{\mu} + \not{E}_{x,y}^{\mathrm{CellOut}} + \not{E}_{x,y}^{\mathrm{Cryo}}. \tag{4.11}$$

The used calibrations stem from the object definitions used in this study (e. g. EM+JES for jets).

5 Analysis

5.1 Motivation

The aim of this analysis is the measurement and understanding of the distribution of the kinematic variables of so-called additional jets in top pair events. In the context of the signal process (semileptonic top pair decays) additional jets are jets that do not stem from the top pair.
The motivation for doing this is manifold. Firstly the comparison between the spectra obtained from analyzing data and Monte Carlo (MC) simulations enables one to draw conclusions about the quality of the MC simulations (e. g. how well kinematic variables are modeled).
Furthermore certain parameters used for the generation of MC samples need to be obtained from data. One example for such a parameter is the fraction of heavy flavored jets (jets that stem from b-hadrons).
A third point motivating this study is the better understanding of the distributions themselves. In searches for electrically charged Higgs bosons in the framework of models with more than one Higgs doublet field, like supersymmetric theories [51], channels are examined which look like top pair decays with additional jets which are decay products of a Higgs boson. Since the properties of the Higgs boson are quite distinctive (presumably high mass etc.) the distributions of the kinematic variables of these jets should be unique. By studying both cases, additional jets in top pair events without Higgs particles and with Higgs particles, one would obtain two sets of distributions for additional jets. Perhaps differences in these distributions could be exploited in order to facilitate Higgs searches. In any case a background estimate for these searches will be obtained.
This analysis is comprised of several steps will be used in order to extract the above-mentioned distributions. In principle every single one of these steps is worth a thorough and careful examination. This study will concentrate on the way top pairs are reconstructed in a given event and evaluate possible ways of optimization.

5.2 Prerequisites

In order to motivate aspects of the following analysis the properties of the semileptonic top pair decay are reviewed in this section.
Figure 5.1 shows the signature of a semileptonic top pair event[1]. It exhibits a charged lepton, a neutrino and four jets of which two are bottom jets.

[1]This is of course just one of many possible Feynman diagrams

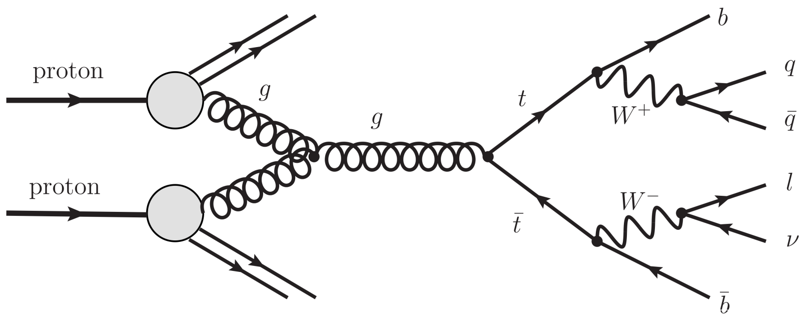

Figure 5.1: General layout of semileptonic top pair event. Sources for additional jets may be initial state radiation e. g. radiation from the gluons, final state radiation emitted from the decay products of the top or remnants from the protons. Additionally pile-up (proton-proton interactions taking place at the same time as the event) may introduce additional jets.

All of these objects originate from the decays of high mass particles (top quarks and W bosons). As these objects are not heavy themselves they should exhibit high momenta.
Apart from these objects there are additional objects in the event with jets that do not stem from the top pair being the main focus in this analysis. In principle these might originate from initial state radiation (ISR), final state radiation (FSR), pile-up or from the debris of the colliding protons.
In the following the measured objects that meet certain identification criteria will be called physical objects. The top quarks and W bosons from which some of these physical objects will be called particles of the event.

5.3 Analysis Algorithm

The algorithm used to achieve the goal of determining the distributions of kinematic variables of the additional jets and their flavor composition is comprised of five individual steps. The first necessary step is to prepare a clean sample of semileptonic top pair events. In the next step an algorithm that reconstructs the top pair is applied on the selected events. As soon as it is known which objects are associated to the top pair it is also known which are not (these are the additional objects).
Therefore a first naive algorithm chain is:

- Definition physical objects,
- Preparation of a clean semileptonic top pair sample,
- Reconstruction of the top pair from its decay products,
- Separation of the physics objects in groups (“objects from top” and “additional”),

- Determination of the kinematic variables and the flavor composition of the additional jets.

For the first two tasks it was decided to follow the prescription regarding object definitions and analysis cuts provided by the top group of the ATLAS experiment (see e. g. [26]). These prescriptions were adopted in order to have a common ground with respect to other top analysis groups making the comparison and evaluation of tools and techniques that differ between the groups easier. There are points in which this study deviates from these prescriptions. These cases will be clearly identified and motivated.
Once the top pair has been reconstructed from its decay products the separation of physics objects in groups and the determination of their kinematic variables is straightforward.
A bit more involved is the determination of the flavor composition of the additional jets since for 2010 data the obtained sample is small. Therefore only statistics inexpensive methods are applicable and the statistical errors of the measurement are expected to be large. Fortunately the data collected in 2011 will be about two orders of magnitude larger which will enable more complicated methods for the determination of the flavor composition. One of these techniques, the template fit method, will be described and applied on 2010 data for a first evaluation whether it presents a viable method or not.
Item three remains, the reconstruction of the top pair from its decay products. As can be seen from the multitude of top pair reconstruction methods that exist (e. g. the KLFitter [52], χ^2 methods et al. [53], etc.) the task at hand is not easily achieved. Studies using crude and simple reconstruction methods such as defining that the top decaying subsequently to three quarks may be reconstructed by choosing the three jets with the highest energy in the event have shown that such criteria seldom provide the right answers.[2]
Therefore more complex algorithms have been developed using both more information provided by the measurement of the physical objects in the event and more assumptions about the particles in the event.
The algorithm used and investigated in this study is the KinFitter [54]. This reconstruction algorithm uses the kinematic variables of the physical objects in the event and evaluates how consistent they are with respect to a given fit hypothesis. The mathematics the algorithm is based on as well as its application, merits and shortcomings will be discussed in Section 5.3.3 after giving more information about the objects definitions and the applied analysis cuts.

5.3.1 Objects definitions

In general not all reconstructed objects in an event are used in an analysis but a subset of them satisfying certain criteria. Depending on the physics process in question the objects in the final state have to fulfill certain characteristics.
An example for such a characteristic is the isolation criterion for muons. In top pair events it is expected that the muons are well isolated since their origin lies in the hard

[2] It should be noted at this point that the notion of a right answer always depends on the way that is used to define the right answer.

scattering process and not from a decay inside a jet. Therefore it is demanded that only small energy deposits exist in their vicinity. On the other hand if one wants to specifically select muons from decays inside jets this criterion may be inverted. The following definitions are the ones recommended by the top reconstruction group of the ATLAS experiment for the winter conferences of 2011 (see e. g. [26]).

Electrons

The Electrons used in the analysis are required to exhibit a high transverse momentum and an associated track. Furthermore the tight quality criteria have to be met.
The transverse energy of the electron which is defined as the energy of the cluster from which the electron was reconstructed divided by the cosinus hyperbolicus of the η component of the track belonging to the electron (short: $E_{\text{cluster}}/\cosh(\eta_{\text{track}})$) has to exceed 20 GeV. The pseudo-rapidity of the cluster from which the electron was reconstructed has to be inside the region of $|\eta_{\text{cl}}| < 2.47$. Furthermore the crack region[3] is excluded, therefore no electrons are accepted with $1.37 < |\eta_{\text{cl}}| < 1.52$.
The isolation criterion demands that the energy deposited in a cone of $\Delta R = 0.2$ around the electron does not exceed 4 GeV.
In order to make sure that the electron only traversed regions which are fully operation, so-called object quality maps (OQMaps) are employed. These are basically maps of the regions where dead EM Calorimeter electronics are situated. If an electron traverses a region that is not fully operational it is discarded.

Muons

The muons used in the analysis were reconstructed by the Muid algorithm. They are required to be combined muons within the tight category. Their transverse momentum has to exceed 20 GeV and they have to be in the eta range of $|\eta| < 2.5$.
There are several requirements concerning inner detector hits:

- The muon has to cause a hit in the b-layer (the innermost sheet of the pixel detector) unless its track indicates that the muon traversed a region where the detector malfunctions.
- The sum of hits in the pixel detector plus the crossed malfunctioning pixel sensors has to exceed 1.
- The sum of hits in the SCT plus the crossed malfunctioning SCT sensors has to be greater or equal to 6.
- The sum of the traversed malfunctioning pixels in the pixel detector plus the traversed malfunctioning SCT sensors has to be lower than 2.

[3] This region spanning $1.37 < |\eta_{\text{cl}}| < 1.52$ contains peripheral infrastructure such as cables for the detectors. Therefore measurements in this region are not possible.

- The requirements on the variable N denoting the sum of hits in the TRT plus TRT outliers[4], depend on η as follows:

  For $|\eta| < 1.9$ N has to exceed 5 and the number of TRT outliers divided by N has to be lower than 0.9 while for $|\eta| \geq 1.9$ it is requaired that if N exceeds 5 the number of TRT outliers divided by N has to be lower than 0.9. There is no requirement for the case where $N \leq 5$ and $|\eta| \geq 1.9$.

In order to ensure the isolation of the muon two isolation criterias are implemented. It is required that the sum of transverse momenta in a cone of $\Delta R = 0.3$ around the muon does not exceed 4 GeV and that the measured transverse energy in a cone of $\Delta R = 0.3$ around the muon is lower than 4 GeV.
Furthermore, muons that lie within $\Delta R < 0.4$ around a jet with a transverse momentum higher than 20 GeV are rejected in order veto muons that might originate from a decay inside a jet.
Additionally muons which might stem from cosmic interactions are discarded by requiring that their point of nearest approach to the beam axis lies in the vicinity of the primary vertex.

Jets

The jets used in this analysis stem from the AntiKt4TopoEMJets collection. Jets inside this collection are reconstructed using the anti-k_t algorithm (see Section 4.1) with topological clusters as input. Furthermore the jets are calibrated using the numerical inversion procedure as described in [41]. From the viewpoint of the objects definitions of the top group the jets only have to satisfy a few constraints. Jets with negative energy[5] are removed. Additionally, jets that lie inside a cone of $\Delta R < 0.2$ around an electron, as defined by the objects definitions of the top group, are discarded, ensuring the removal of overlapping objects.
In this analysis only jets satisfying the conditions $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$ are retained. The p_{T} requirement is motivated by the fact that important calibration studies such as the b-tagging calibration are only made for jets with $p_{\mathrm{T}} > 20$ GeV. The motivation for the η requirement lies in the geometry of the detector. The inner detector that is needed for b-tagging techniques roughly spans the η range of $[-2.5, 2.5]$. Since this study relies on variables coming from b-tagging techniques the η range is restricted by the availability of these techniques.

Missing Transverse Energy

The missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ is calculated in accordance to the object definitions given above. The vector sum of the energy deposits in the calorimeters as measured with respect to the transverse plane is used for the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction.

[4] According to [55] outliers are either TRT drift tubes with a signal which are not traversed by the near-by track or TRT measurements which result in a non-smooth trajectory when combined with Pixel and SCT measurements.

[5] This may happen due to a bug in the jet reconstruction software that was spotted after the data and MC simulation samples had been produced. This does not affect the analysis as there are no jets in the used p_{T} range that exhibit negative energy values.

The energy calibrations of the muons, electrons and jets characterized by the object definitions are used in the E_T^{miss} calculations. Cells that are not associated to physical objects (muons, electrons with $p_T > 10$ GeV, photons, taus, jets and soft jets) contribute with the energy measured at the electromagnetic scale[6].

5.3.2 Event Selection

The selection of top pair events is based on cuts recommended by the top group (see e. g. [56]). These contain requirements for both event selection and event cleaning. The names of the items are the ones found in Figure 5.2. In order of application, these are:

- c1 - Trigger: This cut requires that either the muon trigger or the electron trigger has fired. Due to the increase in instantaneous luminosity during the data taking phase the used trigger changed with time in order to avoid event rates exceeding the technically possibly data recording rate. All the triggers work reliably in the transverse momentum range of the electrons and muons defined by the object definitions.
- c2 - prim vtx: The primary vertex has to be the origin of five or more tracks. In the semileptonic top pair final state there are five charged particles (four quarks plus a charged lepton). Therefore the requirement reflects the rich final state of semileptonic top pair events.
- c3 - No bad jets: If the event contains any badly measured jets, it is discarded. This cut is only applied in data as the variables defining the criteria are not well modeled in MC simulations.
- c4 - leptons > 1: There has to be at least one electron or muon as specified in the objects definitions in the event.
- c5 - leptons == 1: The event contains exactly one electron or one muon. The or is an inclusive or.
- c6 - lepton veto: This cut requires either one electron or one muon. In contrast to the previous cut this is meant as an exclusive or.[7]
- c7 - trigger match: After the previous cut there is only one lepton left in the event. This lepton has to be the one that caused the trigger to fire. The event satisfies the cut if the direction of flight of the lepton matches the region of interest (RoI) that was defined by the trigger.
- c8 - e/mu track share: It is prohibited that electrons and muons near a jet share a track. The information needed for this cut is determined prior to the cuts on the leptons.

[6]This means that the calibration obtained from test beam studies of the detector using electrons is used

[7]The lepton cuts were implemented in three steps in order to study the effects on the background samples more closely. The same reasoning applies for the cuts on the jet multiplicity below.

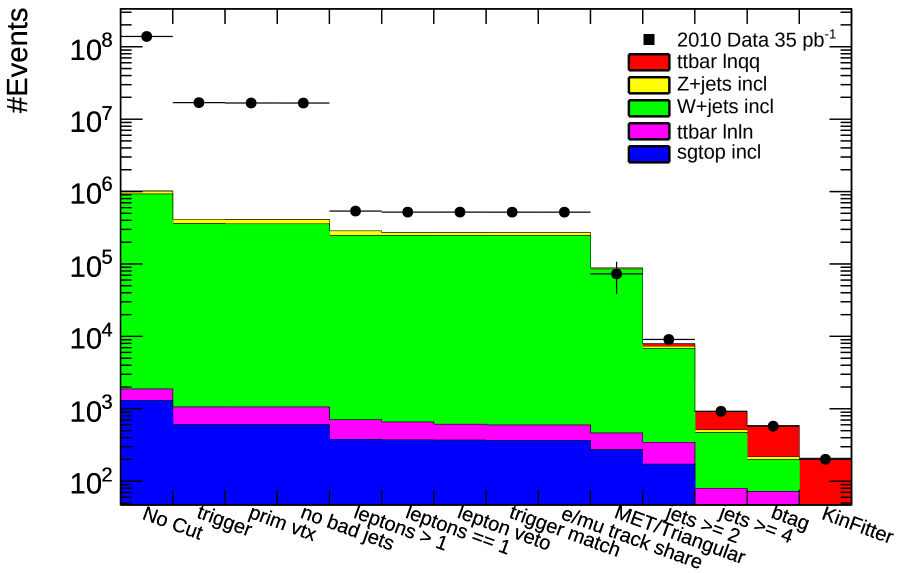

Figure 5.2: Cut flow of the top pair analysis in data and MC simulations. Information on the used MC simulation samples can be found in Section8.2.

- c9 - MET/Triangular: This cut requires that the $E_{\mathrm{T}}^{\mathrm{miss}}$ and the transverse mass of the W boson decaying to leptons which is calculated as

$$\sqrt{2 \cdot p_{\mathrm{T}}^{\mathrm{lep}} \cdot E_{\mathrm{T}}^{\mathrm{miss}} \cdot (1 - \cos(\phi^{\mathrm{lep}} - \phi^{E_{\mathrm{T}}^{\mathrm{miss}}}))} \tag{5.1}$$

  lie above certain thresholds. If the lepton in the event is an electron, the values are 20 GeV for the $E_{\mathrm{T}}^{\mathrm{miss}}$ and 65 GeV for the sum of the transverse mass of the W boson and the $E_{\mathrm{T}}^{\mathrm{miss}}$. In the muon case the $E_{\mathrm{T}}^{\mathrm{miss}}$ has to be larger than 35 GeV and the transverse mass of the W boson has to exceed 25 GeV.

- c10 - jets ≥ 2: There have to be at least 2 jets in the event that are well measured and have a $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.5$.

- c11 - jets ≥ 4: This cut requires 4 jets which have to have the same properties as stated in the previous cut.

- c12 - b-tag: Only events that exhibit at least one jet for which the JetProb algorithm yields a weight of more then 2.05 (70 % efficiency) are retained.

- c13 - KinFitter: This last cut demands that the top pair finding via the KinFitter algorithm was successful.

The resulting cut flow is shown in Figure 5.2. Mismatches between measured data and MC simulations stem partly from the omission of QCD and flaws in the modeling of variables used in the cut flow in MC simulations.
In Table 5.1 and Table 5.2 the event yields after the individual cuts and the cut efficiencies can be found.

channel	no cut	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12	c13
ttbar lnqq	2552	1484	1484	1484	865.8	865.2	864.6	852.6	852.6	658.7	650.4	406.6	371.3	171.5
ttbar lnln	566	452.8	452.8	452.8	327.6	283.2	243.0	230.9	230.9	189.1	172.8	56.36	52.36	9.322
ttbar qqqq	2623	274.6	274.6	274.6	1.991	1.991	1.991	1.991	1.991	0.158	0.158	0.087	0.068	0.022
sgtop	1311	606.7	606.6	606.6	379.4	374.1	369.9	366.5	366.5	274.1	172.1	23.59	19.85	4.813
WJets	923300	360109	357717	357717	247214	247211	247208	247169	247169	84914	6502	389.3	128.5	19.97
ZJets	89932	51322	51107	51107	37866	22962	22924	22565	22565	2313	451.4	42.42	16.43	2.814
Diboson	833.9	623.9	623.7	623.7	426	370.8	358.4	353.9	353.9	228.4	99.53	5.56	1.853	0.355
sum MC	1.0e06	414873	412266	412266	287081	272068	271970	271539	271539	88578	8048	924	590.3	208.8
data	1.0e08	1.7e07	1.7e07	1.7e07	537863	520354	520233	519369	519369	72946	9077	924	576	201

Table 5.1: The event yields after the individual analysis cuts scaled to an integrated luminosity of $35\,\mathrm{pb}^{-1}$. The dominating backgrounds are W + jets, Z + jets, dileptonic top pair events and single top production. All other backgrounds were omitted from the analysis because their contributions are negligible. For details on the used MC simulation samples see Section 8.2.

channel	eff1	eff2	eff3	eff4	eff5	eff6	eff7	eff8	eff9	eff10	eff11	eff12	eff13
ttbar lnqq	0.58	1.00	1.00	0.58	1.00	1.00	0.99	1.00	0.77	0.99	0.63	0.91	0.46
ttbar lnln	0.80	1.00	1.00	0.72	0.86	0.86	0.95	1.00	0.82	0.91	0.33	0.93	0.18
ttbar fullhad	0.10	1.00	1.00	0.01	1.00	1.00	1.00	1.00	0.08	1.00	0.56	0.77	0.33
sgtop	0.46	1.00	1.00	0.63	0.99	0.99	0.99	1.00	0.75	0.63	0.14	0.84	0.24
WJets	0.39	0.99	1.00	0.69	1.00	1.00	1.00	1.00	0.34	0.08	0.06	0.33	0.16
ZJets	0.57	1.00	1.00	0.74	0.61	1.00	0.98	1.00	0.10	0.20	0.09	0.39	0.17
Diboson	0.75	1.00	1.00	0.68	0.87	0.97	0.99	1.00	0.65	0.44	0.06	0.33	0.19
sum MC	0.41	0.99	1.00	0.70	0.95	1.00	1.00	1.00	0.33	0.09	0.11	0.64	0.35
data	0.12	0.99	1.00	0.03	0.97	1.00	1.00	1.00	0.14	0.12	0.10	0.62	0.35

Table 5.2: Efficiencies of the analysis cuts. The last efficiency corresponds to the top pair reconstruction via the KinFitter. For details on the used MC simulation samples see Section 8.2.

5.3.3 Kinematic Fit

The tool used for the reconstruction of the top pair is the KinFitter which was written for ATLAS by Thomas Göpfert and Jan Erik Sundermann and is documented in [54]. The general idea behind reconstruction algorithms such as the KinFitter is that the kinematics of the physical objects in the event depend on the characteristics of the particles they stem from.
For instance the invariant mass distribution of muon pairs can exhibit a peak at the mass of the Z boson if Z decays to muons have contributed. The invariant mass of the muon pair is calculated via the formula

$$M_{\mathrm{inv}}^2 = (\mathbf{p}_{\mu_1} + \mathbf{p}_{\mu_2})^{\mathrm{T}}(\mathbf{p}_{\mu_1} + \mathbf{p}_{\mu_2}) \tag{5.2}$$

where the Lorentz vectors are the bold faced symbols. When applying the hypothesis that in a given event a Z boson is decaying to two muons one would expect that the calculated invariant mass of the muon pair lies near the known mass of the Z boson. Using such an hypothesis events can accepted or rejected depending on the agreement between the measurement and the hypothesis. If there are more than two muons in the event a multitude of unique muon pairs (candidates) may be formed which are then evaluated by applying the fit hypothesis.
The KinFitter evaluates how well the kinematics of the physical objects in the final state match to a fit hypothesis which is characterized by constraints such as the one above. Furthermore it varies the kinematic variables of the physical objects in the range of their errors in order to satisfy the constraints thus correcting the measured variables. The amount of correction that is needed is reflected by a figure of merit which is then used to accept or reject a given event or candidate.
In the following the mathematical framework of the KinFitter will be described and the fit hypothesis for this analysis explained.

Mathematical implementation

The KinFitter employs the method of "Langrange multiplicators" which is for instance described in [57]. In the following discussion of the method the variables are

- $\vec{y}$: vector of measured variables e. g. p_{T}, η, ϕ of a particle,
- $\vec{a}$: vector of unmeasured variables e. g. the component of the momentum parallel to the beam line of a neutrino,
- f_i: various constraints.

The fit hypothesis is represented by a set of m constraints

$$\begin{aligned} f_1(\vec{y}, \vec{a}) &= 0, \\ f_2(\vec{y}, \vec{a}) &= 0, \\ &\dots \\ f_m(\vec{y}, \vec{a}) &= 0. \end{aligned} \tag{5.3}$$

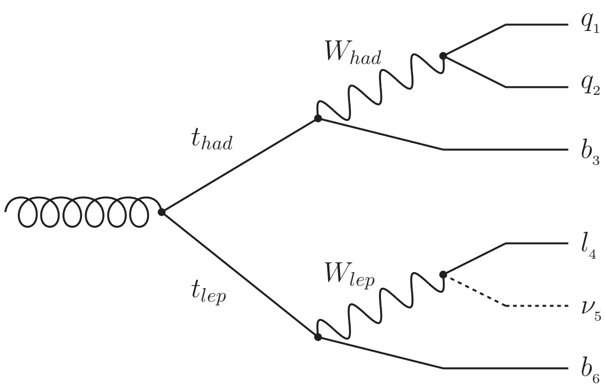

Figure 5.3: Connections between particles in the semileptonic top pair case

These constraints establish connections between the various variables. In general the measured quantities will not fulfill the given constraints. One cause for this are the uncertainties accompanying the measurement. Therefore the measured variables might be altered by $\Delta\vec{y}$ so that one gets new variables $\vec{y}' = \vec{y} + \Delta\vec{y}$ which satisfy the constraints. In addition the unmeasured variables $\vec{a}$ may be determined by employing the constraints f_i.
the size of the corrections $\Delta\vec{y}$ are calculated by minimizing the expression

$$S(\vec{y}) = \Delta\vec{y}^T V^{-1} \Delta\vec{y}, \tag{5.4}$$

where V denotes the covariance matrix of the measured variables $\vec{y}$.
In order to express these steps in one formula m additional parameters λ_k, one for each of the m constraints, are introduced yielding the following formula

$$L(\vec{y}, \vec{a}, \vec{\lambda}) = S(\vec{y}) + 2\sum_{k=1}^{m} \lambda_k f_k(\vec{y}, \vec{a}). \tag{5.5}$$

Searching for extrema of this formula with respect to $\vec{y}$, $\vec{a}$ and $\vec{\lambda}$ is equivalent to minimizing $S(\vec{y})$ under the constraints $\vec{f}(\vec{y}, \vec{a}) = \vec{0}$.
If the constraints are linear then the solution may be calculated directly, otherwise the solution has to be found iteratively linearizing the problem in every iteration of the extremum finding process.

Fit Hypothesis

The fit hypothesis determines the connections between the particles and how key variables are constrained in detail. In semileptonic top pair events, the connections between the particles are as illustrated in Figure 5.3.
The four particles which may be reconstructed from the physical objects in the final state are

- the W boson W_{had} decaying to the quarks q_1 and q_2 called the hadronic W boson[8],

[8] These are naming conventions motivated brevity.

- the W boson W_{lep} decaying to charged lepton l_4 and the neutrino ν_5 called the leptonic W boson,
- the top quark t_{had} decaying to a bottom quark b_3 and the hadronic W boson W_{had} called the hadronic top quark,
- the top quark t_{lep} decaying to a bottom quark b_6 and the leptonic W boson W_{lep} called the leptonic top quark.

These connections provide the prescriptions for the combination of the four vectors of the daughter particles yielding the four vectors of the mother particles

$$p_{\mathrm{W_{had}}} = p_{\mathrm{q_1}} + p_{\mathrm{q_1}}, \tag{5.6}$$

$$p_{\mathrm{W_{lep}}} = p_{\mathrm{l_4}} + p_{\nu_5}, \tag{5.7}$$

$$p_{\mathrm{t_{had}}} = p_{\mathrm{W_{had}}} + p_{\mathrm{b_3}}, \tag{5.8}$$

$$p_{\mathrm{t_{lep}}} = p_{\mathrm{W_{lep}}} + p_{\mathrm{b_6}}. \tag{5.9}$$

With defining the connections between the particles as done in the equations an important assumption about the event is made: there are always four jets stemming from the top pair. The implications of this assumption will be discussed in Section 5.3.3.
The four vectors of the particles that will be reconstructed may be constrained imposing restrictions on the four vectors of the physical objects in the event.
The constraints used in this study stem from assumptions about the masses of the reconstructed top quarks and W bosons.
In the case of the W bosons W_{had} and W_{lep} the mass is fixed to a certain value [12]:

$$M_{\mathrm{W}} = 80.399 \pm 2.085 \pm 0.023 \ \ \mathrm{GeV}. \tag{5.10}$$

The first uncertainty is the natural width of the W boson while the second one is the experimental uncertainty on the mass value. The KinFitter is able to account for these uncertainties by describing the mass peak via a gaussian whose mean is the mass value and whose width is the squared sum of the uncertainties. The reason for incorporating the experimental uncertainty into the gaussian stems from considerations about the comparability of the measured values of the physical objects and the masses of the W bosons. A perfect detector could measure the W boson mass by analyzing its decay products with infinite precision. Imperfections that afflict every machine lead to shifts and uncertainties on the measured values of the decay products. This leads to a possibly shifted value for the mass of the W boson as well as to uncertainties on the mass measurement. If a fixed mass constraint is applied the measured value of the particle in question has to be taken rather than a theoretical one which would spoil the comparability of the mass of the particle and the measured kinematic variables of the physical objects it decays to.
or the W boson the world average value and its uncertainty as provided in [12] was used[9].

[9]Another possibility to solve the problem of comparability between the constraints and the input variables is to correct the input variables to generator-level and use the values as specified in the MC production as was done in [58].

Regarding the mass of the top quark there are two viable possibilities. Again the masses of the t_{lep} and t_{had} might be set to a fixed value

$$M_t = 172.0 \pm 0.9 \pm 1.3 \pm 2.05 \text{ GeV}, \tag{5.11}$$

which defines the fixed mass constraint used in this study. Here the masses of the top quarks are assumed to be independent of each other but they are required to lie inside a given window around a known value.
The first three values used are stemming from the world best average [12] and are in that order the measured mass of the top quark, the statistical uncertainty and the systematical uncertainty of the mass measurement. The fourth value is the decay width as measured by D0 [59]. The squared sum of the three uncertainties is taken as the width of the gaussian describing the mass peak of the top quark.
The second possibility is to require the masses of the t_{lep} and t_{had} to be equal which defines the same mass constraint. Here both masses depend on each other but their numerical value is not fixed. Using this constraint a mass measurement of the top quark mass via the KinFitter can be done.
Both constraints have their vices and virtues. In general the fixed mass constraint performs better than the same mass constraint in terms of background suppression and top pair reconstruction, but using a fixed mass relies heavily on detector understanding.

Parameterization of Objects

The KinFitter takes the four vectors of the physical objects in the event as input for its calculations. A choice has to be made regarding the parameterization of abovementioned four vectors. The choice of the parameterization largely depends on their applicability and how well their uncertainties and covariances are understood.
For jets and charged leptons the measured variables are the transverse momentum p_T, the pseudo-rapidity η and the azimuthal angle ϕ. Since it takes four variables in order to define a four vector, assumptions about the mass of the physical object in question are made. For all light jets the mass is set to zero and for bottom jets it is set to 4.5 GeV. The masses of the charged leptons are set to the values found in [12] while the neutrinos are treated as massless.
After the definition of the parameterization the covariances of the used variables have to be studied. As the masses of the physical objects are fixed masses, only the covariances between the measured variables p_T, η and ϕ are studied.
In the case of the charged lepton the covariance matrix is taken from the fit of the track associated to the charged lepton.
For jets and missing transverse energy a study was performed by Michelangelo Giorgi of the Humboldt University in Berlin (the code may be found in [60], no official documentation is available).
In this study jets exhibiting transverse momenta greater than 5 GeV and satisfy $|\eta| < 5$ are selected. These jets are matched to jets on hadron-level via a ΔR and energy fraction criterium. For all matching pairs the values $x_{reco} - x_{truth}$ and $(x_{reco} - x_{truth})(y_{reco} - y_{truth})$ (with $x, y = p_T$, η or ϕ) are filled into histograms. This is done in dependence to η and p_T yielding a two dimensional map of sets of histograms.

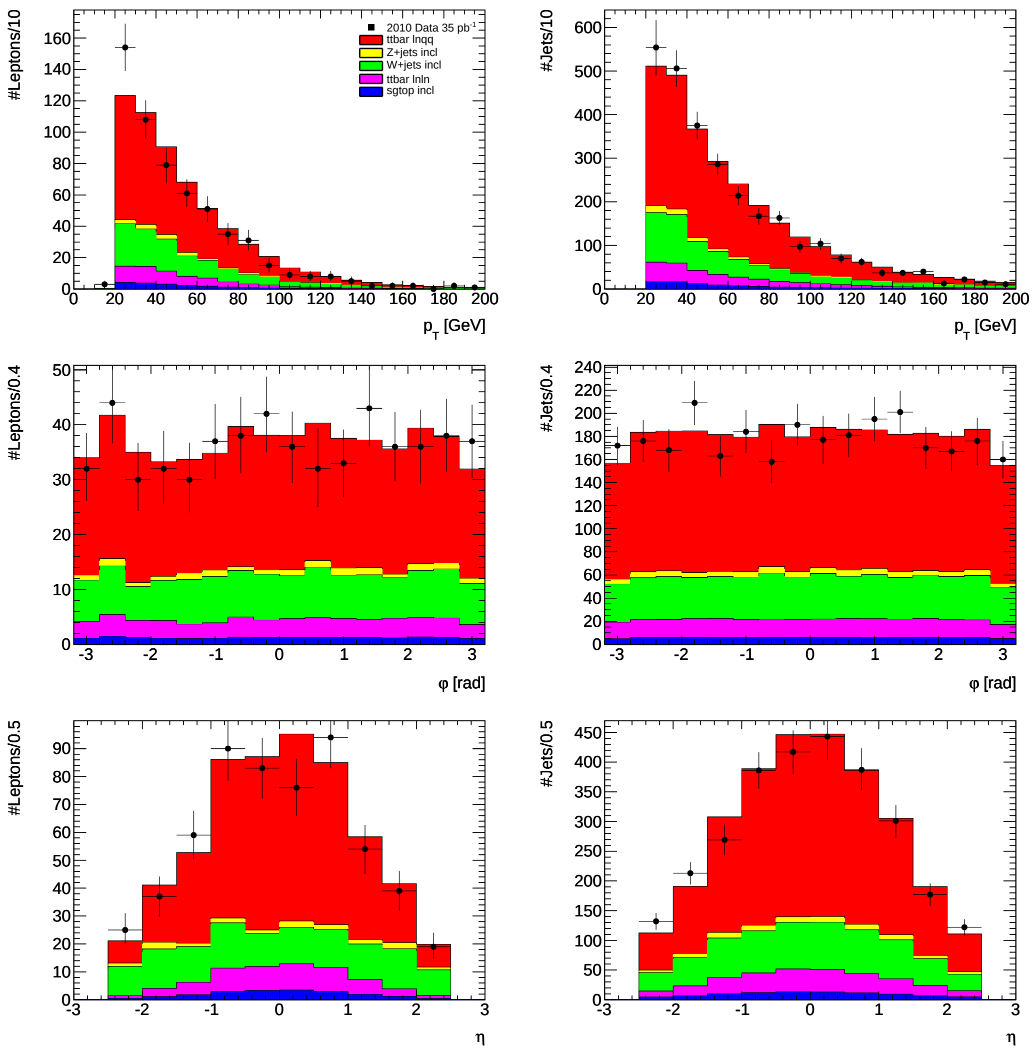

Figure 5.4: The distributions of the input variables for the kinematic fits for the leptons (left) and jets (right). Displayed are transverse momentum (first row), azimuthal angle (second row) and pseudo-rapidity (third row). The legend in the upper left plot applies for all plots.

This data acquisition was done for several MC datasets (namely single top and top pair events). The results are summed with respect to the cross-sections of the different samples. Via a Gaussian fit the covariances and uncertainties of the measured variables are determined from the summed histograms.

For the missing transverse energy the procedure is basically the same but the reference object on truth level is the neutrino with the highest energy since the missing transverse energy is always interpreted as the neutrino in the scenarios where the KinFitter is employed.

For the neutrino the parameterization consists of p_T, θ and ϕ. Here θ was chosen because KinFitter studies [61] revealed that it performs better than η.

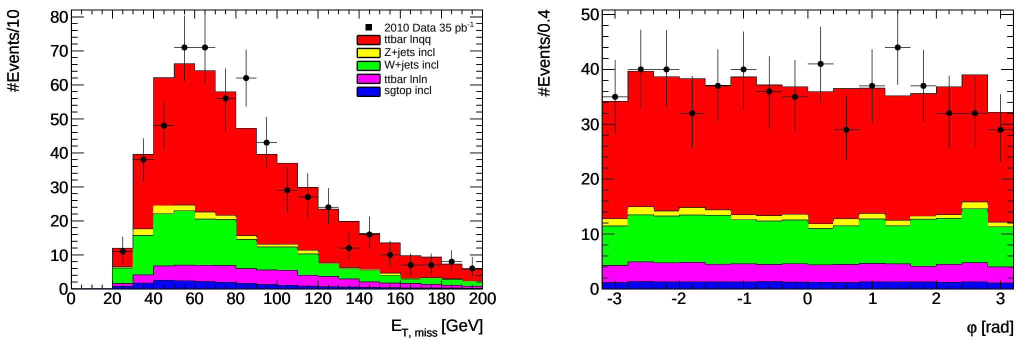

Figure 5.5: The distributions of the input variables for the kinematic fits of the missing transverse energy. Displayed are the magnitude of the missing transverse energy vector (left) and its azimuthal angle (right). The legend in the left plot applies for all plots.

Neutrino Reconstruction

Prior to the application of the KinFitter the neutrino is reconstructed using the information provided by the measurement of the missing transverse energy. Unfortunately this only yields two components of the four vector namely the momenta p_{x} and p_{y}. After applying the assumption that the neutrino may be treated as massless the p_{z} component is still a free parameter. In the fit this free parameter will be determined using the information encoded in the constraints. In order to support the KinFitter an estimate for the p_{z} component of the neutrino is computed. One reason for this is to avoid possible local minima in the fitting process.
The estimation is done by exploiting the connections between the charged lepton, the neutrino and the bottom quark and the $\mathrm{t_{lep}}$ and $\mathrm{W_{lep}}$ established by the fit hypothesis. By solving the equations

$$(p_{\mathrm{l}_4} + p_{\nu_5})^2 - M^2_{\mathrm{W_{lep}}} = 0 \tag{5.12}$$

$$(p_{\mathrm{l}_4} + p_{\nu_5} + p_{\mathrm{b}_6})^2 - M^2_{\mathrm{t_{lep}}} = 0 \tag{5.13}$$

either two roots or a minimum (if the roots do not exist) are obtained from each equation. Now pairs are formed taking one solution of the first equation and one solution of the second equation. The pair where the solutions are nearest to each other is taken and their arithmetic mean is used as the estimate for p_{z}.

Fitting Procedure

After defining the fit hypothesis and the parameterization of the physical objects top pair candidates which will be evaluated by the KinFitter have to be built.
A top pair candidate consists of four jets, one charged lepton and a neutrino. Every object has a certain role e. g. one jet takes the place of the bottom quark coming from the $\mathrm{t_{lep}}$ while another one is one of the quarks from the $\mathrm{W_{had}}$. In an event with four jets and one charged lepton there are already $4! = 24$ possible top pairs. Since the fit hypothesis equation for the $\mathrm{W_{had}}$ is symmetric for q_1 and q_2 top pairs which can be transformed into each other by switching the jets associated to q_1 and q_2 yield

type	ID	p_T [GeV]	η	ϕ [°]	b-tagged	flavor	colorcode
muon		21.981	-0.209	-0.849			
neutrino		214.339	0.0	-50.946			
jet	1	136.814	-0.387	114.610	no	Light	red
	2	58.281	0.683	142.939	no	Light	green
	3	50.377	-1.579	47.939	yes	Bottom	light blue
	4	47.923	0.502	-148.208	no	Bottom	orange
	5	25.004	-1.152	144.928	no	Light	light blue

Table 5.3: Physical objects of event 20186 from the MC@NLO sample describing semileptonic top pair events. Jets are tagged as bottom jets if their b-tagging weight calculated by JetProb exceeds 2.05. The flavor of the jet as obtained from a matching scheme involving particles on generator-level and the jets on detector-level, for details see Footnote 17 in Section 5.5.1.

the same fit result. Therefore the number of unique top pairs for the four jet case is $4!/2! = 12$. This was considered in the analysis as well as exploited.
For the case of an event with five jets there are already $5!/2! = 60$ unique top pairs. These make up five groups of top pair candidates with respect to the jet content. As this study is more concerned with the separation of jets into the groups “additional jets” and “jets from top” the point of primary interest is which jets are belonging to a top pair while the role that each jet fulfills is of secondary importance.
This is best illustrated using a simulated example from the analysis. The event in question is event number 20186 from the MC@NLO [62] sample with run number 105200 which contains semileptonic and dileptonic top pair events[10] as depicted in Figure 5.6. The variables used in the parameterization are given in Table 5.3.
The properties of the jets reveal that although jet 4 is a bottom jet according to the MC simulation it is not tagged as such because the b-tagging weight calculated by JetProb lies below the cut value.
The physical objects from Table 5.3 will be used in the generation of top pair candidates. The lepton will always take the role of the lepton coming from the decay of the W_{lep} and its associated neutrino will be reconstructed from the E_T^{miss}.
Therefore the permutation of the jets is what discerns the top pair candidates from each other. In Table 5.4 all top pair candidates with their fit results and the masses for the reconstructed particles calculated from the initial non-fitted kinematic variables for the event 20186 are given.
Every candidate is assigned a $P(\chi^2)$ value calculated from the χ^2 of the fit and its number of degrees of freedom (NDoF)[11]. The candidate with the best (highest) $P(\chi^2)$

[10] The event number has in principle no relevance in simulated data. It is only used as an identifier and reference. The dataset was taken from the mc10 round of MC generation.

[11] The NDoF is defined as constraints minus unmeasured variables. In the case of the fixed top and W boson mass constraints the NDoF equals 3 while in the same top mass case with fixed W

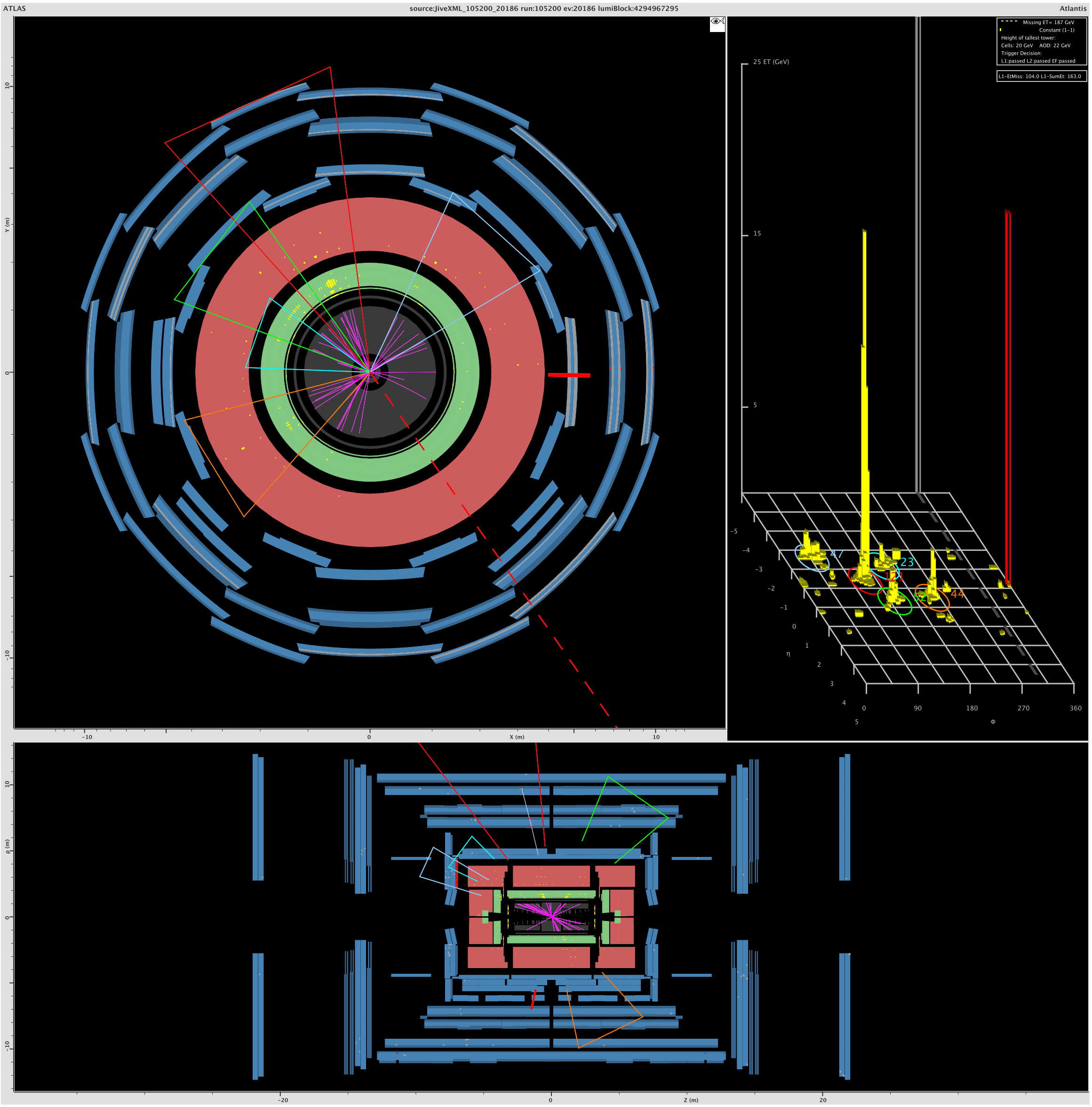

Figure 5.6: Event number 20186 obtained from the top pair sample generated by MC@NLO [62]. The event sports five jets satisfying the jet criteria, a muon and a high amount of missing transverse energy.

cand	grp	q_1	q_2	b_3	b_6	χ^2	$P(\chi^2)$	$m_{W_{had}}$	$m_{W_{lep}}$	$m_{t_{had}}$	$m_{t_{lep}}$
1	1	1	2	4	3	19.99	0.000	109.68	90.33	190.36	195.85
2	2	1	2	5	3	107.66	0.000	109.68	90.33	151.27	195.85
3	1	1	2	3	4	34.23	0.000	109.68	81.10	246.22	185.57
4	2	1	2	3	5	32.00	0.000	109.68	81.03	246.22	170.95
5	1	1	4	3	2	104.81	0.000	143.02	86.47	253.75	250.35
6	1	1	4	2	3	106.16	0.000	143.02	90.33	190.36	195.85
7	3	1	4	5	3	239.81	0.000	143.02	90.33	172.35	195.85
8	3	1	4	3	5	61.82	0.000	143.02	81.03	253.75	170.95
9	2	1	5	3	2	53.16	0.000	65.32	86.47	166.06	250.35
10	2	1	5	2	3	11.14	0.011	65.32	90.33	151.27	195.85
11	3	1	5	4	3	13.48	0.004	65.32	90.33	172.35	195.85
12	3	1	5	3	4	12.40	0.006	65.32	81.10	166.06	185.57
13	1	2	4	3	1	78.95	0.000	62.13	68.08	239.55	363.57
14	1	2	4	1	3	40.90	0.000	62.13	90.33	190.36	195.85
15	4	2	4	5	3	45.06	0.000	62.13	90.33	124.54	195.85
16	4	2	4	3	5	59.89	0.000	62.13	81.03	239.55	170.95
17	2	2	5	3	1	25.90	0.000	81.65	68.08	198.91	363.57
18	2	2	5	1	3	15.72	0.001	81.65	90.33	151.27	195.85
19	4	2	5	4	3	115.54	0.000	81.65	90.33	124.54	195.85
20	4	2	5	3	4	9.06	0.028	81.65	81.10	198.91	185.57
21	3	4	5	3	1	27.60	0.000	71.10	68.08	182.46	363.57
22	4	4	5	3	2	51.52	0.000	71.10	86.47	182.46	250.35
23	3	4	5	1	3	6.41	0.093	71.10	90.33	172.35	195.85
24	4	4	5	2	3	57.94	0.000	71.10	90.33	124.54	195.85

Table 5.4: Top Pair Candidates of event 20186 of the MC@NLO top pair sample. The variable "grp" identifies the group the candidate belongs to according to the jet content of the candidate (compare candidates 1 to 3). The masses of the reconstructed particles are calculated from the initial kinematic variables of the jets and leptons not from the fitted ones. Marked with green are the candidates that satisfy the $P(\chi^2)$ cut of 0.01. The best candidate is number 23.

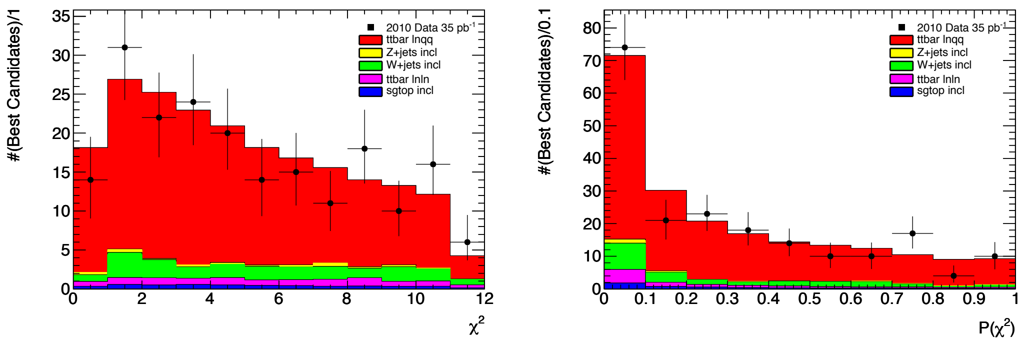

Figure 5.7: The χ^2 (left) and P(χ^2) (right) distributions as obtained from the kinematic fit.

is chosen to be the reconstructed top pair of the event.

There are less than 60 candidates in the event. This is caused by applying a scheme exploiting b-tagging information.

If at least two jets are tagged by the used b-tagging algorithm[12] in the event a list is generated consisting of tagged jets and jets with $|\eta| > 2.5$[13] and only jets from this list are used when assigning the roles of b_3 and b_6. If only one jet that received a tag from the b-tagging algorithm is present then it is required that this jet or a jet with $|\eta| > 2.5$ takes the role of either b_3 or b_6 leaving all other jets free to fulfill any vacant role.

This is done in order to reduce the number of top pair candidates in a given event both for performance reasons (reducing computation time) and helping the KinFitter finding the right candidate. Such additional constraints may exclude the right top pair candidate from the fit. Nevertheless it was observed that applying such constraints increases the performance of the KinFitter with respect to finding the right top pair candidate significantly.

Output of the KinFitter

The output of the KinFitter consists of two figures-of-merit (χ^2 and P(χ^2)) and the altered four-vectors of the physical objects used in the fit. These allow a monitoring of the tool itself as well as providing the kinematic variables of the top quarks and W bosons.

Figure 5.7 shows the χ^2 and P(χ^2) distributions of the best top pair candidates in the selected events. The P(χ^2) distribution is especially interesting as it provides the possibility to check whether the kinematic fit works as intended or not.

The χ^2 distribution is defined as the sum of the squares of independent standard normal random variables Y_i:

boson masses it equals 2.

[12] The JetProb algorithm at 70% b-tag efficiency is used (weight = 2.05).

[13] The prescription takes into account that jets outside the η region where b-tagging is applicable may be bottom jets. It is thought to be safer to include these jets although this increases the number of possible top pair candidates.

$$\chi^2 = \sum_{i=1}^{r} Y_i^2. \tag{5.14}$$

Here r denotes the degrees of freedom. The cumulative distribution function is:

$$\mathrm{P}(\chi^2) = 1 - \frac{\Gamma\left(\frac{1}{2}r, \frac{1}{2}\chi^2\right)}{\Gamma\left(\frac{1}{2}r\right)}. \tag{5.15}$$

Therefore $\mathrm{P}(\chi^2)$ is better suited for comparisons because the number of degrees of freedom r is incorporated. Furthermore the cumulative distribution function is flat if the variables constituting the χ^2 distribution are standard normal random variables. As can be seen in 5.7 the $\mathrm{P}(\chi^2)$ distribution is not flat but exhibits a certain slope. This indicates that the assumption that the uncertainties on the kinematic variables used in the fit are gaussian is not entirely valid. Especially the peak at 0 suggests that the tails of these uncertainties are not described well by a gaussian.
Further interesting distributions are the properties of the reconstructed top quarks and W bosons.
The Figures 5.8 to 5.11 show that the mass distributions of the reconstructed particles are affected the most by the fitting procedure while the distributions of the transverse momentum and the pseudo-rapidity remain relatively untouched. In the case of the pseudo-rapidity this might be due to the fact that the uncertainty on the spatial variables of the jets and leptons used in the fit are generally low compared to the momentum resolutions.
Therefore it might be easier for the KinFitter to vary the momenta rather than the spatial coordinates. Exemplary the relative alterations of the jets that are associated to the quarks of the hadronic W boson are shown in Figure 5.12.

Sources of Combinatorics

As every method the KinFitter has its certain shortcomings that affect the reconstruction of top pairs. The issue that the KinFitter might reconstruct the top pair using a false combination of jets and leptons is called combinatorics. The difficulties with combinatorics stem in part from the method itself, from the fit hypothesis and from the imperfections of the object measurements in general.
To begin with difficulties that may arise because of the uncertainties accompanying the object measurements will be considered. Since the KinFitter uses the p_T, η (θ for neutrinos) and ϕ coordinates of the physical objects in the event, the uncertainties of these variables have to be examined. In general the spatial coordinates are quite well measured as can be seen in Figure 5.12.
For the momenta the uncertainties for charged leptons are also quite small while the uncertainties on the jet and $E_\mathrm{T}^\mathrm{miss}$ momenta are about $10 - 30\,\%$ (taken from the covariance matrices of the study from Michelangelo Giorgi). These large uncertainties make the p_T variable less discriminative because it may be altered in a bigger range in

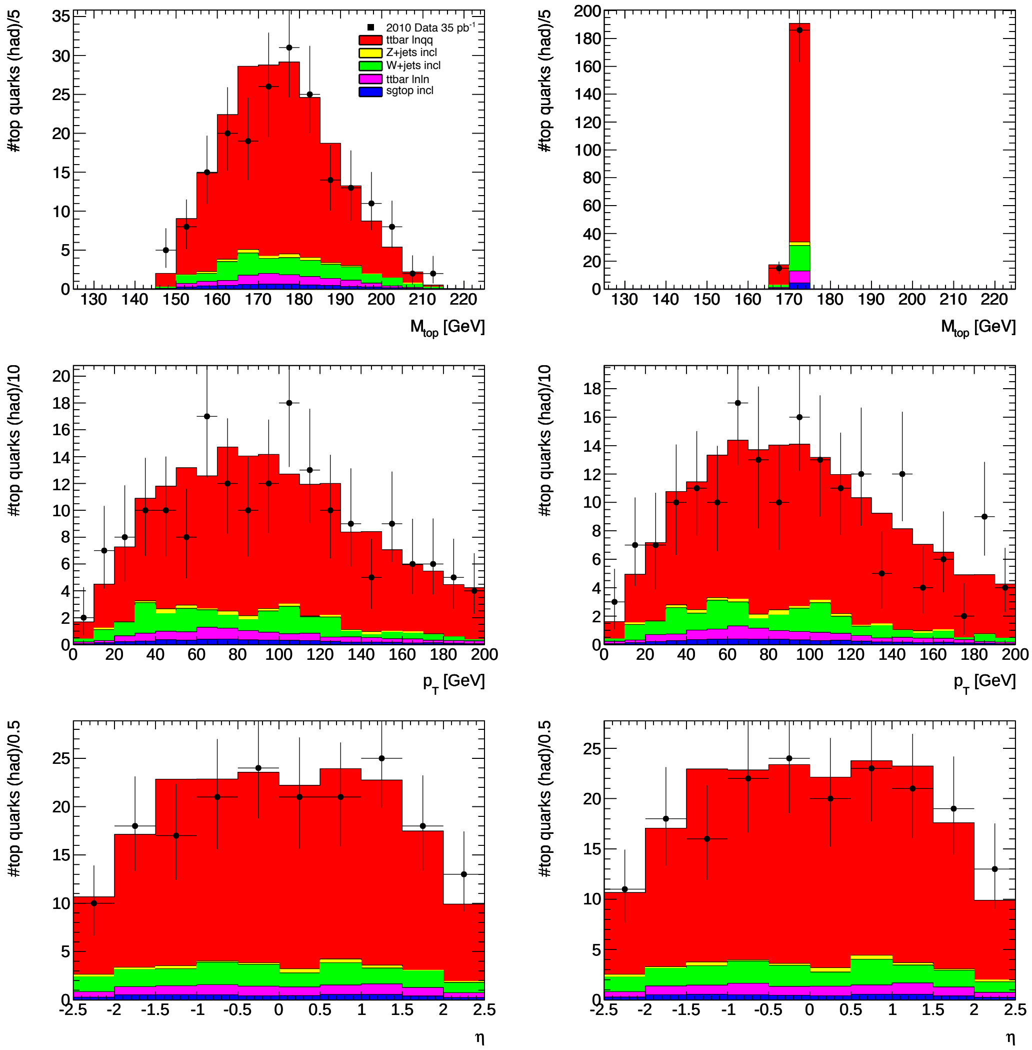

Figure 5.8: The distributions of the mass (first row), the transverse momentum (second row) and the pseudo-rapidity (third row) of the reconstructed top quarks (hadronic). The variables are calculated via the original variables (left) and the fitted variables (right) of the physical objects. The legend in the upper left plot applies for all plots.

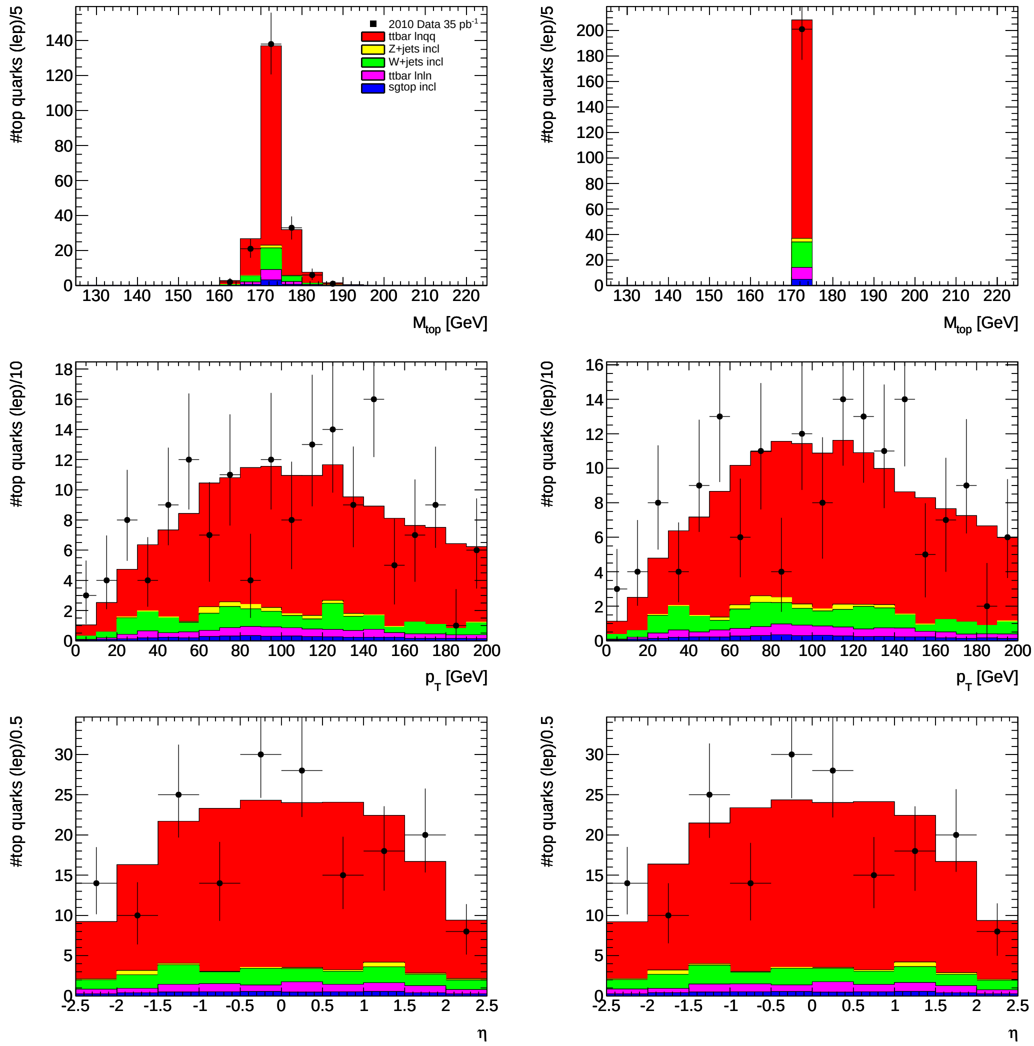

Figure 5.9: The distributions of the mass (first row), the transverse momentum (second row) and the pseudo-rapidity (third row) of the reconstructed top quarks (leptonic). The variables are calculated via the original variables (left) and the fitted variables (right) of the physical objects. The legend in the upper left plot applies for all plots.

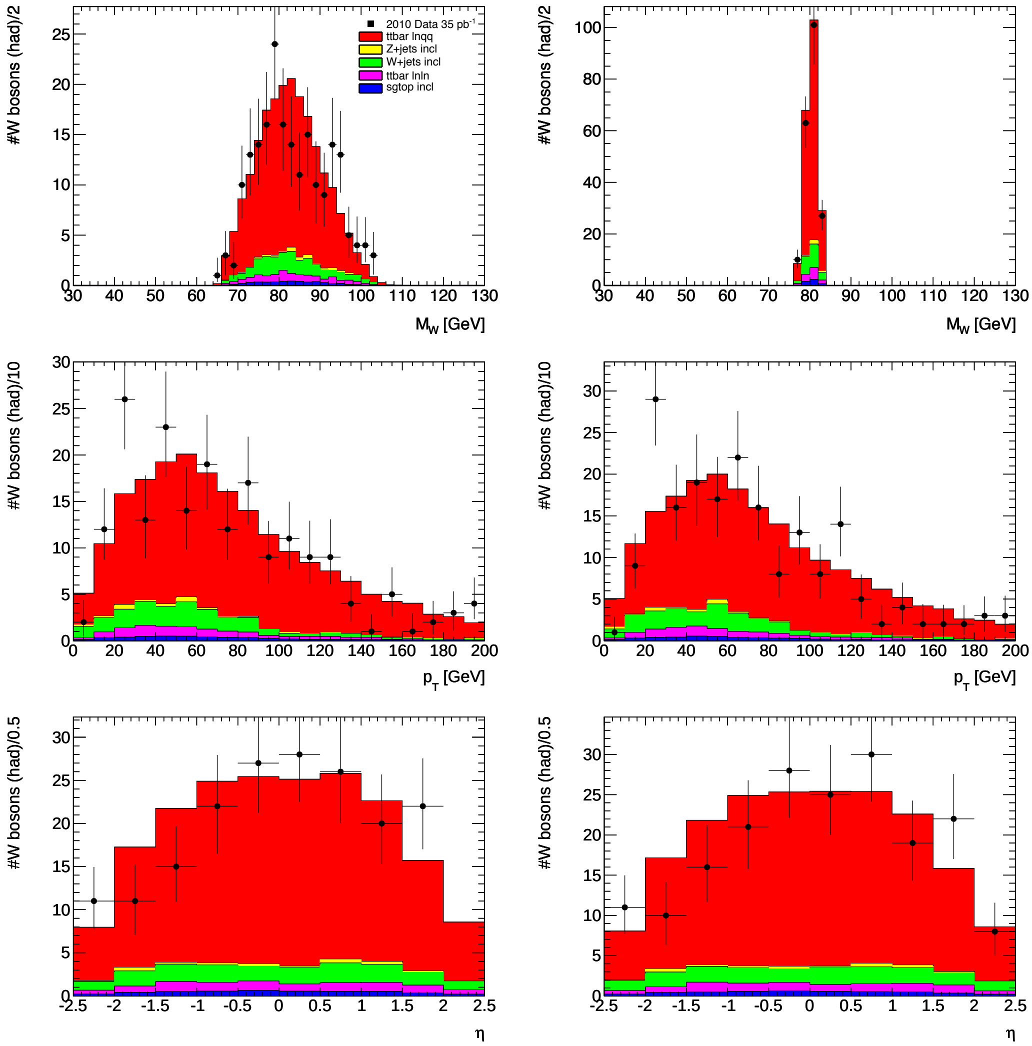

Figure 5.10: The distributions of the mass (first row), the transverse momentum (second row) and the pseudo-rapidity (third row) of the reconstructed W bosons (hadronic). The variables are calculated via the original variables (left) and the fitted variables (right) of the physical objects. The legend in the upper left plot applies for all plots.

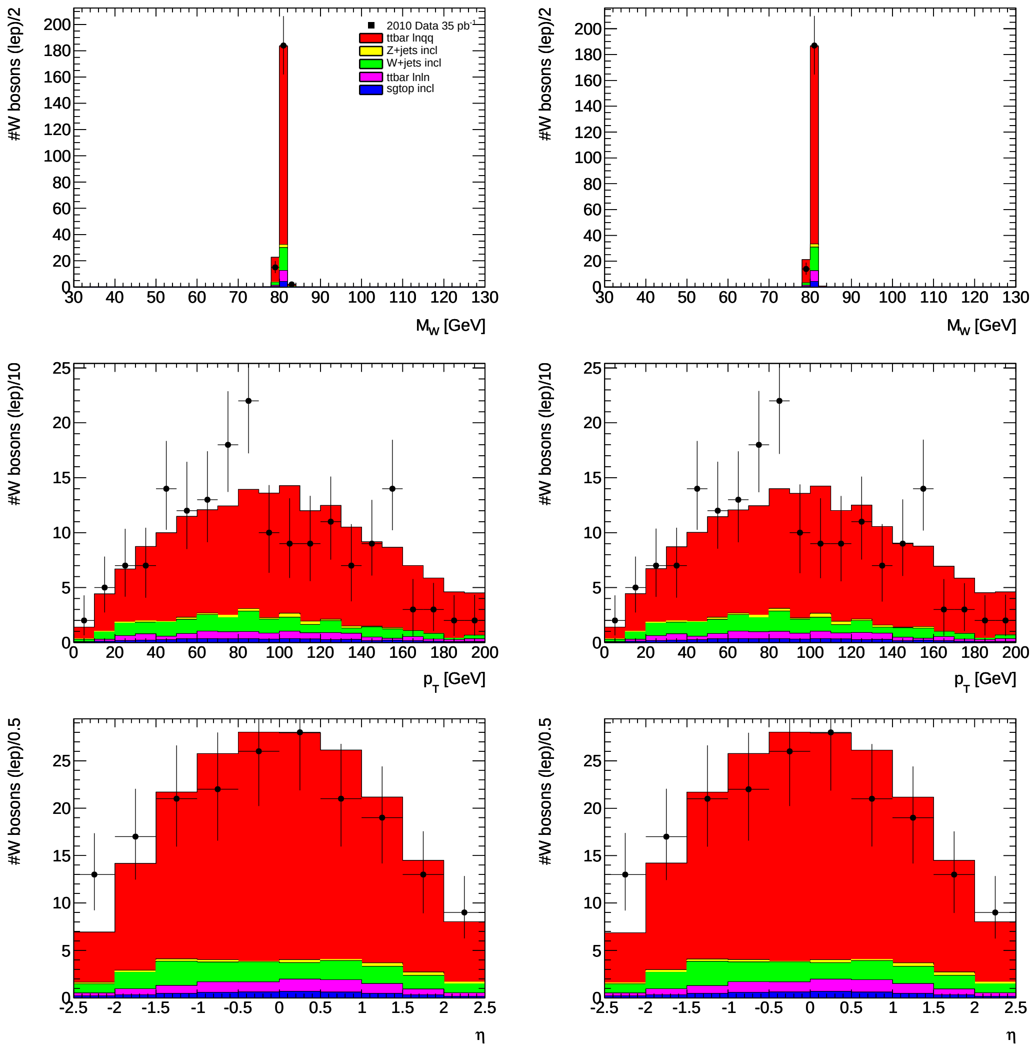

Figure 5.11: The distributions of the mass (first row), the transverse momentum (second row) and the pseudo-rapidity (third row) of the reconstructed W bosons (leptonic). The variables are calculated via the original variables (left) and the fitted variables (right) of the physical objects. The peak for the W boson mass calculated from the original variables stems from the neutrino reconstruction and is not of physical origin. The legend in the upper left plot applies for all plots.

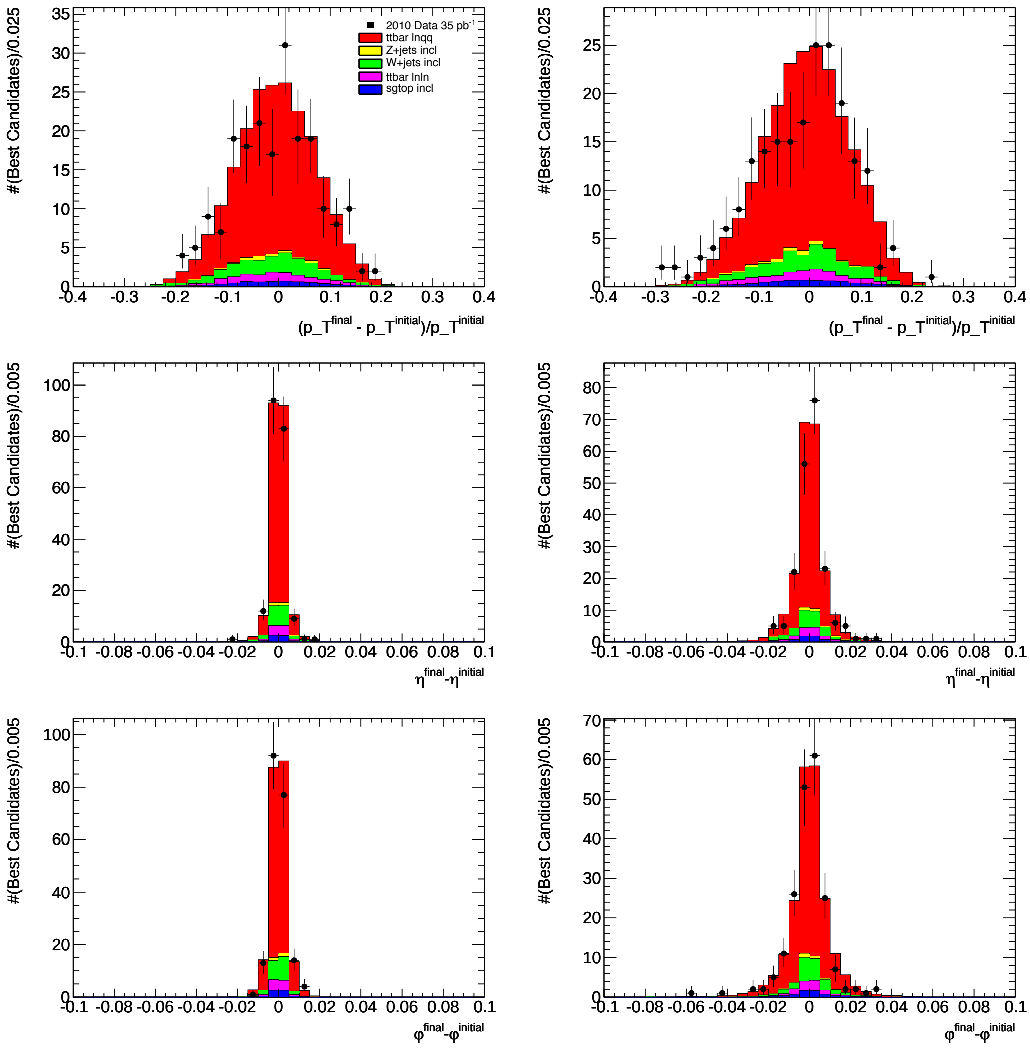

Figure 5.12: The distributions show the alterations of the kinematic variables of the jets from the decay of the hadronic W boson of the best candidate in a given event. The transverse momentum alterations (top row) seem to be significantly larger than those of the pseudo-rapidity (second row) and the azimuthal angle (third row). The plots on the left side are for the first jet the plots on the right for the second one. The legend in the upper left plot applies for all plots.

order to accommodate the given constraints. Four jets and one neutrino reconstructed from the E_T^{miss} are used in the fit making this issue rather important as it affects the majority of the fitted objects. Especially when considering high jet multiplicities the probability that jets not originating from the top pair may be included in the best top pair candidate increases because one of these false top pairs may coincidentally satisfy the constraints better than the correct top pair candidate.

In short the better the resolutions of the kinematic variables used in the fit are the more discriminating become the applied mass constraints. Hopefully with increasing understanding of the detector the resolutions will become smaller thus making the KinFitter more powerful. Then removing additional constraints such as the mentioned b-tagging constraint might be feasible.

Related to this is the possibility that badly measured objects with high uncertainties may cause the KinFitter to pick a false top pair candidate. As an example the following is considered: In an event there are 5 jets, one lepton and a neutrino. All objects are well measured with small uncertainties except for one jet which has a high uncertainty on its transverse momentum. Two top pair candidates built from the objects of the event are compared, one consists of only of the well measured objects while the other one includes the jet with the high uncertainty. Coincidentally the initial kinematic variables of the objects of both top pair candidates result in the same masses for the reconstructed W bosons and top quarks which are a bit off making their correction necessary.

The candidate with the well measured objects will be assigned a higher χ^2 than the one which includes the jet with the large uncertainty on the transverse momentum. This is caused by the high uncertainty because altering the transverse momentum for the badly measured jet does not result in such a high χ^2 as it does for a well measured jet whose transverse momentum is altered by the same value. A possible improvement would be a factor which takes this issue into account and removes this unfair competition advantage based on an undesired feature of an object. With increasing understanding of the detector that results in a homogeneous resolution this should not remain a problem and is even at this point a minor concern but should be mentioned in any case as a caveat.

Another issue is that only top pairs with four jets are reconstructed. In order to illustrate this the case of a jet splitting into two jets is considered. This splitting might be due to a jet emitting a particle which carries away a significant amount of momentum and subsequently forms a new jet or due to jet splitting caused by the reconstruction algorithm. In both cases the variables of the summed four vector would perform better in the fit than each of the two jets alone.

Of course the possibility of employing a fit hypothesis with five jets could be considered. Unfortunately this would severely raise the number of possible top pair candidates thus increasing the probability that a false combination yields the best fit candidate. Another reason is the computational effort that would result in the increasing number of top pair candidates. With this in mind only fit hypotheses with four jets were considered keeping in mind the drawbacks of this choice.

Beside these somewhat artificially introduced problems there are intrinsic drawbacks of the method which have to be dealt with. By applying the fit the information from a 17-dimensional phase space (three components for each of the six physical objects mi-

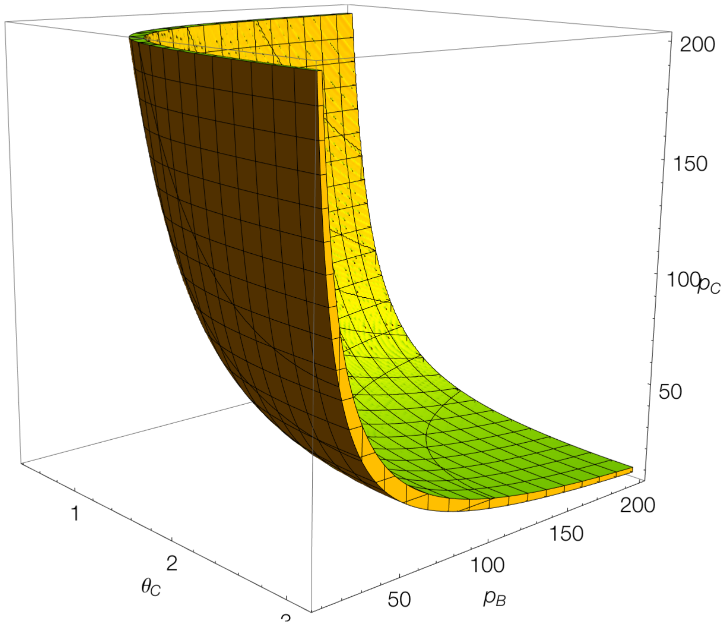

Figure 5.13: Figure showing the possible values for triples of $p_{\rm B}$ (x-axis), $p_{\rm C}$ (y-axis) and $\theta_{\rm C}$ (z-axis) that result in masses from 85 to 95 GeV. All triples lying in the colored volume are well suited for satisfying the given constraint making them equally good candidates.

nus the z-component of the neutrino momentum) is projected onto a one-dimensional cut variable. This involves a loss of information which leads to ambiguities.
The case where a particle A decays into two particles B and C is considered. The invariant mass of particle A is related to the four vectors of particles B and C via

$$M_{\rm A}^2 = P_{\rm A}^2 = (P_{\rm B} + P_{\rm C})^2. \tag{5.16}$$

After expansion and some elementary mathematics

$$M_{\rm A}^2 = M_{\rm B}^2 + M_{\rm C}^2 - 2 \cdot (E_{\rm B} \cdot E_{\rm C} - \vec{p}_{\rm B} \cdot \vec{p}_{\rm C}) \tag{5.17}$$

is obtained.
By applying the parameterization p (momentum), θ and ϕ to the Cartesian coordinates and neglecting the masses of B and C the formula calculating the invariant mass of particle A is

$$M_{\rm A}^2 = 4p_{\rm B}p_{\rm C} \cdot (\sin^2\left(\frac{\theta_{\rm B} - \theta_{\rm C}}{2}\right) + \sin^2\left(\frac{\phi_{\rm B} - \phi_{\rm C}}{2}\right) \sin(\theta_{\rm B}) \sin(\theta_{\rm C})). \tag{5.18}$$

The coordinate system may be adjusted in such a way that the θ component of B is always zero making the formula independent of ϕ. The resulting formula is

$$M_{\rm A}^2 = 4p_{\rm B}p_{\rm C} \cdot \left(\sin^2\left(\frac{-\theta_{\rm C}}{2}\right)\right). \tag{5.19}$$

All combinations of $p_{\rm B}$, $p_{\rm C}$ and $\theta_{\rm C}$ that satisfy the constraint result in an equal goodness of fit. This is illustrated for the case where $M = 90$ GeV in Figure 5.13.

In principle even background processes may exhibit final states where the physical objects have just the right kinematic variables that are necessary in order to obtain a top pair candidate with a very good fit result. Thus the KinFitter is not perfect in terms of background rejection.
Apart from physical background the mentioned issues play the biggest role when it comes to the misidentification of the right top pair. In order to correct both influences, one needs an algorithm that determines the influence from physical background very well and does not suffer from combinatorics.

5.4 Corrections

5.4.1 Correcting the Physical Background

The correction of the physical background relies on MC simulations of the different physics processes. By applying the same analysis to MC simulations predictions for the size of the individual physics processes are obtained which are compared to the results measured in real data. If the agreement between data and MC simulation is good enough it might stipulated that the MC simulations describe the data well and that the analysis behaves the same in data as in MC simulations. The contributions from the physical background as simulated in MC can then be subtracted from the measured data. This yields a data sample that presumably only consists of semileptonic top pair events.

5.4.2 Correction of the Combinatorial Background

Having obtained a clean data sample the issue of combinatorics remains to be solved. The basic task when striving to correct the combinatorial background is to find a method that establishes a clear and unambiguous relationship between a measured jet and a generator-level top quark.[14]
Unfortunately in MC simulations there is no simple connection (e. g. a map of links) between reconstructed jets on detector-level and the particles on generator-level. Therefore a jet cannot be asked whether it originates from the top pair or not. The need arises to define a method that establishes said connection.
The method to be developed has to have certain qualities. Firstly it should be sufficiently different from the KinFitter in order to avoid correlations between the methods which could introduce a bias. Secondly it should be as simple and robust as possible so that other studies may do crosschecks easily. Thirdly the method should operate on the same phase space as the KinFitter in order to maximize the comparability of the two algorithms.
It might be argued that "correctness" of the results (the top pair choice) of the correction method has to be demanded. The link between the reconstructed jets and the generated particles that will be established by the method is a definition of the correct link that cannot be checked for its correctness by other methods as they are themselves just definitions concerning the established link. That the results of the

[14] The leptons were not considered as they were not in the focus of the study.

particle	E [GeV]	η	ϕ [°]
q_1	119.45	- 0.39	113
q_2	57.21	- 1.25	177
b_3	53.66	0.47	- 144
l_4	21.82	- 0.21	- 1
ν_5	358.04	- 1.35	- 46
b_6	124.66	- 1.59	49

Table 5.5: The physical objects on generator level that are associated to the top pair. These data were extracted from event 20186 of the MC@NLo sample.

correction method are right is an assumption that has to be made on the ground of plausibility arguments.

In the following the method that was developed in the course of this study will be described and its advantages and disadvantages will be discussed.

The Proximity Algorithm

The method that was developed was dubbed the Proximity algorithm because it uses a ΔR criterion between generated particles and measured jets. It consists of several steps.

First, the MC genealogy tree of the event is read. From this the quarks and leptons originating from the top pair decay are obtained. This yields the particles on generator-level that will be used when applying the ΔR criterion.

In the next step top pair candidates are generated as described in the Section 5.3.3. Every jet in the top pair candidate has a role e. g. being the first light quark from the W_{had}. Thus it is possible to compare the generated particles and the jets of the top pair candidate by making jet-particle pairs based upon the role they fulfill.

The pairs are evaluated by measuring the ΔR between the generated particle and the measured jet. If the ΔR value of any pair in a top pair candidate is larger than 0.4, the top pair candidate is discarded. Of all top pair candidates that meet this constraint the one whose pairs give the smallest sum of the ΔR values is taken. This top pair candidate is then deemed the true top pair for the event in question.

Event 20186 is revisited for providing an example of the application of the Proximity algorithm. The used jets are the same as specified in Table 5.3.

Top pair candidates are built by permutation of the jets. The spatial coordinates of the jets are compared to those of the particles on generator level listed in Table 5.5. The results of the Proximity algorithm can be found in Table 5.6.

In this event the determination of the top pair via the Proximity algorithm appears unambiguous. Only three candidates have a $\sum \Delta R$ that is smaller than 2. Furthermore only one candidate satisfies the constraint that all jets of the candidate have a $\Delta R < 0.4$ to their associated particle on generator-level.

A comparison of the results of the KinFitter and the Proximity algorithm shows

cand	grp	q_1	q_2	b_3	b_6	$\sum \Delta R$	ΔR_{q_1}	ΔR_{q_2}	ΔR_{b_3}	ΔR_{b_6}
1	1	1	2	4	3	2.147	0.019	2.014	0.088	0.025
2	2	1	2	5	3	3.946	0.019	2.014	1.887	0.025
3	3	1	2	4	5	4.131	0.019	2.014	0.088	2.009
4	5	1	3	4	5	4.395	0.019	2.279	0.088	2.009
5	1	1	4	2	3	3.198	0.019	1.843	1.311	0.025
6	5	1	4	5	3	3.775	0.019	1.843	1.887	0.025
7	4	1	5	4	2	3.217	0.019	0.298	0.088	2.811
8	3	1	5	2	3	1.654	0.019	0.298	1.311	0.025
9	5	1	5	4	3	0.431	0.019	0.298	0.088	0.025
10	1	2	1	4	3	2.679	1.183	1.383	0.088	0.025
11	2	2	1	5	3	4.478	1.183	1.383	1.887	0.025
12	3	2	1	4	5	4.663	1.183	1.383	0.088	2.009
13	6	2	4	5	3	4.939	1.183	1.843	1.887	0.025
14	2	2	5	4	1	3.245	1.183	0.298	0.088	1.676
15	2	2	5	1	3	3.477	1.183	0.298	1.970	0.025
16	6	2	5	4	3	1.595	1.183	0.298	0.088	0.025
17	2	3	5	2	1	4.936	1.651	0.298	1.311	1.676
18	4	3	5	4	1	3.713	1.651	0.298	0.088	1.676
19	6	3	5	4	2	4.849	1.651	0.298	0.088	2.811
20	1	4	1	2	3	4.652	1.933	1.383	1.311	0.025
21	4	4	5	1	3	4.226	1.933	0.298	1.970	0.025
22	6	4	5	2	3	3.567	1.933	0.298	1.311	0.025
23	2	5	1	2	3	3.842	1.123	1.383	1.311	0.025
24	4	5	1	4	3	2.619	1.123	1.383	0.088	0.025
25	3	5	2	4	1	4.901	1.123	2.014	0.088	1.676
26	6	5	2	4	3	3.251	1.123	2.014	0.088	0.025
27	4	5	4	1	3	4.962	1.123	1.843	1.970	0.025
28	6	5	4	2	3	4.302	1.123	1.843	1.311	0.025

Table 5.6: Top pair candidates of event 20186 of the MC@NLO top pair sample. The variable "grp" identifies the group in which the candidate fits which is defined by the jet content of the candidate (compare candidates 1 to 3). The $\sum \Delta R$ is calculated from the individual ΔR. All 120 possible candidates were evaluated but only those with a $\sum \Delta R < 5$ were included in the table to improve readability. Candidate 9 is the best candidate. It corresponds to candidate 11 from Table 5.4. Marked in light green are the candidates with a $\sum \Delta R < 2$.

that their is a certain mismatch. The candidate with the jet signature 1543[15] is favored by the Proximity algorithm while the KinFitter prefers the candidate with the jet signature 4513. Both algorithms clearly disfavor the best candidate of their counterpart. The Proximity algorithm assigns a $\sum \Delta R$ of 4.226 with two jets clearly failing the individual $\Delta R < 0.4$ criterion. In return the KinFitter calculates a $P(\chi^2)$ of 0.004 for the best candidate of the Proximity algorithm, significantly lower than the cut value of 0.01.
Strictly speaking the KinFitter did not find the right result but it assigned the correct jets to the top pair. As it is only demanded that the jet content of the top pair has to be correct with the ordering being of secondary importance the KinFitter found the correct answer in this event.

Assumptions of the Algorithm and Caveats

The basic assumption of the Proximity algorithm is that the direction of flight of the generated particles is approximately the same as that of the measured jets they have caused. This is an often used assumption when generator-level objects are compared to detector-level objects. Measurement-wise this is supported by the very good spatial resolution of the detector. From the physical side this assumption holds for leptons quite well.
For jets the assumption is only valid if the jet retains the direction of flight of the particle it stems from. When the direction of flight of a jet is heavily altered, which may be caused by hard emission of a particle, the ΔR between the jet and the associated generator level particle will be larger. If the ΔR rises above 0.4 the given pair of measured jet and generator-level particle will cause the top pair candidate to be rejected. Thus the Proximity algorithm favors jets that do not exhibit hard radiation. While this affects the efficiency with which top pairs may be found it should not affect the correctness of the top pair determination.
In the case of high jet multiplicities it may happen that a jet coincidentally has the same direction of flight as the generated particle while the jet that originates from the generated particle fails to satisfy the ΔR criterion. This might lead to a false top pair determination.
Another case to be considered is the one of jets which were split by the jet reconstruction algorithm. As mentioned in Section 4.1.1 a jet may be split into two jets that are very close to each other if certain conditions are fulfilled. In principle both jets originate from the top pair but only one will be chosen as the hypothesis for the top pair determination still assumes four jets to be associated with the top. This is a problem similar to the one discussed for the KinFitter where the splitting of jets affects the top pair determination.

Applying the Correction of the Combinatorial Background

Ideally the situation after the event selection would be that both KinFitter and Proximity algorithm reconstruct a top pair for every event. In reality both algorithms provide top pairs only in a certain fraction of the events prohibiting a comparison of

[15] This means that jet 1 is q_1, jet 5 is q_2, jet 4 is b_3 and 3 b_6.

the two algorithms on an event-to-event basis.[16] Furthermore the proximity algorithm is not applicable on measured data because it relies on information about the event genealogy which is (unfortunately) only available in MC simulations.
To be able to still apply a correction the KinFitter is reinterpreted as a link between measured data and MC simulations that makes the application of the combinatorics free Proximity algorithm on measured data possible. The whole correction scheme is expressed via the formula

$$S_{\mathrm{Prox}}^{\mathrm{Data}} = ((S+B)_{\mathrm{KinFitter}}^{\mathrm{Data}} - B_{\mathrm{KinFitter}}^{\mathrm{MC}}) \cdot \frac{S_{\mathrm{Prox}}^{\mathrm{MC}}}{S_{\mathrm{KinFitter}}^{\mathrm{MC}}}. \tag{5.20}$$

Here S denotes the signal (semileptonic top pair events) while B denotes the physical background. The superscripts specify the source of the measurement while the subscripts specify the applied analysis algorithm. The term in parentheses represents the background correction as described in Section 5.4.1 yielding $S_{\mathrm{KinFitter}}^{\mathrm{Data}}$. On this the correction for the combinatorial background is applied using the assumption that both analysis algorithms would behave the same in measured data as in simulated data if the Proximity algorithm were applicable to data.
In practice the histograms that result from the application of both analysis algorithms are divided by another yielding correction factors on a bin-by-bin basis. These factors are then applied (again bin-by-bin) to the histograms from the physical background subtracted data thus providing a result that is free of combinatorics. The results of this correction scheme are situated in Section 6.

5.5 Determination of Heavy Flavor Fraction of Additional Jets

Using the corrected data sample the heavy flavor fraction of the additional jets may be determined.
There are several methods that may be used. While a method such as template fits is desirable it was not applied since the recorded data of 2010 does not sport enough top pairs with additional jets in order to produce reliable results via the method.
With the very promising results of the 2011 data taking (an amount up to two magnitudes higher than 2010 data is expected), methods such as template fitting will become feasible.
In this study the template fit method will be discussed and results will be given as proofs of concept.
More robust methods like a simple counting of bottom jets was not done as the 2011 data will quickly provide more precise results.

[16] It should be noted that the algorithms like the Proximity algorithm are often used for establishing the link between particles on the generator-level and measured objects. In studies where the performance e. g. purity of an analysis algorithm is determined only events can be used where the link providing algorithms delivers an answer. Therefore an evaluation of an analysis algorithm is restricted to a subsample characterized by the link providing algorithm. The performance outside of this subsample has to be assumed to be the same which might not be the case.

5.5.1 Template Fits

Template fitting is a technique used to determine fractions using discriminative variables. The following exemplary case is considered: Three groups of objects A, B and C which have discernible distributions (called templates) in variable x exist. The inclusive distribution of x with unknown fractions of the objects A, B and C is measured. Furthermore it is assumed that the inclusive distribution is fully caused by objects of the types A, B and C. Therefore the inclusive distribution of x should be a weighted sum of the individual distributions. The weights then reflect the amount of the types A, B and C in the sample from which the inclusive distribution was taken. These weights can be determined either using a binned likelihood fit, a χ^2 fit or some other fitting method.
In this analysis it was necessary to find a variable that is differently distributed for light jets, charm jets and bottom jets.
The discriminative variable used is the SV0 mass. This quantity is generated by the SV0 tagging algorithm during the b-tagging procedure. If the reconstruction of the secondary vertex in a given jet is successful, a value for the invariant mass of the secondary vertex is assigned to the jet. Figure 5.14 shows the distributions of the SV0 mass for the different flavors in the AcerMC signal sample (see Section 8.2). Here the determination of the flavor was done by accessing the Truth-flavor variable[17].
The fitting procedure used is described in [63] and implemented in the TFractionFitter class of ROOT [64]. It is basically an binned likelihood fit that takes into account that the templates themselves are statistically limited. Therefore the templates may be altered in the range of their statistical errors.
This can be dangerous if the binning of the templates is chosen too fine and the statistics in the individual bins becomes very low. Then the templates may be heavily altered to fit the inclusive distributions making the result unreliable.
Since this method is not yet trustworthy for additional jets because of the low statistics it is demonstrated here using all jets that passed both cut flow and Kinematic Fit. The templates (Figure 5.14) are obtained from the AcerMC signal sample. The template for jets whose Truth-flavor variable indicated that they were τ leptons were not in included as it severely lacked statistics. The inclusive distribution is the SV0 mass spectrum of jets from top as obtained from the AcerMC sample employing the KinFitter. It was chosen because it is known that the fraction of jets from b-hadrons has to be about 50 % providing a test for the template fit.
In Figure 5.15 the result of the template fit are given. The fractions have the expected sizes with the bottom fraction being close to 50 % and the light fraction about three to four times higher than the charm fraction indicating that the W boson decays equally

[17]The Truth-flavor variable is determined as follows: In a cone of $\Delta R = 0.3$ around the jet axis of a given jet the MC particles on generator-level are collected. These particles are checked with regard to their flavor in the order bottom quarks, charm quarks, tau leptons. If the set contains a bottom quark, the jet is deemed a bottom jet according to the Truth-flavor variable. If no bottom quark but a charm quark is found, the jet is deemed a charm jet. If neither a bottom nor a charm quark is found but a tau lepton, the jet is deemed a jet from a tau. The last case is where none of the checked flavors were found, then the jet is deemed a light jet meaning that it originates from light quarks, gluons or possibly unidentified sources.

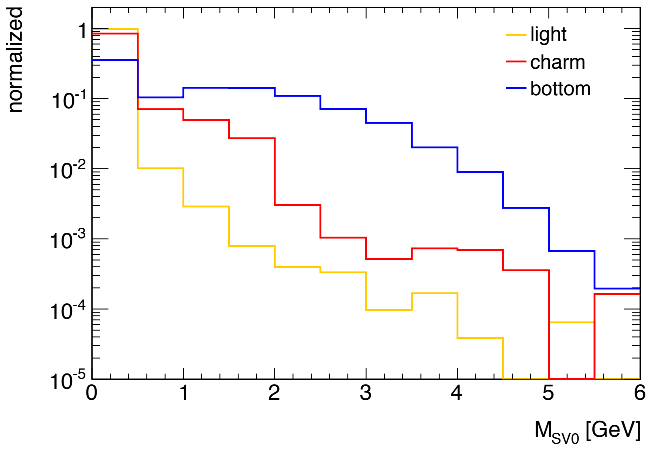

Figure 5.14: The templates as obtained from the AcerMC signal sample. All templates were normalized for shape comparison.

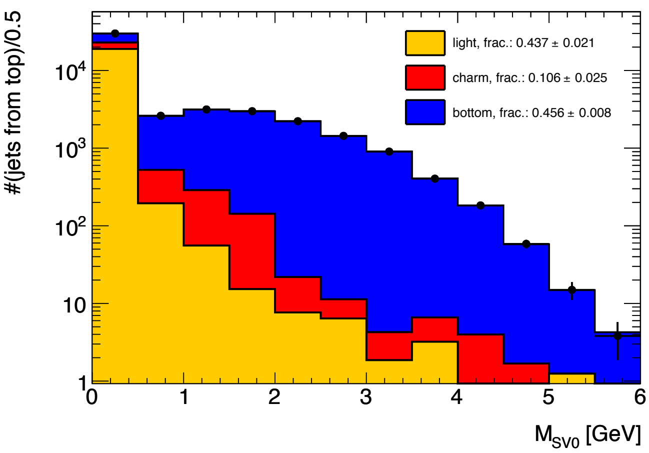

Figure 5.15: The result of the template fit applied on the SV0 mass spectrum of jets from top. The spectrum was obtained from the AcerMC signal sample via the KinFitter.

to pairs of up and down and strange and charm quarks.[18]
In the fit the constraint was applied that the three fractions have to sum up to one. Therefore a connection between the fractions is introduced. The parameter correlation coefficients are −0.96 between the light and charm fraction, 0.544 between the light and bottom fraction and −0.656 between the charm and the bottom fraction.
The fractions might be influenced by combinatorics e. g. the bottom fraction might be artificially lowered because another jet that did not originate from a b-hadron was tagged by the JetProb algorithm and resulted in a top pair without two jets from b-hadrons. Coincidentally the charm fraction might be increased as jets from charm quarks may fake jets from bottom quarks.
Further results using template fits can be found in Section 6.4.

5.6 Systematics

Apart from statistical uncertainties that affect an analysis there are systematic uncertainties which have to be kept in mind Especially in top pair analyses where high statistics will be reached in the near future, systematic effects will dominate.
Most of these uncertainties arise from a not yet perfect understanding of the detector and therefore the modeling of the detector response in MC simulations is flawed. These issues are being worked on and the systematic effects are getting smaller.
Some of the effects such as mismatches in reconstruction efficiencies are corrected by tools that are executed on MC samples prior to the analysis resulting in smaller systematical errors. In the following the systematics in question will be described, the application and existence of correction methods discussed and explained how the remaining systematic uncertainty was derived.
All systematics were evaluated by comparing the altered output caused by the application of the systematic with the nominal case. The deviations were taken as the systematical error. In cases where a systematic effect such as the one on the jet energy scale are evaluated by applying a set of alterations e. g. a down-shift and an up-shift the largest deviation of all alterations was taken.

5.6.1 Electron Systematics

Electrons both exhibit mismatches in energy resolution and reconstruction efficiency. The energy resolution in MC simulations is better than that observed in data. Therefore a tool is provided by the Electron/Gamma group called “EnergyRescaler” that applies corrections through a smearing procedure on the MC simulations. A factor binned in η of the cluster associated to the electron in question determines the width of a gaussian which is used for the smearing. The remaining uncertainty is evaluated by varying said factor in boundaries given by the tool.
For the reconstruction efficiency the top group has maintained its own tool due to the special isolation criteria. The mismatch between MC simulation and data is reduced by altering the event weights of the simulated events. This is achieved by determining a scale factor consists of three individual contributions concerning trigger

[18] Neglecting effects from the CKM matrix or other influences.

efficiency, reconstruction efficiency and identification efficiency. The first two are mere numbers independent of the kinematic variables of the electron in question. The third identification factor is binned in η and the energy of the cluster associated to the electron. The identification and reconstruction factors are added in quadrature The resulting RecoID scale factor and the scale factor for the trigger are then each multiplied with the event weight yielding an altered event weight.

The effect of the systematic uncertainty on the scale factor is evaluated by applying a smearing via a gaussian whose width is the uncertainty on the scale factor.

Corrections on the energy scale were not applied because they are small and the incorporation of the "EnergyRescaler" into the A++ framework (see Section 8.3) was postponed as it will no longer be applied in the 2011 MC simulations). The evaluation of the energy scale uncertainty was done using the values from the CSC analyses. Two analysis runs were done with the energies of all electrons multiplied with 0.995 and 1.005 respectively.

5.6.2 Muon Systematics

For muons the situation is similar to that of the electrons.

Mismatches in the transverse momentum resolution and scale of the muons are accounted for by applying a tool provided by the Muon Combined Performance Group. The smearing factors for the muon spectrometer and the Inner Detector are calculated from factors binned in η of the muon in question. The overall smearing of the combined muon is then calculated from these two smearing factors and applied to the transverse momentum of the muon.

The uncertainties accompanying these corrections are evaluated by varying the individual smearing factors within their errors.

The correction for the reconstruction efficiency is the aim of a top group based tool. This tool functions similar to that in the electron case. Here the reconstruction and the identification scale factors are mere numbers while the trigger scale factor is binned in η and ϕ of the muon.

The uncertainties on the scale factor are evaluated as in the electron case.

5.6.3 Jet Systematics

The jet energy resolution was corrected via the "JERProvider" tool. Depending on the η variable of a jet a reference histogram is consulted yielding the numbers for the noise, stochastic and constant terms that are used in the calculation for the smearing of the transverse momentum of the jet. The systematic on this correction is evaluated by switching the tool off and comparing the results.

Systematics affecting the jet energy scale (JES) could not be evaluated as the implementation of the JesUncertaintyProvider tool into the A++ was not finished in time.

The jet reconstruction efficiency was evaluated by randomly discarding jets with a probability of 2 % using the p_T independent value for the ICHEP conference 2010.

B-Tagging Systematics

The mismatch of the tagging efficiencies observed in data and MC is corrected via scale factors. A calibration file exists with the necessary information for the SV0 tagger with a b-tagging weight cut of 5.85 (50 % tagging efficiency) and the JetProb algorithm with weight cuts at 2.05 (70 % tagging efficiency) and 2.83 (50 % tagging efficiency).
Each jet in the energy region 20 GeV $< E <$ 500 GeV is evaluated and assigned a scale factor according to reference histograms. The product of the individual jet scale factors yields an event scale factor that is multiplied to the event weight. The remaining systematic effect is evaluated by varying the error on the scale factors inside their boundaries. Information on the calibration schemes employed may be found in [45].

6 Results

In this chapter the results of the analysis will be presented, beginning with the KinFitter analysis. Here all plots will show a comparison between the situation encountered in data and Monte Carlo (MC) simulations. Following this will be a discussion of the correction factors obtained from MC simulations and their application. Eventually the corrected data will be compared to the output of the Proximity algorithm in the semileptonic top pair sample obtained from MC simulations. Details on the uses MC simulation samples can be found in Section 8.2.
All plots containing data are normalized to $35\,\mathrm{pb}^{-1}$ if not stated otherwise.

6.1 Results of the KinFitter Analysis

First results stemming from the KinFitter had been shown in 5.3.3. Of central interest are the kinematic variables of the jets found in the event accepted by the KinFitter. In Figure 6.1 the jet multiplicity and transverse momentum distributions of these jets are shown. The discrepancies between the shapes of the jet multiplicities in MC simulations and data may stem from the fact that the AcerMC generator used for the top pair sample operates only at the leading order of perturbation theory. Another possible explanation may be flaws in the modeling of pile-up in MC simulations. Overall about two thirds of the events exhibit more than four jets.[1]
The transverse momentum distribution on the other hand seems to be described quite well in MC simulations. It is comparable to the transverse momentum distribution of

[1]These jets still satisfy the constraints that their transverse momentum exceeds 20 GeV and that their pseudo-rapidity lies in $|\eta| < 2.5$.

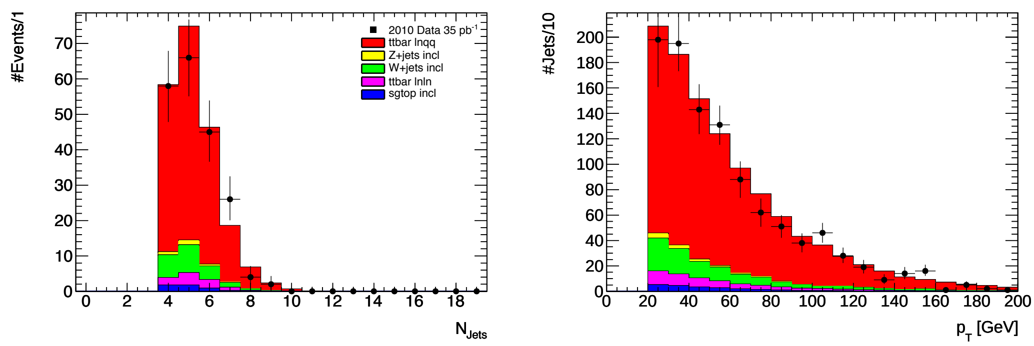

Figure 6.1: The distributions show the jet multiplicity (left) and the transverse momentum (right) of all jets in the events accepted by the KinFitter. The legend in the left plot applies for all plots.

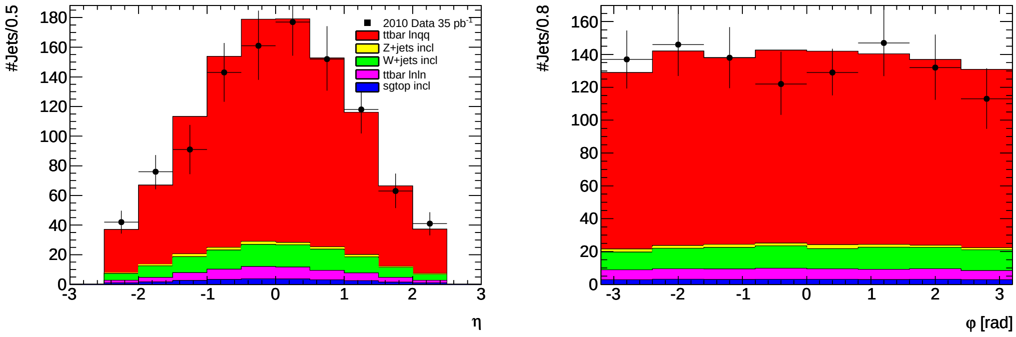

Figure 6.2: The pseudo-rapidity (left) and azimuthal angle (right) distributions of all jets in the events accepted by the KinFitter. The legend in the left plot applies for all plots.

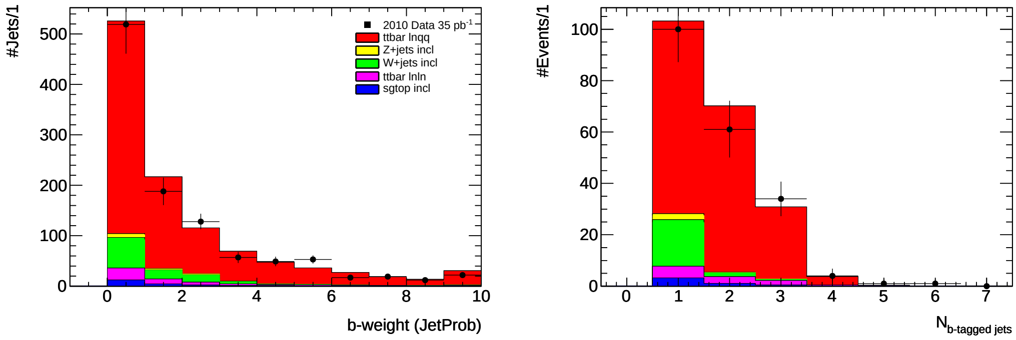

Figure 6.3: On the left is the distribution of the b-weights as calculated by JetProb of all jets in events accepted by the KinFitter. The multiplicity of the tagged jets is on the right for said events. The legend in the left plot applies for all plots.

the jets that comprise the input for the kinematic fit (see section 5.3.3) but seems to be slightly softer.

The spatial variables η and ϕ of the jets shown in Figure 6.2 exhibit a good agreement between MC simulations and data and behave as expected. While the jets from top pairs (in the following called jet from top) are centered around $\eta = 0$ the additional jets are distributed more flatly. The observed distribution is a superposition of both cases with jets from top being dominant causing the peak.

The symmetry of the detector around the azimuthal angle is reflected in the distribution which is essentially flat.

The topology of top pair decays exhibits two jets from b hadrons. In Figure 6.3 the b-weights of the jets as well as the number of tagged jets in the events accepted by the KinFitter are presented. All jets with a b-weight above 2.05 are counted as jets from b-hadrons. The discrepancies encountered in the jet multiplicity as seen in Figure 6.1 seem to translate to the multiplicity of the tagged jets. In general the variables sensitive to the flavor of the jets seem well modeled in MC simulations.

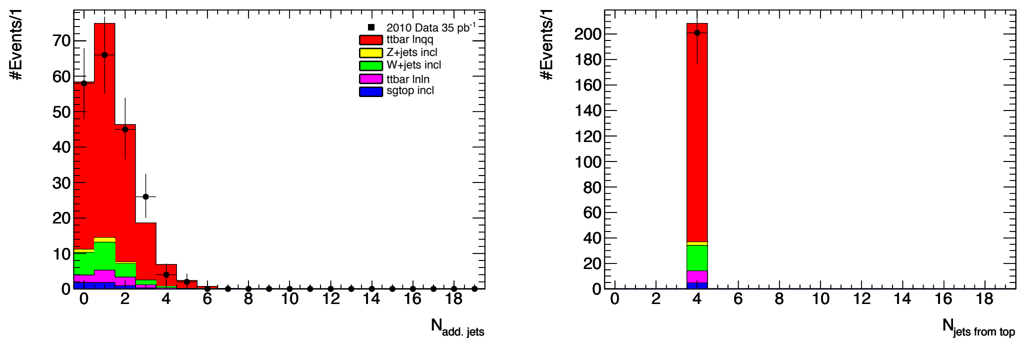

Figure 6.4: The jet multiplicity distributions for the additional jets (left) and the jets associated to the top quarks (right). The legend in the left plot applies for all plots.

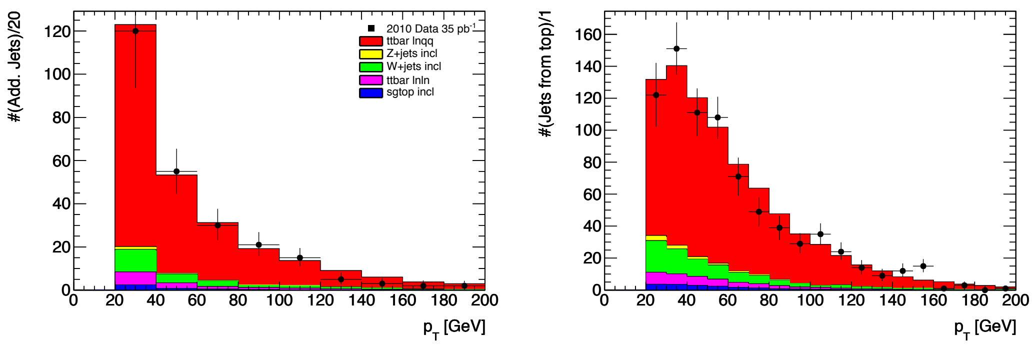

Figure 6.5: The distributions for the transverse momentum for the additional jets (left) and the jets associated to the top quarks (right). The legend in the left plot applies for all plots.

6.1.1 Results for Additional Jets and Jets From Tops

The main use of the KinFitter in this analysis is the separation of the jets in the accepted events into the groups "additional jets" and "jets from tops". After the reconstruction of the top pair these two groups of jets can be compared.

In Figure 6.4 the jet multiplicities of the two groups are shown. As in the multiplicity plot of the tagged jets the additional jets exhibit a discrepancy in shape between MC simulation and data. Most of the events include zero, one or two additional jets encouraging the assumption that the combinatorial background might manageable. In events where both top finding methods yield a result it is observed that with the setup chosen in this analysis the two algorithms agree in $\approx 50\,\%$ of the cases. As mentioned before this does not necessarily extrapolate to the whole phase space.

In the plot of the jet multiplicity for the jets from tops the fit hypothesis is reflected by the fact that there are only entries in the 4-jet bin.

The transverse momenta of the two jet groups are presented in Figure 6.5. The jets from top have on average higher transverse momenta than the additional jets. This might be explained by the origins of the respective jets. The jets from tops originate

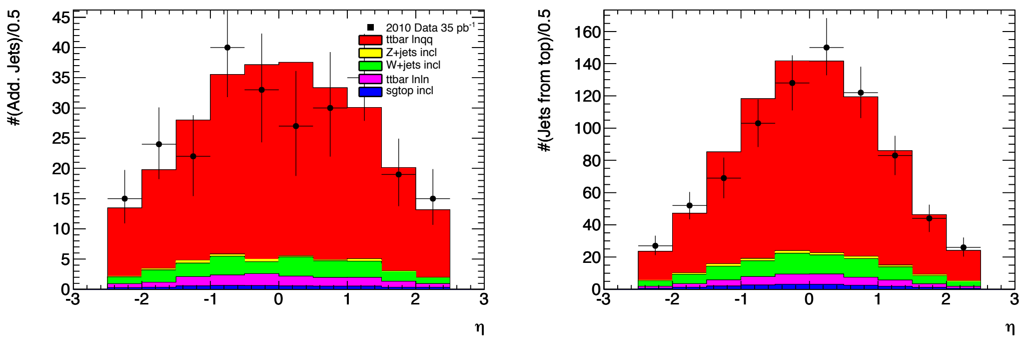

Figure 6.6: The distributions for the pseudo-rapidity for the additional jets (left) and the jets associated to the top quarks (right). The legend in the left plot applies for all plots.

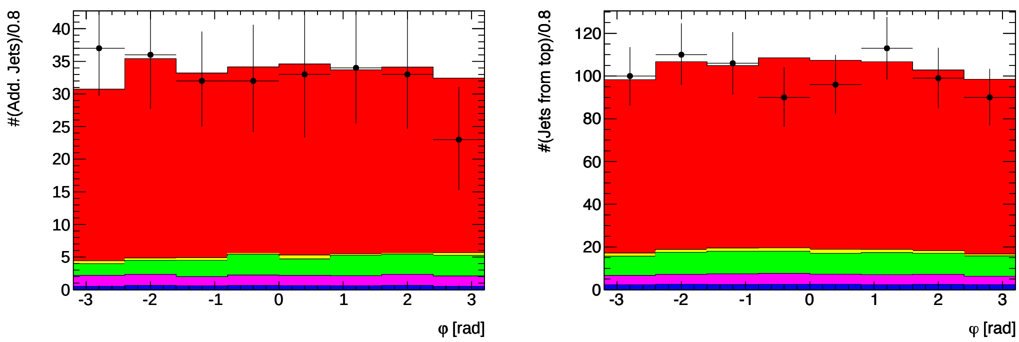

Figure 6.7: The distributions for the azimuthal angle for the additional jets (left) and the jets associated to the top quarks (right). The legend in the left plot applies for all plots.

from top and W boson decays. Due to the high masses of these particles the jets are produced with high momenta. Furthermore the KinFitter might prefer jets which have a high transverse momentum as the top pair reconstruction may be easier using such jets (see Figure 5.13).

The distributions of the pseudo-rapidity η (Figure 6.6) also show clear differences between the two jet groups. Both groups show a peak around zero but in the jets from tops case it is far more pronounced. This again hints at the origin because the hard scattering process produces jets that are concentrated in the central η region while jets from other sources such as initial state radiation are distributed more flatly in η.

The only distribution without any major differences between the two groups is the one of the azimuthal angle ϕ shown in Figure 6.7. Here both groups exhibit the expected behavior that the distributions are flat.

By construction the jets from top should on average have higher b-tagging weights than the additional jets because the KinFitter accepts only candidates that contain

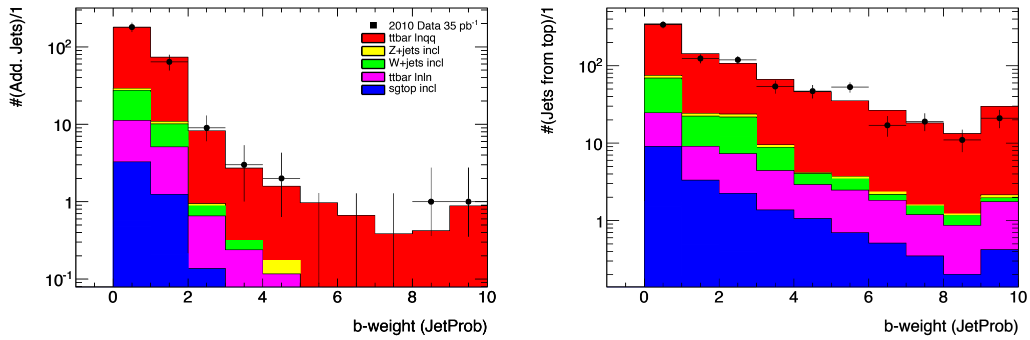

Figure 6.8: The distributions of the b-tagging weights as calculated by the JetProb algorithm for additional jets (left) and jets from tops (right). The legend in the left plot applies for all plots.

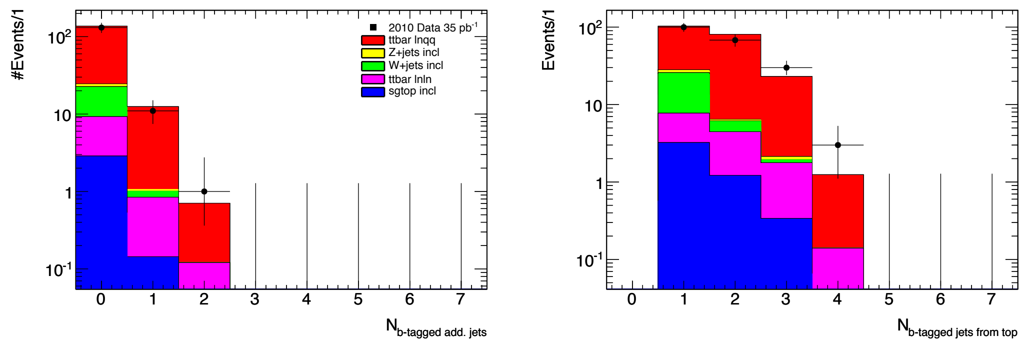

Figure 6.9: The jet multiplicity distributions of tagged jets for the additional jets (left) and the jets associated to the top quarks (right). The legend in the left plot applies for all plots.

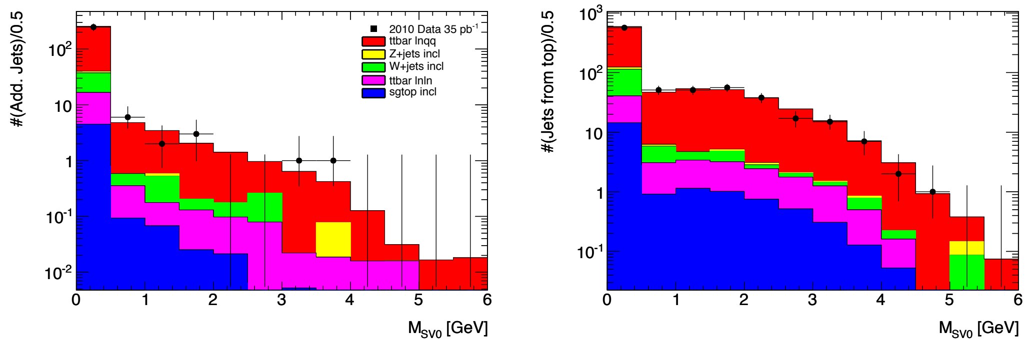

Figure 6.10: The SV0 mass distributions for the additional jets (left) and the jets associated to the top quarks (right). The legend in the left plot applies for all plots.

at least one jet meeting the b-weight cut. The validity of this assumption is backed by Figures 6.8 and 6.9.

Again it should be noted that in terms of the determination of the heavy flavor fraction of additional jets the usage of b-tagging criteria in the analysis may lead to biases. At the current point where systematics on the kinematic variables, especially the calibration of the jets, are not fully understood the application of a b-tagging information in the top pair candidate building is still necessary (see discussion in Section 5.3.3).

The core of this problem is finding the balance between a too hard b-tagging criterion that artificially depletes the heavy flavor content of the set of additional jets and no criterion at all which leads to a bad reconstruction of top pairs and thus to an artificially increased heavy flavor content through migration of bottom jets into the group of additional jets.

It has been observed during the studies that the performance of the KinFitter depends only weakly on the chosen b-tagging efficiency of the tagging algorithm in question. In general it suffices to apply any b-tagging technique at all to raise the fraction of correctly reconstructed top pairs by about 10 %.

In the future when the detector understanding has improved the kinematic variables of the physical objects might be measured with sufficient precision to afford the omission of the b-tagging criterion.

Figure 6.10 shows the SV0 mass distributions for the two groups of jets. In both cases the measured data and the MC simulation are in good agreement. This is especially important when applying the method of template fits since the templates taken from MC simulations are fitted to the inclusive distribution taken from measured data. A good description of the variables used in the fit is key.

The lack of statistics that is apparent in the plot for the additional jets makes the application of the template fit method in 2010 data impossible. The prospects for 2011 data on the other hand look very promising as the integrated luminosity of the acquired data will be larger by a factor of about 50.

On the other hand the statistics for distribution of the jets from top are sufficiently high for template fitting. In Section 6.4 the results from the a test of the technique

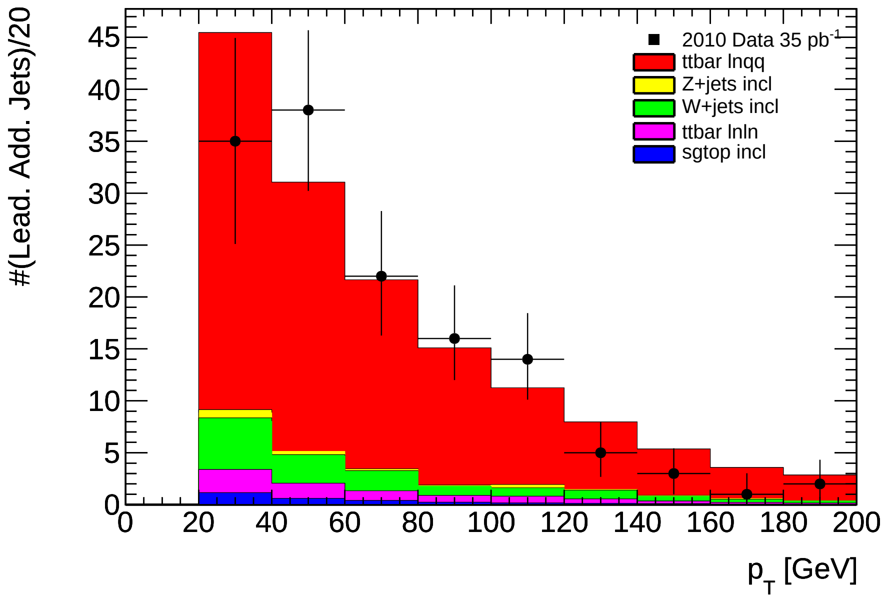

Figure 6.11: The transverse momentum of the leading additional jet.

applied to the jets from top will be shown.

6.1.2 Results for the Leading Additional Jet

The following presentation of the properties of the leading additional jet will conclude the first part of the results.

In Figure 6.11 the transverse momentum of the leading additional jet is depicted. Although somewhat softer than that of the jets from top the momenta exhibited are in a viable range for faking jets stemming from a top pair decay. In [58] an attempt was made to apply further cuts on the transverse momentum of the jets used for the top pair determination in order to both reduce computing time and combinatorial background.

This was not done in this study as it might have introduced additional shifts in the transverse momentum distribution towards lower values because jets that not meeting the cut for a jet stemming from a top are automatically deemed additional although they might contribute to a good top pair candidate (if not the best in the event).

The spatial variables η and ϕ of the leading additional jet are shown in Figure 6.12. The azimuthal angle distribution behaves as expected and is flat across the whole range.

The η distribution exhibits a peak around the center. In its relative height the peak lies between those of the additional jets and the jets from tops as can be seen in Figure 6.6. This might stem from combinatorics or a correlation between the height of the transverse momentum and the origin of the additional jet. It can be argued that additional jets stemming from processes such as final state radiation might have a generally higher transverse momentum than additional jets from other sources whose

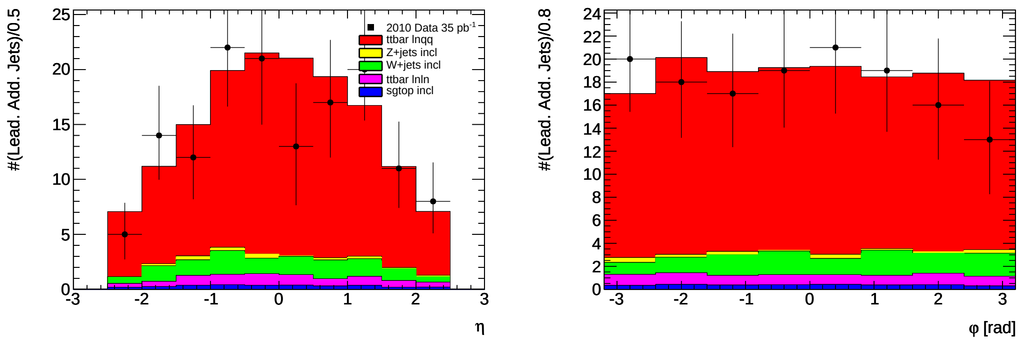

Figure 6.12: The spatial variables η (left) and ϕ (right) of the leading additional jet. The legend in the left plot applies for all plots.

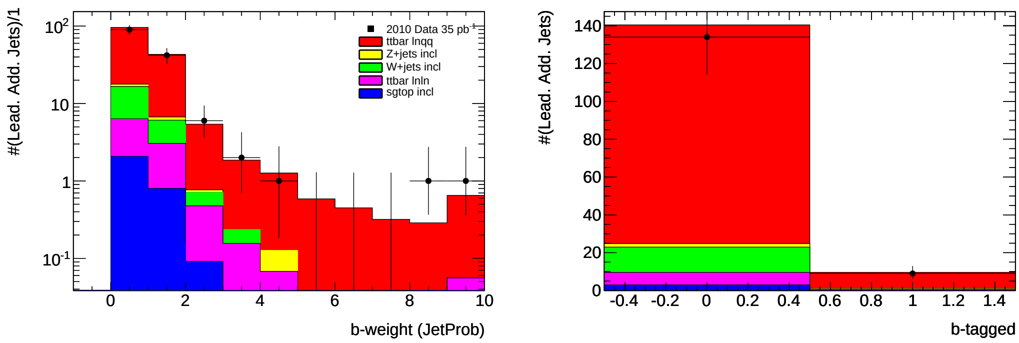

Figure 6.13: The b-tagging weights computed by JetProb (left) and tagging decision (right) for the leading additional jet. The legend in the left plot applies for all plots.

η distribution does not exhibit a peak.

The tendency of additional jets towards lower b-tagging weights is also reflected by the leading additional jet as can be seen in Figure 6.13. In general it can be said that if an additional jet is accepted by the b-tagging algorithm the probability is high that it is the leading additional jet. Nevertheless the additional jets are dominated by light jets[2].

Both distributions of the SV0 mass shown in Figure 6.14 discourage the application of the template fit method on the leading additional jets. Again the statistics in 2010 data is too low but this issue will be gone in an analysis of 2011 data.

Figure 6.15 concludes the results on the leading additional jet. The distribution shows the smallest ΔR between the leading additional jet and the nearest jet of any kind (respectively the nearest jet stemming from the top pairs). Both show that the final states of top pair events are rich with a multitude of objects present.

[2]In the following, jets originating from bottom quarks will be called bottom jets whereas jets from lighter quarks (up, down, strange) and gluons will be called light jets. If not mentioned otherwise jets from charm quarks will count as light jets. In all other cases they will be called charm jets.

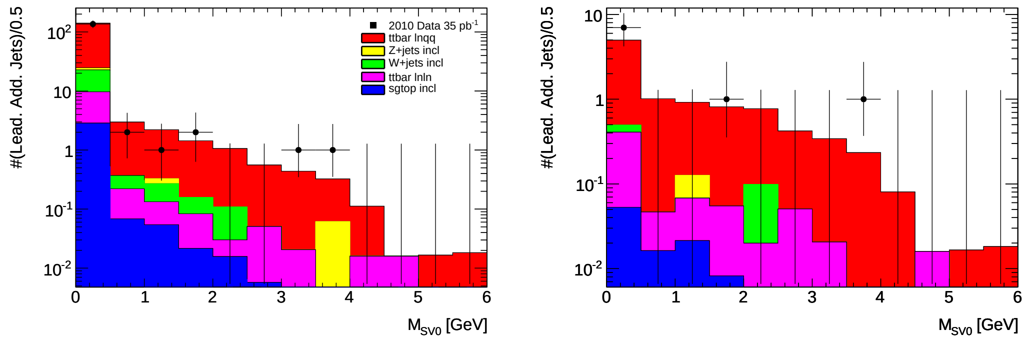

Figure 6.14: The SV0 mass distributions for the leading additional jets without b-tagging applied (left) and with a b-tagging criterion (right). The legend in the left plot applies for all plots.

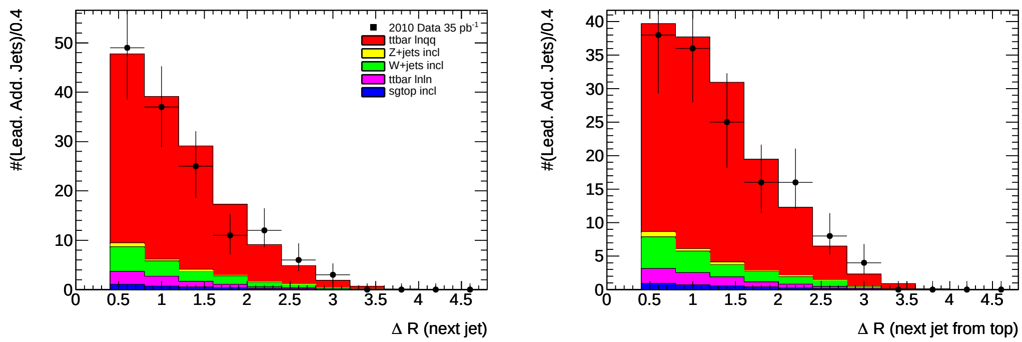

Figure 6.15: The smallest spatial distance of the leading additional jet to any other jet (left) or the nearest jet stemming from the top (right). The legend in the left plot applies for all plots.

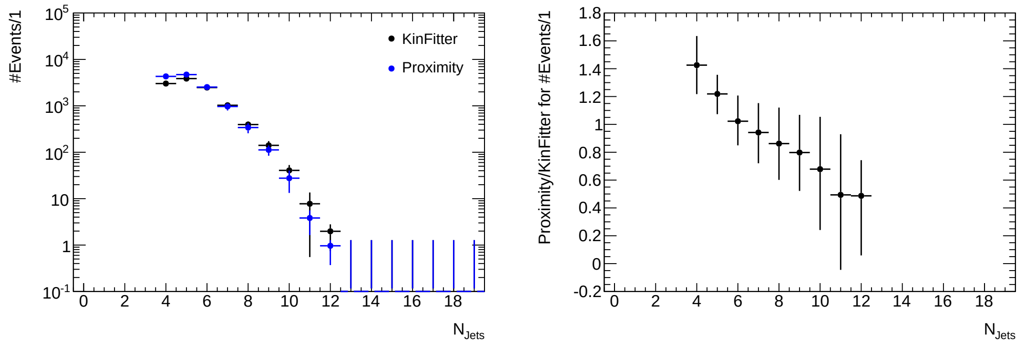

Figure 6.16: Comparison of the jet multiplicity of all jets returned by the top reconstruction algorithms (left) and the ratio of the outputs (right).

6.2 Comparison of the Output of the KinFitter and the Proximity Algorithm

The following plots contain the outputs of the KinFitter and the Proximity algorithm. First the results concerning all jets in the events that were accepted by the algorithms will be compared. By studying these outputs general statements about the behavior of the KinFitter with respect to the Proximity algorithm can be made.
Both algorithms were applied on the semileptonic top pair sample simulated by AcerMC. The number of semileptonic top pair events was initially 163337 and 24504 after application after the cut based selection omitting the decision of the top reconstruction algorithm.

6.2.1 Distributions for All Jets in the Events Accepted by the Algorithms

The jet multiplicity of all jets in events accepted by both algorithms is displayed in Figure 6.16. The KinFitter found a top pair in 10979 events while the Proximity algorithm yielded a result for 13007 events. Therefore an overall ratio between the two algorithms of 1.18 is computed. Here and in the following the ratio is defined as the value returned by the Proximity algorithm as the nominator and the KinFitter result as the denominator. In the right-hand plot of Figure 6.16 the declining character of the ratio with increasing jet multiplicity is clearly visible.
This suggests that in high multiplicity events the KinFitter may accept more events because there is such an abundance of top pair candidates that at least one of them satisfies the $P(\chi^2)$ criterion though it may not be correct top pair.
The situation in the four jet bin is reversed. Here the Proximity algorithm yields more events. The KinFitter might reject the right top pair candidates[3] because all fits yield low $P(\chi^2)$ values.

[3]There are only right ones as the correct top pair is defined by its jets content not by the ordering. Events with four jets are in principle a trivial case for the KinFitter.

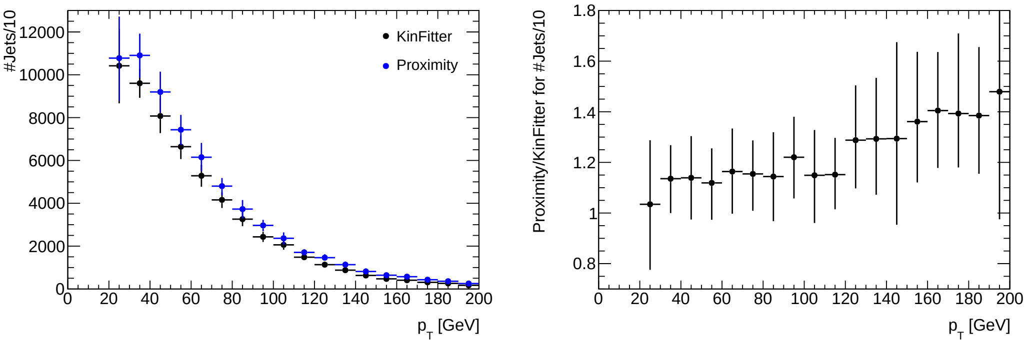

Figure 6.17: Comparison of the transverse momentum of all jets returned by the top reconstruction algorithms (left) and the ratio of the outputs (right).

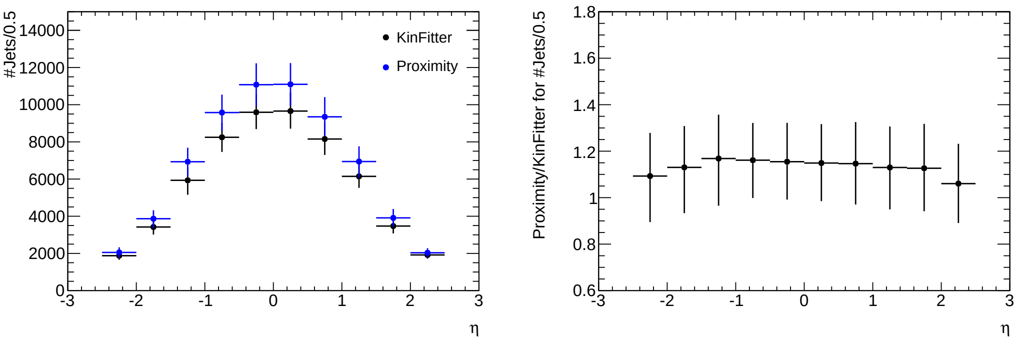

Figure 6.18: Comparison of the pseudo-rapidity of all jets returned by the top reconstruction algorithms (left) and the ratio of the outputs (right).

Figure 6.17 shows the transverse momentum distribution of the jets accepted by the individual algorithms. The KinFitter finds 58419 jets in the events where it found a top pair while Proximity Algorithm yields 66838 jets giving an overall ratio of 1.14. The shapes of both algorithms are slightly different. The spectrum as extracted by the Proximity algorithm is slightly harder while the KinFitter seems to prefer jets with intermediate p_{T}. This is possibly linked to the way the KinFitter works. Jets with intermediate momenta have a higher freedom concerning their spatial coordinates. In Figure 5.13 the shell that indicates viable triples of values is considerably thicker for intermediate values than for extreme ones. Therefore it might be easier for the KinFitter to reconstruct top pairs with a high enough $\mathrm{P}(\chi^2)$ using jets with moderate transverse momenta.

In general it is assumed that the pseudo-rapidity of jets from top pairs is distributed more centrally while additional jets exhibit a broader and flatter distribution. Figure 6.18 shows the η coordinate for all jets in the events accepted by the algorithms. The output of the Proximity algorithm exhibits a slightly more pronounced peak than the KinFitter. The observed η spectrum is a superposition of both individual distributions and its concrete shape is governed by the ratio between jets from top and additional jets. Here the jets in the events accepted by the KinFitter seem to have a higher fraction of additional jets compared to the ones from the Proximity algorithm.

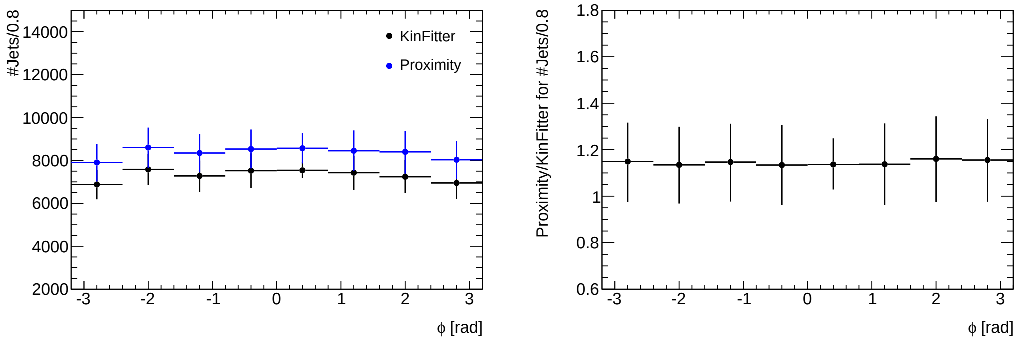

Figure 6.19: Comparison of the azimuthal angle of all jets returned by the top reconstruction algorithms (left) and the ratio of the outputs (right). The legend in the left plot of Figure 6.18 applies for the plots.

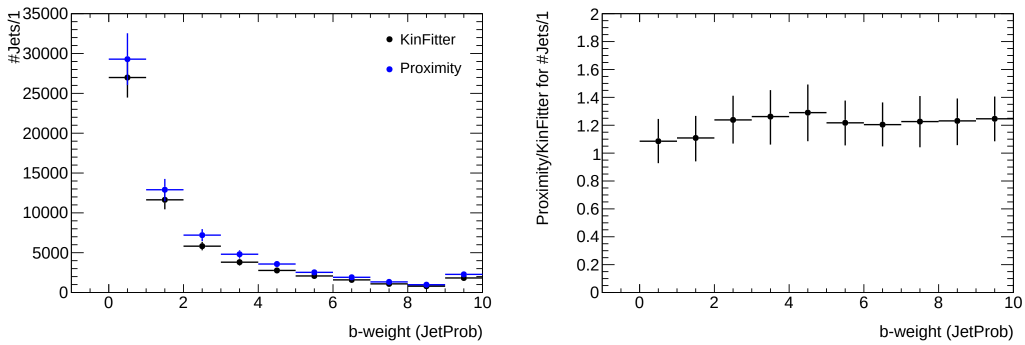

Figure 6.20: Comparison of the b-tagging weights of all jets returned by the top reconstruction algorithms (left) and the ratio of the outputs (right).

The symmetry around ϕ is reflected in Figure 6.19. Both algorithms yield flat distributions.

Beside the kinematic variables the information concerning b-tagging are of central importance as this study is also geared towards a determination of the heavy flavor fraction of additional jets in semileptonic top pair events. In Figure 6.20 the weights of all jets in the accepted events as computed by the JetProb algorithm are depicted. The jet content selected by the KinFitter seems to have a higher fraction of light jets as indicated by the ratio plot. This seems to be again an effect that stems from the fact that the KinFitter prefers events with high jet multiplicities. Additional jets are predominantly light jets. The jets from the top pairs are two light jets and two jets from b-hadrons. Therefore the more additional jets there are in an event the more the fraction between light jets and bottom jets will be shifted towards light jets which in turn affects the b-tagging weight distribution.

Closely related to this is the multiplicity of b-tagged jets in the events accepted by both algorithms depicted in Figure 6.21. The mismatch between the algorithms in the 1-jet bin is particularly interesting. A possible source for this mismatch is that bottom quarks might lie outside the jet acceptance region of $|\eta| < 2.5$. Therefore the jet that belongs to the bottom quark is not accessible by the Proximity algorithm

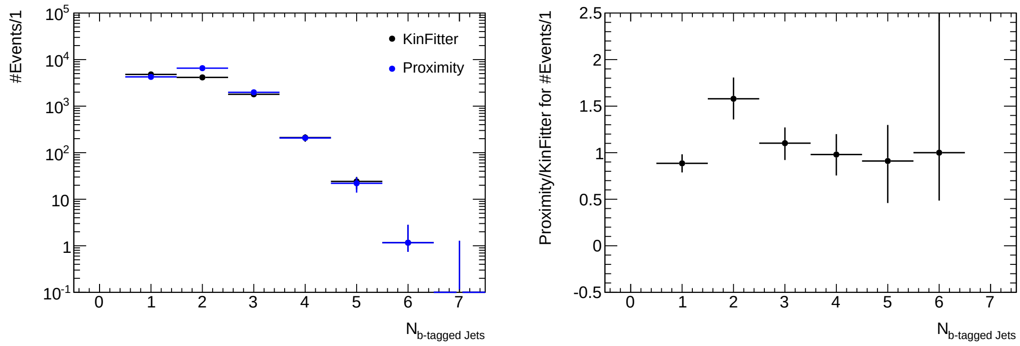

Figure 6.21: Comparison of the multiplicities of jets tagged by the JetProb algorithm in events accepted by the top reconstruction algorithms (left) and the ratio of the outputs (right).

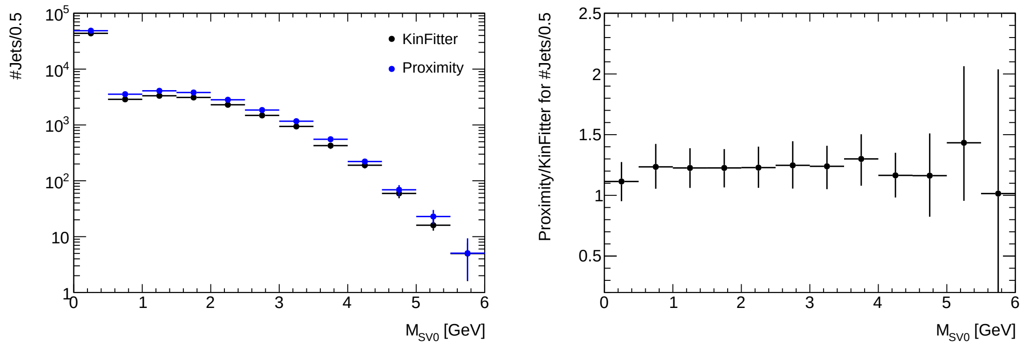

Figure 6.22: Comparison of the SV0 masses of all jets in events accepted by the top reconstruction algorithms (left) and the ratio of the outputs (right).

which causes the algorithm to reject the event as it cannot match any jet to the bottom quark.

In the 2-jet bin the situation is reversed. Here the proximity algorithm yields more events than the KinFitter. The two b-tagged jets suggest that there are two bottom quarks in the η range where b-tagging techniques are applicable. Therefore the argument for the 1-jet bin may be reversed: With the bottom quarks lying inside the jet acceptance region the Proximity algorithm is able to reconstruct the top pair. The KinFitter on the other hand may have difficulties reconstructing a top pair because the event has a low jet multiplicity (see Figure 6.16).

It is surprising that the ratio of the two algorithms for the SV0 mass of the jets is quite flat given the deviations observed in the other b-tagging related variables. The main difference seems to lie in the first bin that is mostly populated by light jets. There jets are collected where either no SV0 mass could be determined because the SV0 tagging algorithm failed to reconstruct a secondary vertex or the SV0 mass is very small.

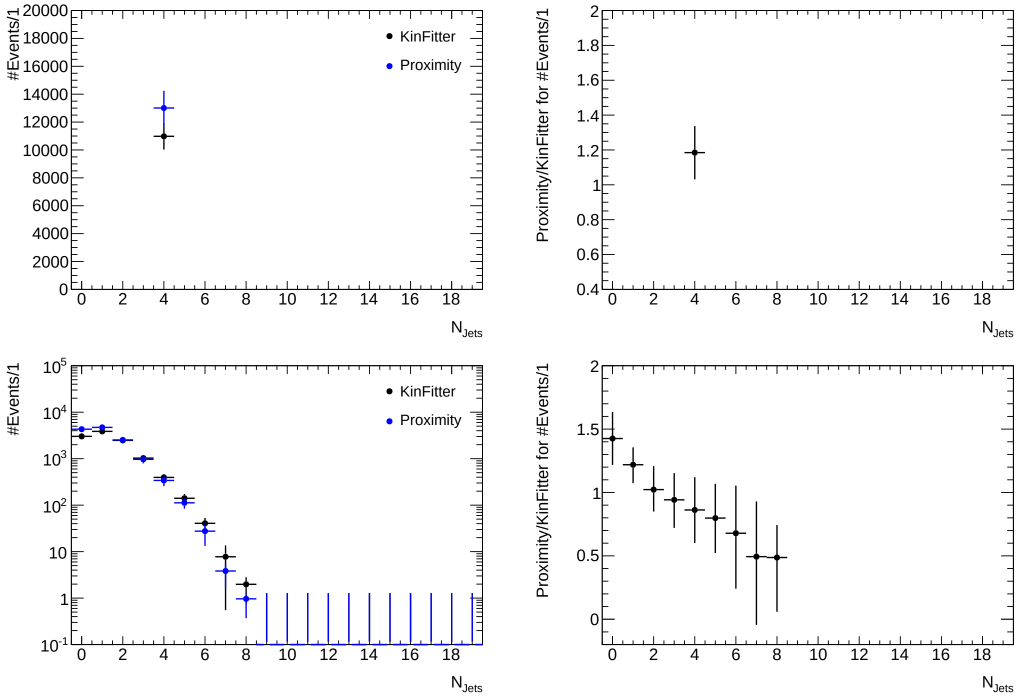

Figure 6.23: Comparison of the jet multiplicity of the jets from top (upper row) and additional jets (lower row) in events accepted by the top reconstruction algorithms (left) and the ratio of the outputs (right).

6.2.2 Distributions for the Jets from Top and Additional Jets

While the distributions of the variables of all jets in the events accepted by the top finding algorithms gave information about the general character of the selected events the comparison of the distributions of the jets from top and additional jets provides insights on the top finding process itself.
Again the starting point is the jet multiplicity (Figure 6.23). In the upper row the jet multiplicity of the jet from tops shows anew the influence of the assumption that four jets belong to the top pair. Here the overall ratio of about 0.9 between the jets from top selected by the KinFitter and the Proximity algorithm can clearly be seen. In the lower row the same situation as in (Figure 6.16) is observed and the same explanation theories apply.
For the jets from top 43917 jets are present in the events selected by the KinFitter while the yield for the Proximity algorithm is 52030 resulting in an overall ratio of 1.18. In the case of additional jets the KinFitter found 14502 jets and the Proximity algorithm returned 14808 jets which translates to a ratio of 1.02. Thus the KinFitter selects events yielding more additional jets than the Proximity algorithm in comparison.
The transverse momentum distributions for both algorithms and jet groups can be found in Figure 6.24. The Proximity algorithm exhibits a harder transverse momentum distribution for the jets from top than the KinFitter. The explanation found in the treatment of the transverse momentum distribution for all jets seems to apply here too.

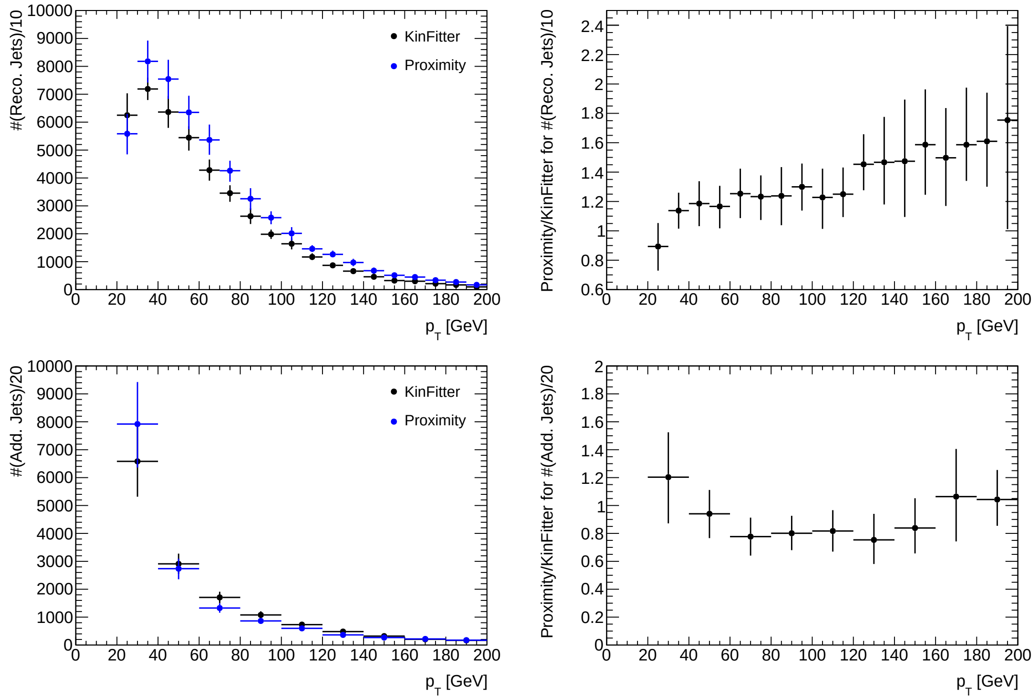

Figure 6.24: Comparison of the transverse momentum distributions of the jets from top (upper row) and additional jets (lower row) in events accepted by the top reconstruction algorithms (left) and the ratio of the outputs (right).

In the additional jet case the KinFitter selects more jets with high transverse momentum. Some of these might be jets from top where the false top pair candidate was found.

Similar effects where combinatorics causes jets from top to be identified as additional jets by the KinFitter seem to apply for the pseudo-rapidity distributions (Figure 6.25). The additional jets determined by the Proximity algorithm exhibit a much more flat spectrum than those found by the KinFitter. Jets that do not stem from the hard scattering are believed to be distributed equally in pseudo-rapidity (at least those produced in QCD interactions) while those from the hard scattering have a clear peak around zero. This is reflected more clearly by the output of the Proximity algorithm than that of the KinFitter.

Such distinctions cannot be made for the azimuthal angle distributions depicted in Figure 6.26. All distributions and ratios are flat.

In Figure 6.27 the b-tagging weights as calculated by JetProb are shown for the two groups of jets. The shapes for the jets from top look quite similar resulting in a relatively flat ratio plot. In contrast the distributions for the additional jets show large deviations between the top finding algorithms. Here the proximity algorithm classifies much more jets with intermediate b-tagging weights as additional jets.

These mismatches are clearly reflected by the multiplicities of b-tagged jets (according to he JetProb algorithm) in Figure 6.28. Overall the KinFitter select 14502 additional

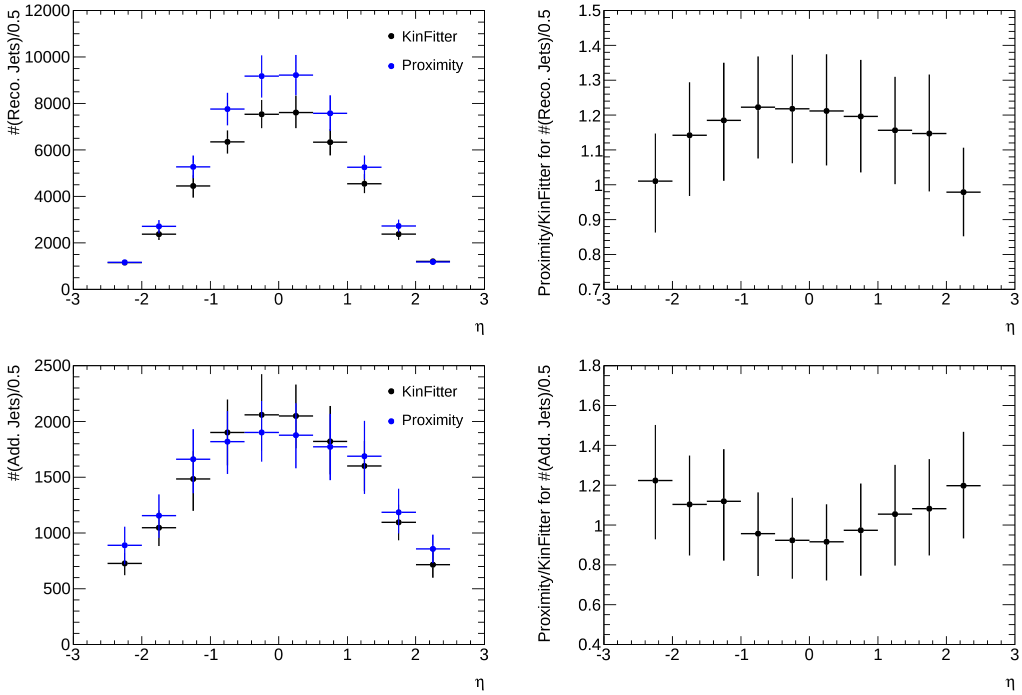

Figure 6.25: Comparison of the pseudo-rapidity distributions of the jets from top (upper row) and additional jets (lower row) in events accepted by the top reconstruction algorithms (left) and the ratio of the outputs (right).

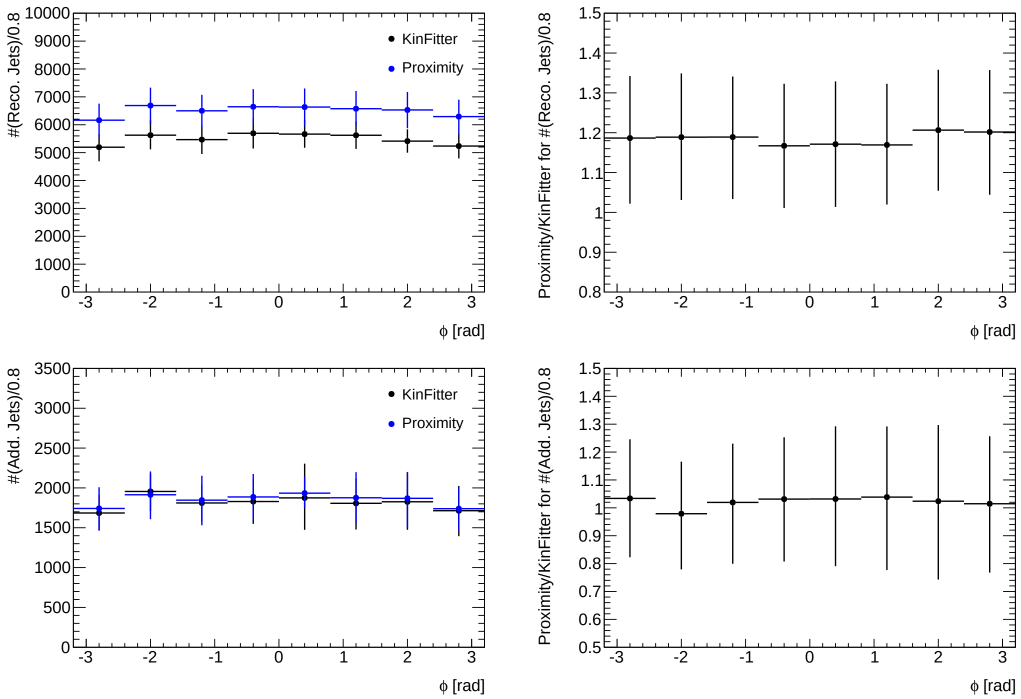

Figure 6.26: Comparison of the azimuthal angle distributions of the jets from top (upper row) and additional jets (lower row) in events accepted by the KinFitter and Proximity algorithm (left) and the ratio of the outputs (right).

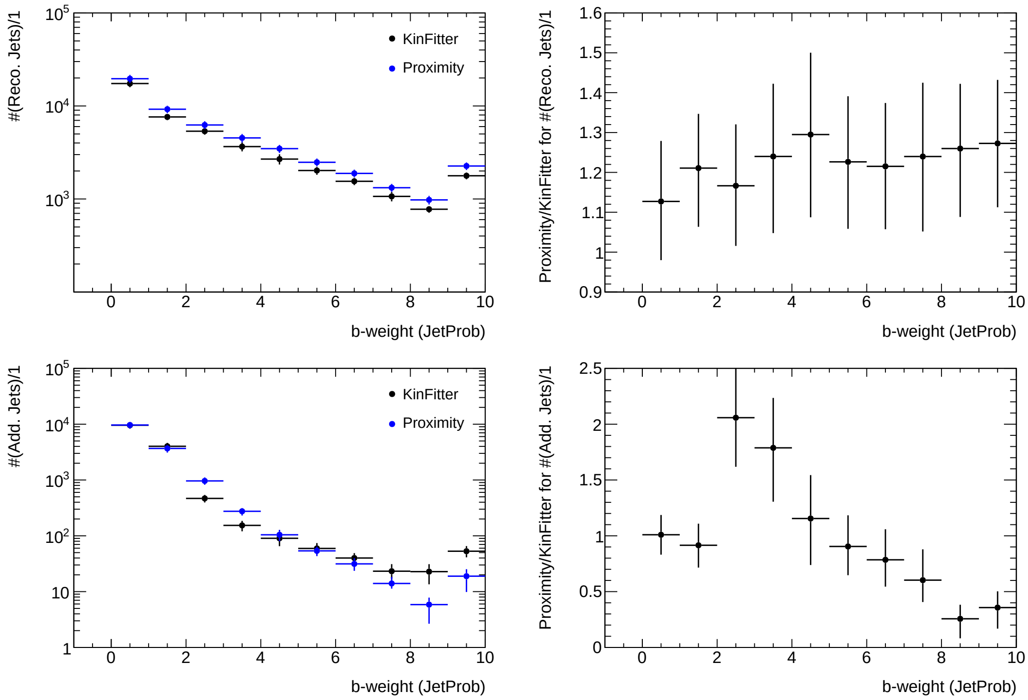

Figure 6.27: Comparison of the b-tagging weight (JetProb) distributions of the jets from top (upper row) and additional jets (lower row) in events accepted by the top reconstruction algorithms (left) and the ratio of the outputs (right). All jets with a b-tagging weight above 2.05 are tagged as jets from b-hadrons.

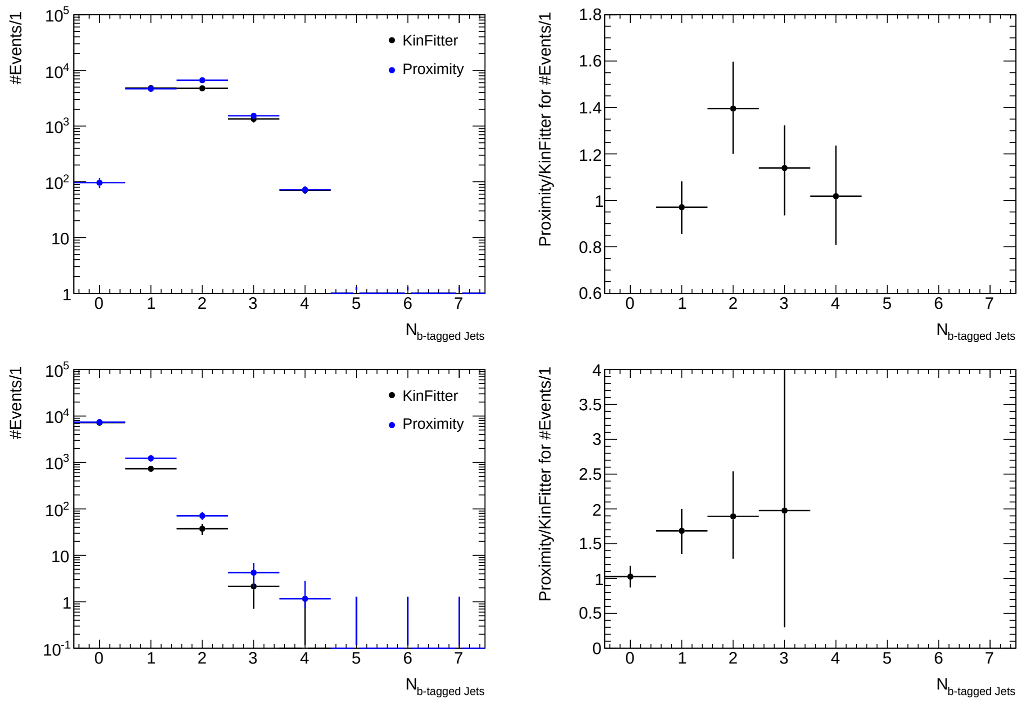

Figure 6.28: Comparison of the multiplicities of jets tagged by the JetProb algorithm of the jets from top (upper row) and additional jets (lower row) in events accepted by the top reconstruction algorithms (left) and the ratio of the outputs (right).

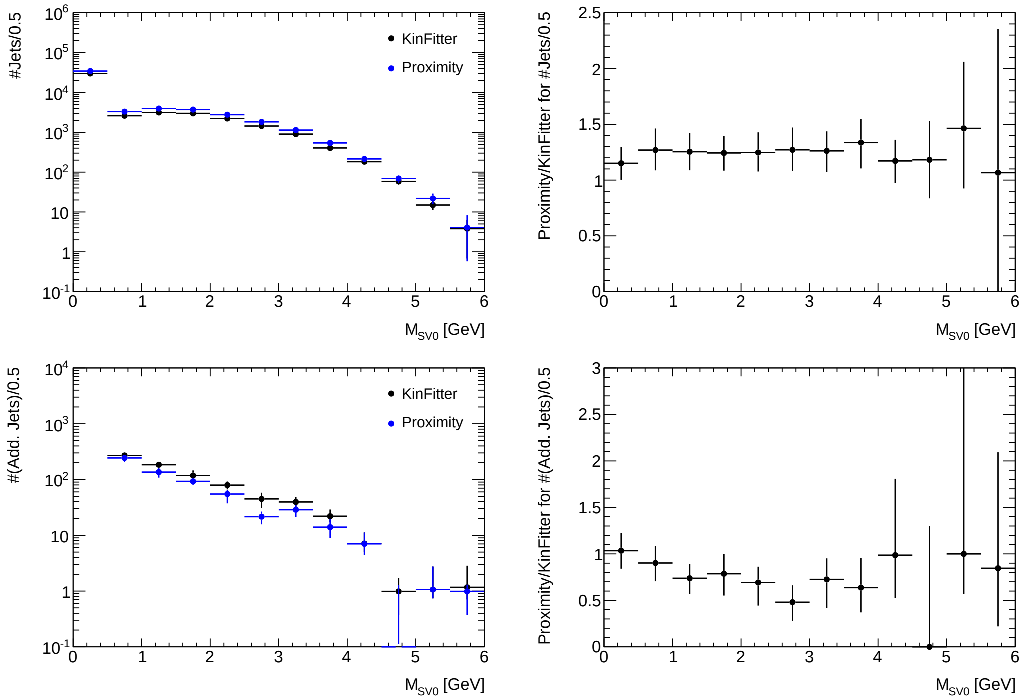

Figure 6.29: Comparison of the multiplicities of jets tagged by the JetProb algorithm of the jets from top (upper row) and additional jets (lower row) in events accepted by the top reconstruction algorithms (left) and the ratio of the outputs (right).

jets of which 814 are tagged by the JetProb algorithm. At the same time the Proximity algorithm yields 14808 additional jets of which 1393 were tagged by the JetProb algorithm. This yields a ratio of 0.056 in the KinFitter case and 0.094 in the Proximity algorithm case.

The last plot in the comparison is the SV0 mass (Figure 6.29). While the distributions of the jets from top case look very similar and the obtained ratio is flat the additional jet case shows a quite large mismatch between the methods. This is insofar troublesome that the heavy flavor fraction obtained via the template fit method might lead to misleading values. Perhaps the determination of the heavy flavor fraction using all jets in the events selected by the methods yields a better and more trustworthy result which then may be projected on the additional jets using assumptions such as that the top pairs always contain two bottom jets and two light jets.

6.2.3 Comparison of the Top Finding Methods for the Leading Additional Jet.

This section summarizes the findings for the additional jet with the highest transverse momentum. In principle all statements that were given in Section 6.2.2 apply here.

Figure 6.30 exhibits the kinematic variables for the leading additional jet. The KinFitter yields 7957 leading additional jets while the Proximity algorithm returns 8699

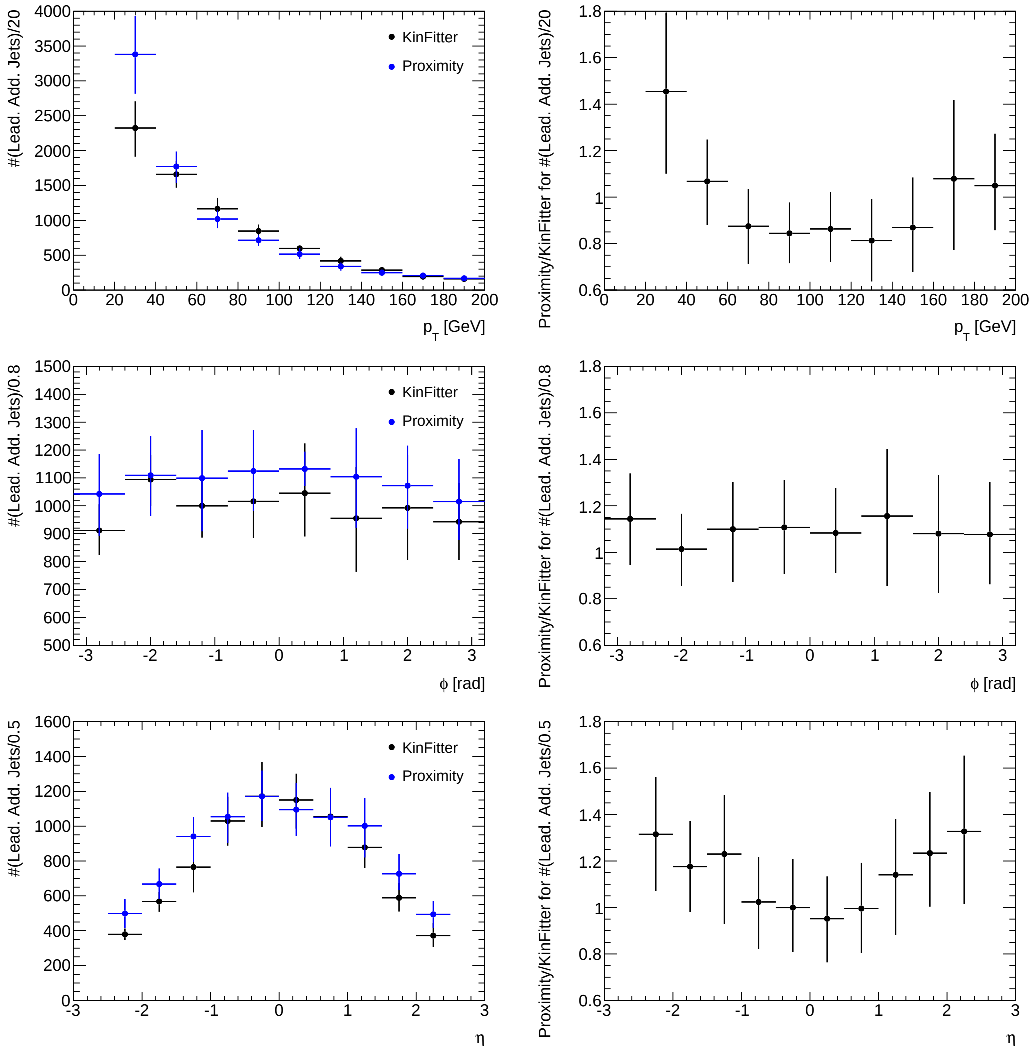

Figure 6.30: The plots depict the transverse momentum (first row), azimuthal angle (second row) and pseudo-rapidity (third row) of the leading additional jet. The distributions for the KinFitter and Proximity algorithm (left) and the ratio between both algorithms (right) are shown.

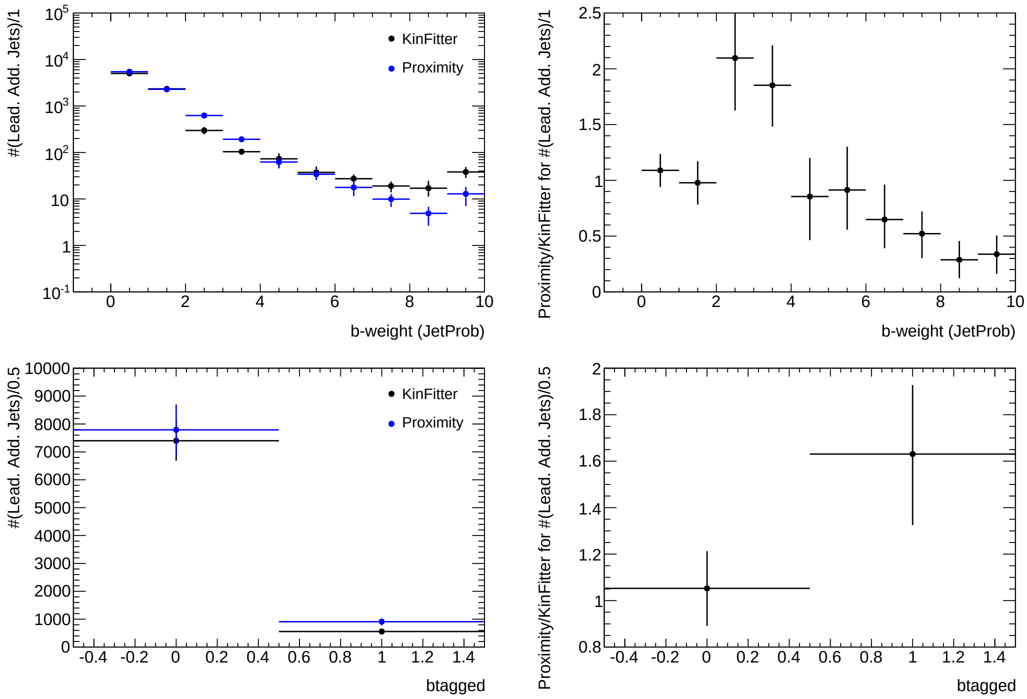

Figure 6.31: The Plots depict the b-tagging weights as calculated by the JetProb algorithm (upper row) and the resulting tagging decision (lower row). The distributions for the KinFitter and Proximity algorithm (left) and the ratio between both algorithms (right) are shown.

leading additional jets corresponding to an overall ratio of 1.09.

All distributions show the same behavior as in the case where all additional jets were examined.

The b-tagging weights and the tagging decision are shown in Figure 6.31 Again large deviations in the b-tagging weights can be observed. Consequently the resulting tagging ratios are quite different for the both algorithms. In the case of the KinFitter 558 out of the 7957 leading additional jet were tagged by the JetProb algorithm resulting in a tagging ratio of 0.07. This is lower than the tagging ratio of 0.105 that was returned by the Proximity algorithm (out of 8699 leading additional jets 910 jets were tagged.)

Figure 6.32 presents the SV0 masses for the leading additional jet with and without a tag of the JetProb algorithm required. Two things are apparent, the deviations between the algorithms are quite high and there is only a small amount of statistics in both cases. This puts limits on the applicability of both template fitting and the correction method.

The last plot in this section depicts the distance of the leading additional jet to the next jet (Figure 6.33). There seems a small amount of jets with a distance lower than 0.4 which might indicate that these are jets that were split by the anti-k_t algorithm. Compared to the entries in the other bins the problematic of split jets might be a negligible concern. Apart from that both exhibit similar shapes.

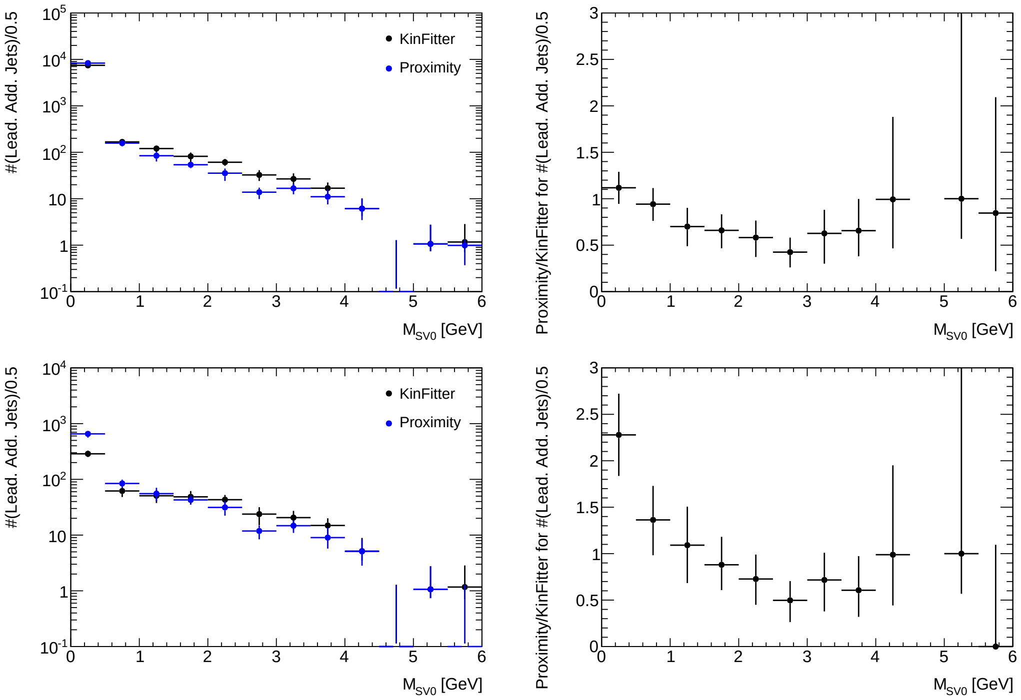

Figure 6.32: The SV0 mass of all leading additional jets (upper row) and for leading additional jets that were tagged by the JetProb algorithm (lower row) are shown. On the left are the distributions for the KinFitter and Proximity algorithm and on the right the ratio between both algorithms.

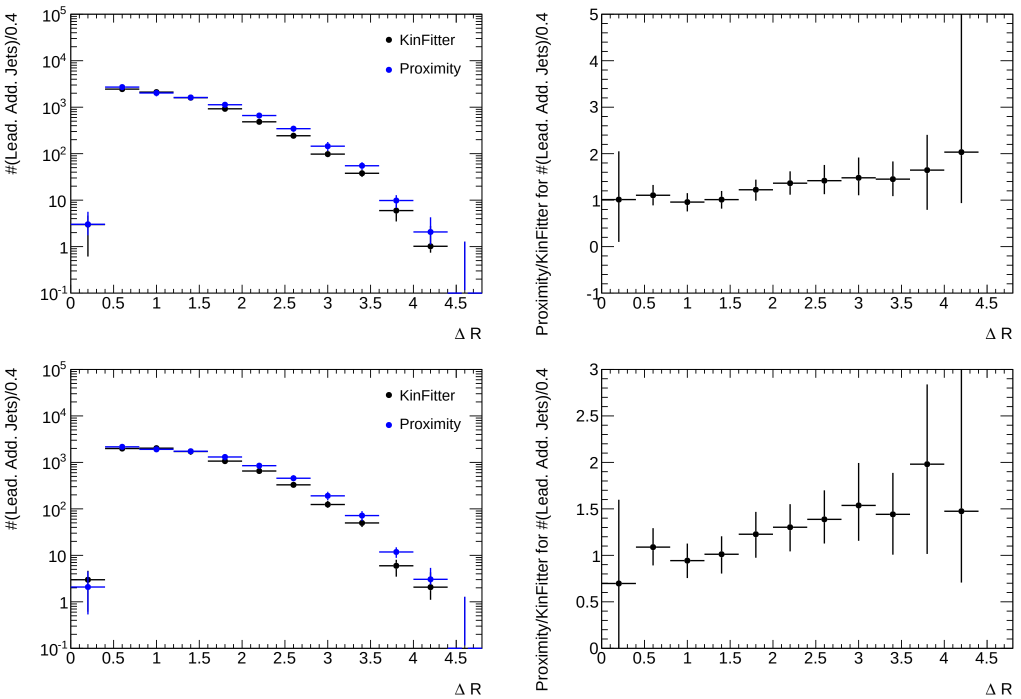

Figure 6.33: The Distance of the leading jet to the next jet (upper row) and the next jet from top (lower row). On the left are the distributions for the KinFitter and Proximity algorithm and on the right the ratio between both algorithms.

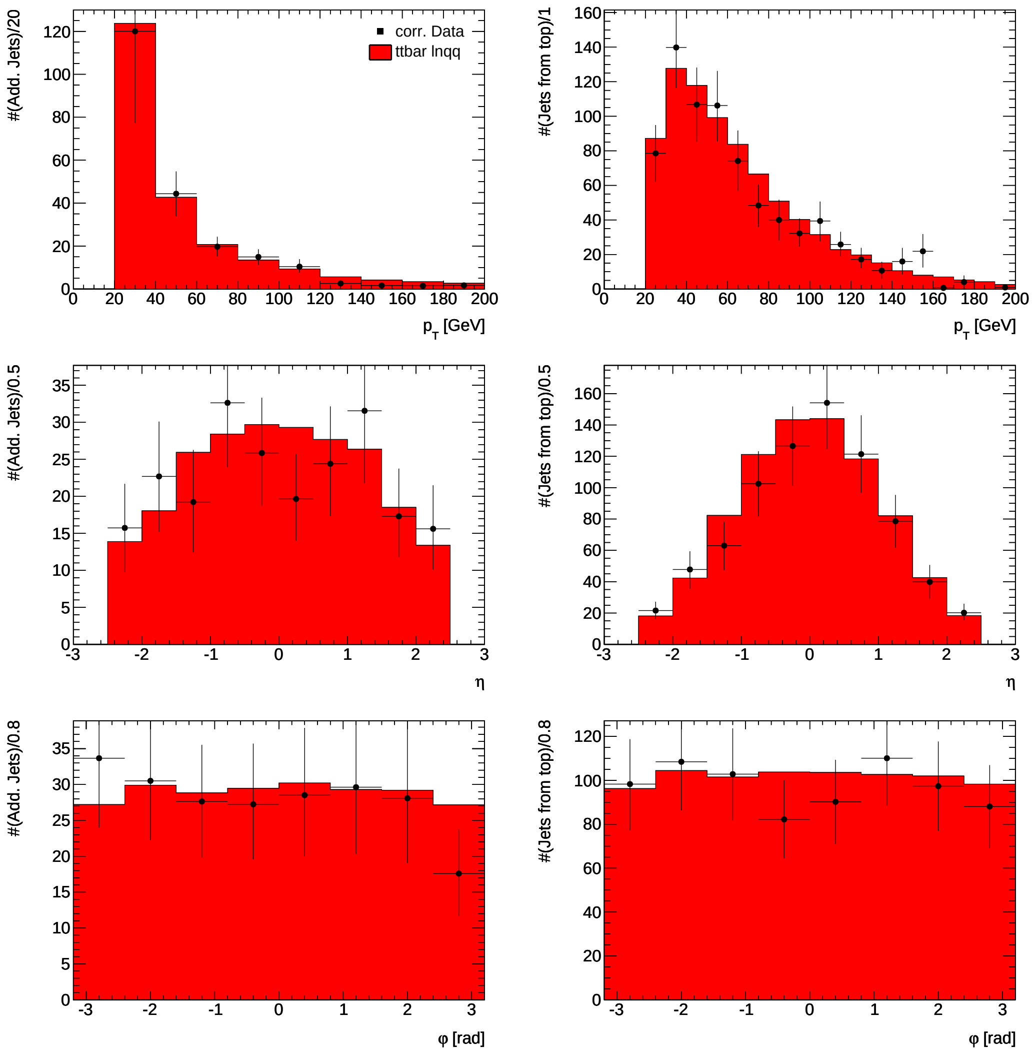

Figure 6.34: The distributions of the transverse momentum (first row) the pseudo-rapidity (second row) and the azimuthal angle (third row) for the additional jets (left) and the jets associated to the top quarks (right). The legend in the upper left plot applies for all plots.

6.3 Comparison of Corrected Data and Proximity Algorithm Output

After comparing the output of the KinFitter in MC simulations and measured data and reviewing the correction factors obtained from the AcerMC signal sample the corrections on data are applied.
First, the physical background as observed in MC simulations is subtracted from the measured data. Following this the correction factors are multiplied to the cleaned data set. The result of this procedure is then compared to the output of the proximity algorithm obtained from the AcerMC signal sample.
In Figure 6.34 the kinematical variables of the additional jets and jets from top found in the corrected data are compared to the output of the Proximity algorithm. The

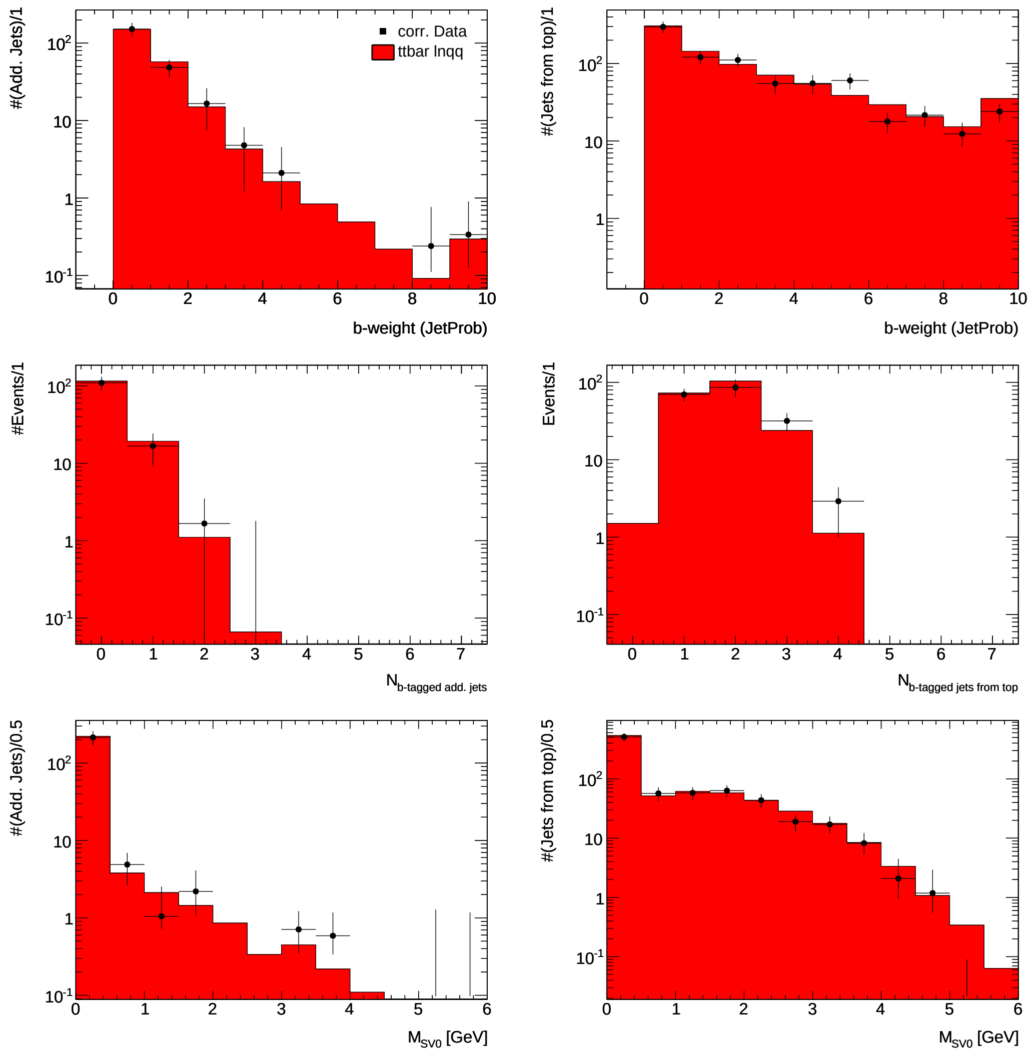

Figure 6.35: The distributions for the b-tagging weights as calculated by JetProb (first row), the multiplicity of tagged jets (second row) and the SV0 mass (third row) for the additional jets (left) and the jets associated to the top quarks (right). The legend in the upper left plot applies for all plots.

overall normalization seems to be correct and the shapes match fairly well.

The same is observed concerning the b-tagging variables shown in Figure 6.35.

The reason for the good agreement of the corrected data and the Proximity algorithm lie in the very good agreement of the outputs of the KinFitter in MC simulations and measured data. This is reflected by the Equation 5.20 which describes the correction method.

In principle the applied corrections should depend on a certain parameterization e. g. the transverse momentum of the leading additional jet. Unfortunately the statistics in 2010 data as well as in the AcerMC signal sample do not permit this. Therefore it is highly desirable to obtain larger samples both in MC simulations as well as in measured data to bring the correction method to full fruition.

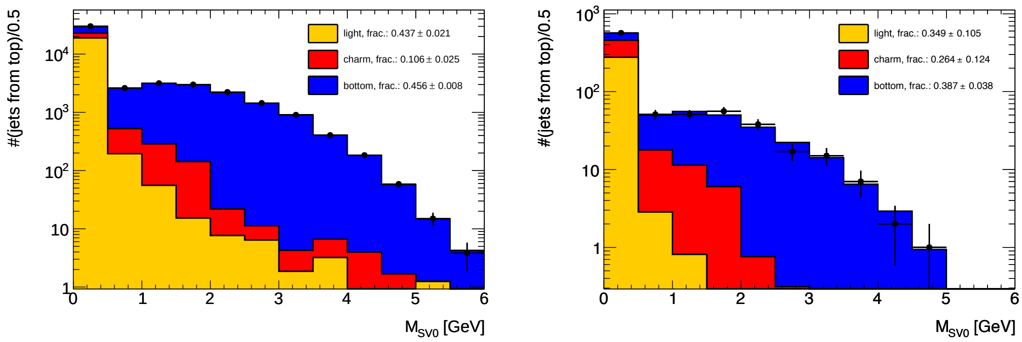

Figure 6.36: The results of the template fit in signal MC (left) and data (right) for jets from top obtained by the KinFitter. The templates used in the fit were taken from signal MC using all jets that passed cut flow and KinFitter.

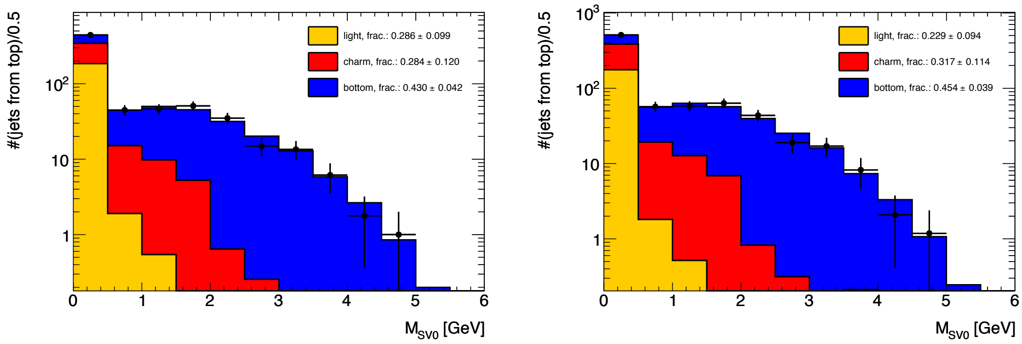

Figure 6.37: The results of the template fit in data with background subtraction applied (left) and fully corrected data (right) for jets from top obtained by the KinFitter. The templates used in the fit were taken from signal MC using all jets that passed cut flow and KinFitter.

6.4 Template Fit Results in Data

This section is designed as a proof of concept that the template fit method may return viable results given a reasonable amount of statistics.

In Figure 6.36 the outcome of the fitting procedure can be seen. Here the template fit in the AcerMC signal sample was plotted again for comparison. The result from data exhibits larger uncertainties than that from the AcerMC signal sample. In contrast to the 804 jets from top found in 2010 data there are 43917 jets form top in the AcerMC signal sample. Therefore the available statistics in MC simulation is 55 times larger than in data.

Nonetheless the template fit as executed in data looks promising keeping in mind that the data are neither corrected with regards to the physical and combinatorial background.

These corrections were applied one by one with the outcome plotted in Figure 6.37. On the left-hand side only the physical background was subtracted. While the bottom fraction rises the light fraction is diminished indicating that the subtracted back-

ground is dominated by light jets.
The subtraction of the physical background decreases the number of jets from top to 656. In principle the resulting uncertainties should be higher because of possible statistical limitations in MC simulations. These effects were not propagated to the inclusive distribution therefore the fit uncertainties might be about the same as in the fit of the uncorrected data.
The result of the last correction step, the combinatorial background correction, can be seen on the right-hand side in Figure 6.37.
The charm fraction is still very high in comparison to that in simulated data but the sum of light and charm fraction is about the same in both cases. It will be interesting to see if this effect is due to limited statistics (here the sum of entries was about 780) or if there are other reasons causing the observed mismatch.
In any case the template fit methods seems to be a potent and feasible method for the determination of flavor fractions in top pair events and its application in 2011 data will return much more reliable results.

7 Summary and Outlook

In this thesis the properties of additional jets in semileptonic top pair events via a kinematic fit were examined. The obtained results were compared in data and Monte Carlo (MC) simulation. Building on the examination of the kinematic fit algorithm strategies for the correction of combinatorial background were developed using the Proximity algorithm. After the subtraction of the physical background derived from MC simulations the combinatorial background was removed using the correction factors from the comparison of the KinFitter to the Proximity algorithm. The resulting data was compared to the distributions that were extracted by the Proximity algorithm from a MC simulation of the signal process.
The application of the KinFitter to measured data shows very promising results. Especially the very good agreement between data and MC simulations encourages the further use of this tool as it proves to be both useful for the selection of top events and the reconstruction of top pairs. With improving understanding of the detector the KinFitter will even increase its performance. Possible new applications could lie in the study of the fully hadronic top pair events where the reconstruction of the top pair may provide ways to discriminate between signal and background a task still very challenging for this process.
Although the KinFitter is a sophisticated tool it may still select the wrong top pair candidate in a given event. Therefore the Proximity algorithm was used to correct for the combinatorial background of the KinFitter analysis. The first tests comparing 2010 data and the MC simulation of the top pair signal as extracted by the Proximity algorithm show promising results. At the moment the correction procedure is limited by statistics in both data and MC simulation. It is desirable to be able to parameterize the correction factors in order to account for possible effects that could not be incorporated due to the small size of the used MC simulation sample (AcerMC). Higher statistics in 2011 data and larger MC simulation samples may improve the reliability of the correction method significantly. Furthermore an examination of other techniques for the matching between particles on generator level and the measured jets would be useful for quantifying the effect of different matching strategies on the obtained correction factors.
A precise measurement of the heavy flavor fraction of additional jets has to be postponed due to the lack of statistics in 2010 data. Nonetheless the implementation and first tests of the template fit technique have been achieved with promising first results. The next logical step will be the application of the derived algorithms on 2011 data.

8 Appendix

8.1 Data Samples

The data used in this study were accumulated during the data taking period of 2010 at the LHC. The data runs went from late March until late November 2010 and were divided into periods. These periods are listed in Table 8.1.

As can be seen in Figure 8.1 the integrated luminosity has risen with increasing pace over the course of the year with data periods G to I making up the majority of the acquired data volume. Between the data periods new setups resulting in higher instantaneous luminosities had been successfully tested. This caused changing prescriptions for the different periods with respect to certain analysis elements e. g. the set of used triggers. At the same time calibration efforts and detector studies took place resulting in correction tools and an improved detector understanding.

In late 2010 the top group decided that it is recommended to only use periods E4 to I for analyses because the preceding data periods are too small in terms of luminosity to justify the effort it takes to keep all prescriptions consistent. For practical reasons this analysis uses only periods F to I which sport a combined luminosity of $43.35\,\mathrm{pb}^{-1}$ before applying the good run list of the top group. With the good run list applied this value decreases to $35\,\mathrm{pb}^{-1}$.

period	L in pb^{-1}
A	0.0004
B	0.009
C	0.0095
D	0.320
E	1.118
F	1.98
G	9.07
H	9.3
I	23

Table 8.1: Data taking periods of the year 2010. The periods marked in green were used in this analysis. The remaining periods were discarded being small in luminosity.

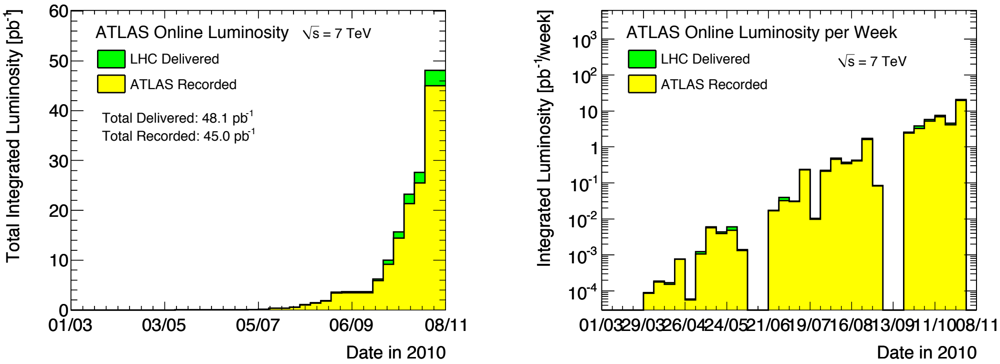

Figure 8.1: The integrated luminosity for the year 2010 (left) and per week (right). During the year 2010 the pace at which data is accumulated has risen steadily. [3]

top pair semileptonic decays (signal) + dileptonic decays (background)
mc10_7TeV.105205.AcerMCttbar.merge.AOD.e574_s933_s946_r1831_r2040/
W + jets (background)
mc10_7TeV.104990.Sherpa010103W4jetstoenu.merge.AOD.e550_s933_s946_r1831_r2040/
mc10_7TeV.104991.Sherpa010103W4jetstomunu.merge.AOD.e550_s933_s946_r1831_r2040/
mc10_7TeV.104992.Sherpa010103W4jetstotaunu.merge.AOD.e550_s933_s946_r1831_r2040/
Z + jets (background)
mc10_7TeV.104993.Sherpa010103Z4jetstoee.merge.AOD.e550_s933_s946_r1831_r2040/
mc10_7TeV.104994.Sherpa010103Z4jetstomumu.merge.AOD.e550_s933_s946_r1831_r2040/
mc10_7TeV.104995.Sherpa010103Z4jetstotautau.merge.AOD.e550_s933_s946_r1831_r2040/
top pair fully hadronic (background)
mc10_7TeV.105204.TTbar_FullHad_McAtNlo_Jimmy.merge.AOD.e598_s933_s946_r1831_r2040/
single top (background)
mc10_7TeV.108340.st_tchan_enu_McAtNlo_Jimmy.merge.AOD.e598_s933_s946_r1831_r2040/
mc10_7TeV.108341.st_tchan_munu_McAtNlo_Jimmy.merge.AOD.e598_s933_s946_r1831_r2040/
mc10_7TeV.108342.st_tchan_taunu_McAtNlo_Jimmy.merge.AOD.e598_s933_s946_r1831_r2040/
mc10_7TeV.108343.st_schan_enu_McAtNlo_Jimmy.merge.AOD.e598_s933_s946_r1831_r2040/
mc10_7TeV.108344.st_schan_munu_McAtNlo_Jimmy.merge.AOD.e598_s933_s946_r1831_r2040/
mc10_7TeV.108345.st_schan_taunu_McAtNlo_Jimmy.merge.AOD.e598_s933_s946_r1831_r2040/
Diboson (background)
mc10_7TeV.105985.WW_Herwig.merge.AOD.e598_s933_s946_r1831_r2040/
mc10_7TeV.105986.ZZ_Herwig.merge.AOD.e598_s933_s946_r1831_r2040/
mc10_7TeV.105987.WZ_Herwig.merge.AOD.e598_s933_s946_r1831_r2040/

Table 8.2: The MC samples used in the analysis

8.2 Monte Carlo Samples

The MC samples that were used in this study are given in Table 8.2.
The selection of MC datasets was primarily determined by practical reasons. The entire analysis was done using the WLCG (Worldwide LHC Computing Grid, in the following called grid). As the privately produced data sets (see Section 8.3) were too big to be stored and analyzed locally the use of the grid became mandatory. One drawback of the private production of data sets is that once a grid site that contains a needed sample is offline it is simply not available. Therefore having as few MC samples to manage was desirable. This enforced the use of the SHERPA [65] samples in the case of W+jets and Z+jets as the MC samples representing the processes are only six compared to the Alpgen [66] samples are about 60 in number. The same reasons motivated the use of HERWIG [67] in the diboson case.
The choice for the AcerMC sample resulted from problems when reading the information of the MC particles on generator-level. For applying the correction algorithm it was needed to follow the decay chain of the particles on generator-level from the top quark to the W boson and its decay products. Unfortunately for all generators except AcerMC there were many events (in general about 40 %) where this information was not readable in a automated way.
Reasons were that some decay chains simply ended prematurely, that the W boson only decayed to one quark or that decay chains exhibited odd intermediate steps (e. g. particles which were labelled as clusters) which confused the reading algorithm. On which level these problems were introduced (e. g. in the generations of the AOD or in the translation of AODs to A++ files) is under investigation.
While AcerMC exhibits very well readable information regarding the particles at generator-level it suffers from the fact that it does have about one tenth of the statistics compared to the MC@NLO signal samples which are mainly used in top pair analyses.

8.3 The A++ Framework

This study was carried out with the help of the A++ framework [68]. This is an analysis framework mainly developed by the Experimental Particle Physics group of the physics faculty of the Humboldt university, Berlin. Around ten people are actively contributing to the framework with Oliver Maria Kind being the lead author.
A++ was designed to be independent of Athena. The only interface to Athena is a converter which converts AODs to A++ unique, object-oriented file format, the A++ file. By introducing this conversion process disk space requirements for the data files can be lowered as the A++ file does not contain all of the information that AODs do (similar to what is done with D3PDs). In addition, information that is almost always calculated for events such as information on object quality maps may be written to the file facilitating analyses. Another advantage of the A++ file is its object-oriented layout. Here the information concerning a certain physical object is not scattered on many arrays and tables but is modeled as an object itself. This object-oriented approach enables the user to write elegant and readable code thus minimizing bugs and facilitates debugging.

List of Figures

List of Tables

Bibliography

[1] L. Evans and P. Bryant, *LHC Machine*, Journal of Instrumentation **3** (2008) no. 08, S08001.

[2] http://lhc.web.cern.ch/lhc/News.htm. visited on July 26th, 2011.

[3] https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults. visited on July 23rd, 2011.

[4] CDF Collaboration, F. Abe et al., *Observation of top quark production in anti-pp collisions*, Phys. Rev. Lett. **74** (1995) 2626–2631, `arXiv:hep-ex/9503002`.

[5] D0 Collaboration, S. Abachi et al., *Observation of the top quark*, Phys. Rev. Lett. **74** (1995) 2632–2637, `arXiv:hep-ex/9503003`.

[6] M. Flechl, *ttbar backgrounds in charged Higgs boson searches*, Tech. Rep. ATL-PHYS-PROC-2010-149, CERN, Geneva, Dec, 2010.

[7] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (Nov, 1967) 1264–1266.

[8] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, **2** (Oct., 1970) 1285–1292.

[9] S. L. Glashow, *Partial-Symmetries of Weak Interactions*, Nuclear Physics **22** (1961) no. 4, 579–588.

[10] G. Altarelli and M. W. Grunewald, *Precision electroweak tests of the standard model*, Phys. Rept. **403-404** (2004) 189–201, `arXiv:hep-ph/0404165`.

[11] G. Bertone, D. Hooper, and J. Silk, *Particle dark matter: Evidence, candidates and constraints*, Phys. Rept. **405** (2005) 279–390, `arXiv:hep-ph/0404175`.

[12] K. Nakamura et al., *The Review of Particle Physics*, Journal of Physics G **37** (2010) 075021.

[13] S. Tomonaga, *On a Relativistically Invariant Formulation of the Quantum Theory of Wave Fields*, Progress of Theoretical Physics **1** (1946) no. 2, 27–42.

[14] J. Schwinger, *On Quantum-Electrodynamics and the Magnetic Moment of the Electron*, Phys. Rev. **73** (1948) no. 4, 416–417.

[15] R. P. Feynman, *Mathematical Formulation of the Quantum Theory of Electromagnetic Interaction*, Phys. Rev. **80** (1950) no. 3, 440–457.

[16] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, Phys. Rev. Lett. **30** (1973) no. 26, 1346–1349.

[17] H. D. Politzer, *Asymptotic Freedom: An Approach to Strong Interactions*, Physics Reports **14** (1974) no. 4, 129 – 180.

[18] D. J. Gross and F. Wilczek, *Asymptotically Free Gauge Theories. I*, Phys. Rev. D **8** (1973) no. 10, 3633–3652.

[19] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (Jun, 1963) 531–533.

[20] M. Kobayashi and T. Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, Progress of Theoretical Physics **49** (1973) no. 2, 652–657. http://ptp.ipap.jp/link?PTP/49/652/.

[21] S. W. Herb, D. C. Hom, L. M. Lederman, J. C. Sens, H. D. Snyder, J. K. Yoh, J. A. Appel, B. C. Brown, C. N. Brown, W. R. Innes, K. Ueno, T. Yamanouchi, A. S. Ito, H. Jöstlein, D. M. Kaplan, and R. D. Kephart, *Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions*, Phys. Rev. Lett. **39** (Aug, 1977) 252–255.

[22] M. Jezabek and J. H. Kühn, *QCD corrections to semileptonic decays of heavy quarks*, Nuclear Physics B **314** (1989) no. 1, 1 – 6. http://www.sciencedirect.com/science/article/pii/0550321389901089.

[23] S. Catani, *Aspects of QCD, from the Tevatron to the LHC*, `arXiv:hep-ph/0005233`.

[24] *Searches for Single Top-Quark Production with the ATLAS Detector in pp Collisions at* $\sqrt(s) = 7$ TeV, Tech. Rep. ATLAS-CONF-2011-027, CERN, Geneva, Mar, 2011.

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

[26] *Measurement of the top quark-pair cross-section with ATLAS in pp collisions at* $\sqrt(s) = 7$ TeV *in the single-lepton channel using b-tagging*, Tech. Rep. ATLAS-CONF-2011-035, CERN, Geneva, Mar, 2011.

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

[28] T. D. H. Project.

[29] A. Sherstnev and R. S. Thorne, *Parton Distributions for LO Generators*, Eur. Phys. J. **C55** (2008) 553–575, `arXiv:0711.2473 [hep-ph]`.

[30] A. Quadt, *Top quark physics at hadron colliders*, The European Physical Journal C - Particles and Fields **48** (jan, 2006) 835–1000.

[31] The ATLAS Collaboration, G. Aad et al., *The ATLAS Experiment at the CERN Large Hadron Collider*, Journal of Instrumentation **3** (2008) no. 08, S08003.

[32] The CMS Collaboration, S. Chatrchyan et al., *The CMS Experiment at the CERN LHC*, Journal of Instrumentation **3** (2008) no. 08, S08004.

[33] A. Augusto et al., *The LHCb Detector at the LHC*, Journal of Instrumentation **3** (2008) no. 08, S08005.

[34] The ALICE Collaboration, K. Aamodt et al., *The ALICE Experiment at the CERN LHC*, Journal of Instrumentation **3** (2008) no. 08, S08002.

[35] The TOTEM Collaboration, G. Anelli et al., *The TOTEM Experiment at the CERN Large Hadron Collider*, Journal of Instrumentation **3** (2008) no. 08, S08007.

[36] The LHCf Collaboration, O. Adriani et al., *The LHCf Detector at the CERN Large Hadron Collider*, Journal of Instrumentation **3** (2008) no. 08, S08006.

[37] W. Lampl, S. Laplace, D. Lelas, P. Loch, H. Ma, S. Menke, S. Rajagopalan, D. Rousseau, S. Snyder, and G. Unal, *Calorimeter Clustering Algorithms: Description and Performance*, Tech. Rep. ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, Geneva, Apr, 2008.

[38] G. P. Salam and G. Soyez, *A practical Seedless Infrared-Safe Cone jet algorithm*, JHEP **05** (2007) 086, `arXiv:0704.0292 [hep-ph]`.

[39] G. P. Salam and M. Cacciari, *The anti-kt jet clustering algorithm*, JHEP 0804:063 (2008) , `arXiv:0802.1189v2`.

[40] The ATLAS Collaboration, G. Aad et al., *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics*, `arXiv:hep-ex/0901.0512`.

[41] D. Lopez Mateos, E. W. Hughes, and A. Schwartzman, *A Simple p_T- and η-Dependent Monte Carlo-Based Jet Calibration*, Tech. Rep. ATL-PHYS-INT-2009-077, CERN, Geneva, Aug, 2009.

[42] *Properties of Jets and Inputs to Jet Reconstruction and Calibration with the ATLAS Detector Using Proton-Proton Collisions at $\sqrt{s} = 7$ TeV*, Tech. Rep. ATLAS-CONF-2010-053, CERN, Geneva, Jul, 2010.

[43] V. Kostioukhine, *Secondary vertex based b-tagging*, tech. rep., 2003. ATL-PHYS-2003-033.

[44] K. Zenker, M. Kobel, and M. zur Nedden, *Development of a data based algorithm for measuring the mistag rate of b-quarks at the ATLAS experiment. oai:cds.cern.ch:1311861.* PhD thesis, Dresden, TU Dresden, Dresden, 2010. Presented on 01 Dec 2010.

[45] G. Aad, S. Aoun, J. Boudreau, N. Bousson, B. Clément, T. Delemontex, L. Feligioni, A. Ferretto Parodi, F. Filthaut, J. Fleckner, F. Garberson, B. Guo, T. Golling, L. Heinrich, F. Hirsch, C. Jung, A. Khanov, C. Liu, A. Lucotte, T. Merugo, M. Muller, J. Mueller, M. zur Nedden, F. Parodi, F. Rizatdinova, L. Rossi, K. Sapp, M. Saleem, C. Schiavi, J. Schwindling, P. Sinervo, P. Skubic, S. Strandberg, M. Talby, N. Tannoury, K. Uchida, L. Vacavant, M. Verducci, G. Watts, C. Weiser, and J. Yu, *Calibrating the b-Tag Efficiency and Mistag Rate in 35 pb*$^{-1}$ *of Data with the ATLAS Detector*, Tech. Rep. ATLAS-COM-CONF-2011-082, CERN, Geneva, Apr, 2011.

[46] D. Buskulic et al., *A precise measurement of* $[\gamma]Z \to b/[\gamma]Z \to hadrons$, Physics Letters B **313** (1993) no. 3-4, 535 – 548. http://www.sciencedirect.com/science/article/pii/037026939390028G.

[47] *Performance of the ATLAS Secondary Vertex b-tagging Algorithm in 7 TeV Collision Data*, Tech. Rep. ATLAS-CONF-2010-042, CERN, Geneva, Jul, 2010.

[48] S. Hassani, L. Chevalier, E. Lançon, J.-F. Laporte, R. Nicolaidou, and A. Ouraou, *A muon identification and combined reconstruction procedure for the ATLAS detector at the LHC using the (MUONBOY, STACO, MuTag) reconstruction packages*, NIM A **572** (2007) no. 1, 77 – 79. http://www.sciencedirect.com/science/article/pii/S0168900206019863. Frontier Detectors for Frontier Physics - Proceedings of the 10th Pisa Meeting on Advanced Detectors.

[49] T. Lagouri, D. Adams, K. Assamagan, M. Biglietti, G. Carlino, G. Cataldi, F. Conventi, A. Farilla, Y. Fisyak, S. Goldfarb, E. Gorini, K. Mair, L. Merola, A. Nairz, A. Poppleton, M. Primavera, S. Rosati, J. Shank, S. Spagnolo, L. Spogli, G. Stavropoulos, M. Verducci, and T. Wenaus, *A muon identification and combined reconstruction procedure for the ATLAS detector at the LHC at CERN*, Nuclear Science, IEEE Transactions on **51** (dec., 2004) 3030 – 3033.

[50] *Muon reconstruction efficiency in reprocessed 2010 LHC proton-proton collision data recorded with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2011-063, CERN, Geneva, Apr, 2011.

[51] J. Wess and B. Zumino, *Supergauge transformations in four dimensions*, Nuclear Physics B **70** (1974) no. 1, 39 – 50. http://www.sciencedirect.com/science/article/pii/0550321374903551.

[52] J. Erdmann, K. Kroeninger, O. Nackenhorst, and A. Quadt, *Kinematic fitting of ttbar-events using a likelihood approach: The KLFitter package*, Tech. Rep. ATL-COM-PHYS-2009-551, CERN, Geneva, Sep, 2009.

[53] E. Cogneras and D. Pallin, *Reconstruction of t events without b-tagging within ATLAS*, Tech. Rep. ATL-COM-PHYS-2009-083, CERN, Geneva, Mar, 2009.

[54] T. Goepfert and J. E. Sundermann, *KinFitter - A Kinematic Fit with Constraints*, Tech. Rep. ATL-COM-SOFT-2009-014, CERN, Geneva, Sep, 2009.

[55] *Determination of the muon reconstruction efficiency in ATLAS at the Z resonance in proton-proton collisons at sqrt(s)=7 TeV*, Tech. Rep. ATLAS-CONF-2011-008, CERN, Geneva, Feb, 2011.

[56] *Measurement of the Top-Quark Mass using the Template Method in pp Collisions at root(s)=7 TeV with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2011-033, CERN, Geneva, Mar, 2011.

[57] V. Blobel and E. Lohrmann, *Statistische und numerische Methoden der Datenanalyse.* Teubner Verlag, 1 ed., 1998.

[58] T. Göpfert, M. Kobel, and H. Lacker, *Studies of Charged Higgs Boson Observability in the $H \to tb$ Decay at the ATLAS Experiment within the Minimal Supersymmetric Standard Model. oai:cds.cern.ch:1316223.* PhD thesis, Dresden, Technische Universität Dresden, Dresden, 2010. Presented on 15 Dec 2010.

[59] D0 Collaboration, V. M. Abazov et al., *Determination of the width of the top quark*, Phys. Rev. Lett. **106** (2011) 022001, `arXiv:1009.5686 [hep-ex]`.

[60] O. M. Kind et al. http://isscvs.cern.ch/cgi-bin/cvsweb.cgi/APlusPlus/?cvsroot=huberlin. visited on July 28th, 2011.

[61] P. Rieck, *Entwicklung eines kinematischen Fits zur Untersuchung elektroschwacher Top-Quark-Produktion bei ATLAS.* Master thesis, Humboldt Universität zu Berlin, 2010.

[62] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, JHEP **06** (2002) 029, `arXiv:hep-ph/0204244`.

[63] R. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, Computer Physics Communications **77** (1993) no. 2, 219 – 228. http://www.sciencedirect.com/science/article/pii/001046559390005W.

[64] F. Rademakers and R. Brun, *ROOT - An Object Oriented Data Analysis Framework*, Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. Meth. in Phys. Res. A **389** (1997) 81 – 86. http://root.cern.ch/.

[65] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007, `arXiv:0811.4622 [hep-ph]`.

[66] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, *ALPGEN, a Generator for Hard Multiparton Processes in Hadronic Collisions*, `arXiv:hep-ph/0206293v2`.

[67] G. Corcella et al., *HERWIG 6.5 release note*, `arXiv:hep-ph/0210213`.

[68] O. M. Kind et al. http://www-eep.physik.hu-berlin.de/ATLAS/a_plusplus. visited on July 23rd, 2011.

Danksagung

So wie es zum Großziehen eines Kindes ein ganzes Dorf benötigt, so ist auch eine Diplomarbeit nie das Ergebnis der Arbeit eines Einzelnen.
Zuvorderst möchte ich Herrn Prof. Michael Kobel danken, der durch seine spannende und motivierende Art, Wissen zu vermitteln, mich zur Teilchenphysik gebracht hat und es mir ermöglichte, diese Diplomarbeit im Bereich der Experimentalteilchenphysik anzufertigen.
Dafür, dass er allzeit ein offenes Ohr hatte und diese Arbeit mit vielen wertvollen Hinweisen und Anregungen erst zu dem gemacht hat, was sie ist, möchte ich meinem Betreuer Herrn Dr. Martin zur Nedden sehr herzlich danken.
Herrn Dr. Ulrich Husemann möchte ich großen Dank dafür aussprechen, dass er sich zur Begutachtung dieser Arbeit bereit erklärt hat.
Allen, die am Institut für Kern- und Teilchenphysik arbeiten und forschen möchte ich für die vielen Gespräche, anregenden Diskussionen und einfach die gute Zeit danken.
Gemäß der Weisheit "Ohne Admin - keine Competition" möchte ich mich bei allen Administratoren (Rainer Schwierz, Andreas Petzold, Thomas Göpfert, Wolfgang Mader und Anja Vest) dafür bedanken, dass sie unermüdlich die Rechnersysteme überwachen und am Leben halten.
Für die Möglichkeit das A++ Framework mitgestalten zu können und für die fruchtbare Zusammenarbeit möchte ich mich bei den Physikern der Humboldt Universität zu Berlin, insbesondere Oliver Maria Kind, Ruth Herrberg, Michelangelo Giorgi, Lukas Heinrich und Rocco Mandrysch bedanken.
Ganz speziell sei allen gedankt, die diese Arbeit vorab gegengelesen haben.
Meiner Familie danke ich für die bedingungslose Unterstützung und die schöne Zeit, die war, ist und sein wird.
Ganz besonders danken möchte ich zum Schluss Juliana, die eine unerschöpfliche Quelle für Alles ist, was das Leben schön macht.

Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Dresden, den 30.06.2011

Felix Thomas