GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Fakultät für
Physik

Master's Thesis

Reconstruction studies for $t\bar{t}H(H \to b\bar{b})$ events with the ATLAS experiment at the LHC

Rekonstruktionsstudien von $t\bar{t}H(H \to b\bar{b})$ Ereignissen mit dem ATLAS Experiment am LHC

prepared by

Veit Peter Dahlke

from Magdeburg

at the II. Institute of Physics

Inventory Nr.: II.Physik-UniGö-MSc-2013/06

Thesis period: 30th May 2013 until 29th November 2013

First Referee: Prof. Dr. Arnulf Quadt

Second Referee: Prof. Dr. Ariane Frey

Abstract

In this thesis, a kinematic likelihood fitter is used to reconstruct the decay of top-quark pairs produced in association with a Higgs boson that decays into two b-quarks. A simulated data sample produced in the environment of the ATLAS experiment is used to evaluate the reconstruction efficiencies for different modified versions of the likelihood fitter compared to a basic approach used in a 2012 ATLAS analysis.
It was found that more sophisticated model assumptions can positively improve the reconstruction of the $t\bar{t}H(H \to b\bar{b})$ decay signature. Furthermore, the dependence of the reconstruction efficiencies on number of b-tagged jets used during the kinematic fit is studied.

Zusammenfassung

In dieser Arbeit wird ein kinematisches fitting tool verwendet um den Zerfall von top-quark Paaren, welche in Kombination mit einem Higgs-Boson erzeugt wurden. Hierfür werden simulierte Daten verwendet die im Kontext des ATLAS Experiments erzeugt wurden, um die Rekonstruktionseffizienzen für verschiedene modifizierte Versionen des kinematischen Fitters im Vergleich zu einem grundlegendem Model,l welches in 2012 zur Analyse dieses Zerfalls genutzt wurde, zu untersuchen. Es konnte gezeigt werden das ein aufwändigeres Modell die Rekonstruktion von $t\bar{t}H(H \to b\bar{b})$ Zerfällen verbessern kann. Außerdem wurde die Abhängigkeit der Rekonstruktionseffizienz von der Anzahl der Jets, welche vom Zerfall von b-quarks hervorgerufen wurden, untersucht.

Contents

1 Introduction

Since the times of ancient Greece, humankind wondered about the constituents of the matter surrounding them. The Greek philosopher Demokrit was the first to assume that everything should be composed of small particles that can not be divided into even smaller particles. Over 2500 years later we know that he was right in a sense that there are indeed smallest undividable particles, the so called elementary particles. The field of elementary particle physics tries to identify those particles and deduce rules for the interactions between them.
The discovery of the electron in 1897 could be seen as a starting point for modern particles physics. In the years following groundbreaking works like Rutherfords realization that atoms consist of small localized nuclei with electrons surrounding it, lead to a theoretical description, which peaked in the Standard Model of particle physics developed in the 1970s. The Standard Model combined three well-recognized theories describing particle interactions, namely quantum electrodynamics, quantum flavourdynamics and quantum chromodynamics into one unified theory. The Standard Model is in accordance withall known particles, like quarks and leptons, all of which have been found in the past decades strengthening the confidence in the model.
Nowadays elementary particle physics is mainly pursued at particle accelerators that collide highly relativistic particles, in order to provoke different interactions between them and to measure the reaction outcome. The most advanced accelerator at the moment is the Large Hadron Collider (LHC) at CERN located near Geneva. Its main focus is on the detection of the last missing piece of the Standard Model, the Higgs boson, which has been discovered in 2012 by the two main experimental groups, namely ATLAS [1] and CMS [2], individually. But even after this extraordinary discovery, there are still a wide array of interesting physics processes to be studied at the LHC. The new particles traits have to be measured to make sure it is indeed the particle the Standard Model predicts. Furthermore, in collisions with rising center of mass energies at the LHC completely unexpected observations might

be possible, which are usually referred to as physics beyond the Standard Model. Another important field is the measurement of top-quarks in order to confirm its attributes, which is especially difficult due to its very short lifetime. Moreover the interaction between top quark and the Higgs boson is an opportunity to measure the Yukawa coupling. This would allow a calculation of the elementary particle's masses, since the coupling strength is proportional to the particles mass.
This thesis explores a reconstruction algorithm for events including the mentioned particles, the associated production of a pair of top quarks together with a Higgs boson ($t\bar{t}H$), decaying into a special signature. This is a very rare process, which usually is hard to distinguish from various other processes referred to as background, thus requiring a preferably efficient reconstruction method in order to maximize the knowledge gained by the few processes detected.
This thesis is organized as follows. Chapter 2 focuses on a short recapitulation of the underlying physics model, i.e. the Standard Model, followed by a description of the experimental environment in Chapter 3, namely the LHC and the ATLAS detector. In Chapter 4 a kinematic fitting tool, the KLFitter, used for reconstruction of the events is described and the modifications implemented during my studies are discussed. To allow an evaluation of the reconstruction efficiencies, data are used generated by a Monte Carlo simulation. The generation process and the selection of events used for the analysis is described in Chapter 5. Also a short overview over the state of $t\bar{t}H$ measurements by the ATLAS collaboration is given in Chapter 6. The results obtained during the different analysis are presented in Chapter 7. Eventually, in Chapter 8 a conclusion and an outlook to future implementations of the KLFitter in studies of $t\bar{t}H$ is given.

According to custom in elementary particle physics the constants of the speed of light c and the reduced Planck constant $\hbar$ are set to 1 in natural units $c = \hbar = 1$. This implies that the energies, momenta and masses of particles are all given in the same unit, the electronvolt eV.

2 The Standard Model

The Standard Model (SM) of particle physics is a theory that gives a description of all known elementary particles and also the interactions amongst them. The SM model incorporates the different theories describing three of the four elementary forces excluding gravity: the weak interaction, the strong interaction and the electromagnetic interaction. Quantum electrodynamics (QED), quantum flavourdynamics (QFD) and their unification the Glashow-Weinberg-Salam theory of electroweak interaction [3–5] cover, as the name suggests, the electromagnetic and the weak interactions, whereas by quantum chromodynamics (QCD) [6–8] the strong interaction is explained. Since the SM was devised by the scientific community in the 1970's theorists and experimentalist worked hand in hand to verify the predictions made by the theory.
An important feature of the SM is that all the underlying physics concepts require a local gauge invariant theory. This chapter is supposed to give a short overview over the different interactions within the SM, in particular the role of the top quark and the Higgs boson and especially their combined production mechanism. Also the implications of an observation of this combined production are discussed shortly.

2.1 Quarks and Leptons

The constituents (or particles) of the SM are divided into two groups. The fermions, which represent the actual particles which form heavier particles and the matter that surrounds us, and the so called (gauge-)bosons, which describe interactions between the elementary particles and as such even interactions with the bosons themselves. The fermions are once again divided into quarks and leptons, that are again separated into three *generations*. Higher generation particles carry the same charge and spin as the earlier generations, but differ substantially in their rest mass (Tab. 2.1). In total there are six leptons and quarks, respectively (Tab. 2.2). When talking about quarks one usually distinguishes six different *flavours* that represent the six

different quarks. The quarks are classified by their charge (Q) and quantum numbers that resemble the different flavors i.e. upness (U), downness (D), strangeness (S), charm (C), beauty or bottomness (B) and finally truth or topness (T) [9]. On top of this quarks also carry another charge-like property called colour, which can take three different values red (r), green (g) and blue (b).

leptons		Quarks	
	Mass m in $[\frac{\text{MeV}}{\text{c}^2}]$	Flavour	Mass m in $[\frac{\text{MeV}}{\text{c}^2}]$
e	0.511	*up*	2
ν_e	$< 2\text{x}10^{-6}$	*down*	5
μ	106	*strange*	100
ν_μ	< 0.2	*charm*	1200
τ	1777	*bottom*	4200
ν_τ	< 18	*top*	174000

Table 2.1: Masses of the fermions as given in [9].

The leptons can be classified in a similar way again with charge (Q) and also the quantum numbers electron number (L_e), muon number (L_μ) and tau number (L_τ) [9]. All fermions carry a spin of $s = \frac{1}{2}$. Every quark and lepton also has an antiparticle partner with the same mass, but opposite electric charge, with the exception of the neutrinos. This results in a total of 48 possible fermions. Important to note in the context of this thesis is the role of the top quark among those particles due to its very short lifetime it is the only quark that does not pair with other quarks to formbound states the so called *hadrons*. An example for a hadron is the proton. The hadrons are again divided into *baryons* and *mesons* containing three and two quarks, respectively. The top quark and its properties will be discussed further in Chapter 2.3.

Particles				$Q[e]$	C	s
Generation	I	II	III			
Quarks	$\begin{pmatrix} u \\ d \end{pmatrix}$	$\begin{pmatrix} c \\ s \end{pmatrix}$	$\begin{pmatrix} t \\ b \end{pmatrix}$	$\begin{pmatrix} +\frac{2}{3} \\ -\frac{1}{3} \end{pmatrix}$	$\begin{matrix} r,g,b \\ r,g,b \end{matrix}$	$\begin{matrix} \frac{1}{2} \\ \frac{1}{2} \end{matrix}$
Leptons	$\begin{pmatrix} \nu_e \\ e \end{pmatrix}$	$\begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}$	$\begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}$	$\begin{matrix} 0 \\ -1 \end{matrix}$	$\begin{matrix} -- \\ -- \end{matrix}$	$\begin{matrix} \frac{1}{2} \\ \frac{1}{2} \end{matrix}$

Table 2.2: The different fermion generations in the Standard Model [9].

2.2 Interactions and their Bosons

As already explained, the SM combines three elementary forces: weak, strong and electromagnetic. Their respective describing theories: quantum electrodynamics, quantum flavourdynamics and quantum chromodynamics are all derived using the same assumption of local gauge invariance of the Lagrangian $\mathcal{L}$ of the underlying particle interaction.

2.2.1 Local Gauge Theory

A field theory is called locally gauge invariant if its Lagrangian is invariant under a continuous group of local transformations. Those transformations are mathematically described by Lie groups [10], i.e. there are three groups corresponding to the symmetries of the elementary forces noted before: The unitary group U(1), and the two special unitary groups SU(2) and SU(3), for the electromagnetic, weak and strong interaction, respectively.
Starting with the free Lagrangian of the corresponding system, a local gauge tranformation, for example $\phi(x) \rightarrow e^{i\Phi(x)}\phi(x)$ (ϕ representing the wave function of a particle), is required to leave the Lagrangian invariant. In order to do so, it usually has to be extended in a suitable way. This leads to a Lagrangian that contains additional terms representing the particle interaction in the form of additional gauge fields representing *bosons* mediating the different forces. This can be done in a straightforward way for the case of a transformation which is induced by the U(1) group (like in the example before), but becomes increasingly difficult when accounting the other to Lie groups. For the SU(2) and SU(3), three and eight matrices, respectively, instead of just one scalar parameter generate the transformation eventually leading to additional terms representing an interaction of the force mediators with themselves. This is due to the fact that the Algebra of the SU(2) and SU(3) is non-Abelian, whereas U(1) is Abelian.

Quantum Electrodynamics

The theory of quantum electrodynamics describes the interaction of charged particles via the electromagnetic force. Its gauge boson is massless, carrys no charge and is called photon. Since the photon itself is electrically neutral, it is understandable that it can not interact with itself. Another important feat of QED is that the

closer the interacting particles are, the stronger the forces between the particles get. This behavior is analogous to observations of the classic electromagnetic force, where larger charge values are measured when coming closer to a test charge. In particle physics this is denoted with the term *running coupling constant* [9]. This can qualitatively be understood by picturing the screening of the original charge by the polarization of the vacuum surrounding it (Fig. 2.1). In a Feynman diagram this is visualized by a loop of a lepton and its corresponding antilepton.

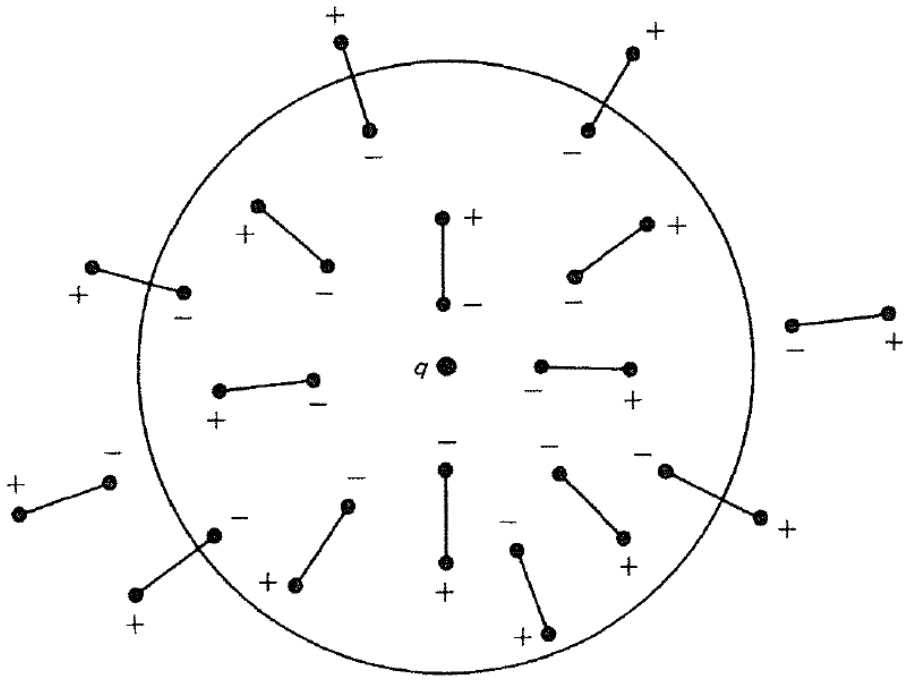


Figure 2.1: The vacuum polarization screening of a positive test charge q. Pairs of electrons and positrons are generated in the vacuum. The negative electron is attracted by the test charge while the positive positron is repelled, thus partially screening the test charge and reducing its field. Taken from [9].

Quantum Chromodynamics

The interaction of all particles carrying colour-charge is described by QCD. There are, contrary to QED, a grand total of eight interacting particles instead of just one. The force mediators are called gluons and are massless like the photon. The gluon states constitute combinations of colour and anticolour and as such gluons also interact with themselves. This self-interaction is the reason why the strong

force becomes stronger the further the particles interacting are apart and in return gets weaker the closer particles are together. A consequence of this fact is that quarks can not be observed on their own. With increasing distance, the energy in the field between strongly interacting particles becomes large enough to create quark anti-quark pairs forming hadrons. This is called *confinement*. On the other hand for small distances, i.e. high energies, the strong force diminishes and the particles become essentially free referred to as asymptotic freedom.

Quantum Flavourdynamics

The last field theory incorporated in the SM, the QFD describing the weak interaction, comes with three mediating bosons, the charged $W^{\pm}$ and the neutral Z. These particles contrary to the cases before are massive particles with masses $m_W = 80.399 \pm 0.023$ GeV and $m_Z = 91.1876 \pm 0.0021$ GeV. This implies a violation of local gauge invariance on the first view. This problem can be solved introducing the *Higgs Mechanism* [11–13] leading to a spontaneous breaking of symmetry. It proposes the instroduction of a complex scalar field ϕ in the Lagrangian, the *Higgs field*, with the potential of the form

$$U(\phi) = -\mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \tag{2.1}$$

For values $\mu^2 < 0$ and $\lambda > 0$ the global minima lie on a circle with radius $\sqrt{\frac{\mu}{2\lambda}}$

$$\phi^\dagger \phi = v = \sqrt{\tfrac{\mu}{2\lambda}}. \tag{2.2}$$

Expanding the Lagrangian about a specific ground state v (see Fig. 2.2) leads to a breaking of the interaction symmetry, but also results in mass terms for the otherwise massless gauge fields of the $W^{\pm}$ and Z. With a specific choice of gauge this results in an additional massive, scalar particle, the *Higgs boson*. This is now in accordance with the observation of the massive particles $W^{\pm}$ and Z in 1983 [14, 15]. The Higgs field also interacts with fermions via a Yukawa coupling. When the symmetry is broken by choosing a non zero vacuum expectation value for the Higgs field, this coupling splits into two parts, one representing a mass term for the fermion fields and the other representing the interaction. This interaction is believed to give fermions

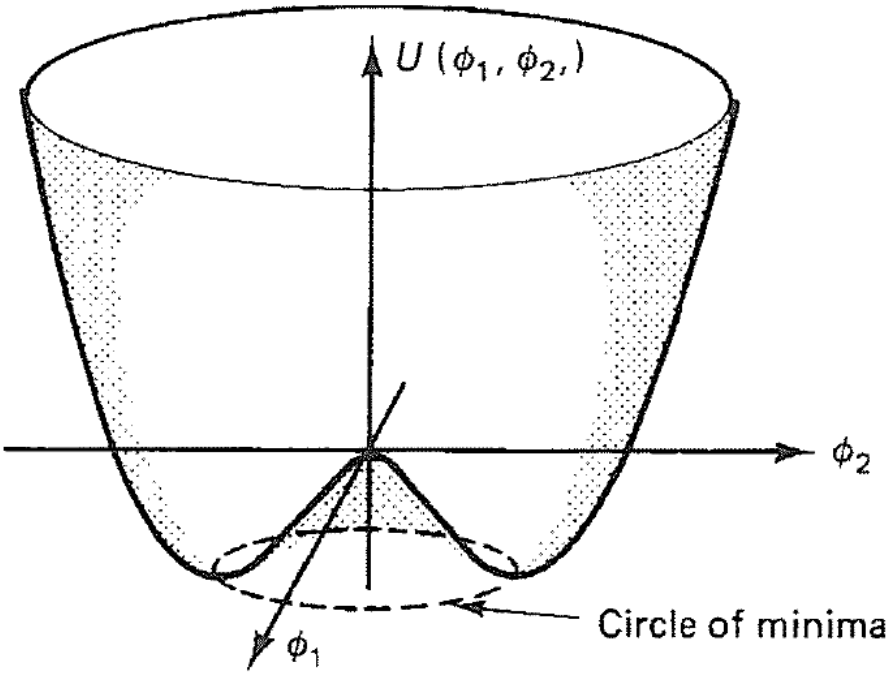


Figure 2.2: The Higgs potential $U(\phi)$ for the case of $\mu^2 < 0$. By choosing a ground state in the minimum the symmetry is spontaneously broken. Taken from [9].

their mass. A SM Higgs boson was reported to be discovered during 2012 by the ATLAS and CMS collaborations [1, 2].

Elektroweak Unification

At energies of more than 100 GeV, the electromagnetic and weak force *unify* according to the Glashow-Weinberg-Salam (GWS) theory to the electroweak force, where the electromagnetic and weak force each on their own are only appearances of the same fundamental interaction. The gauge transformation for this field theory is induced by the group $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$. The indices L and Y denote the kind of particles the unified interaction acts on; left handed particles, that are the only one that carry a weak isospin, and particles with weak hypercharge defined as $Y = 2(Q - I_3)$, where I_3 is the third component of the weak isospin and Q the electric charge.
This results in four gauge fields, that through linear combination, form two neutral bosons, the massless photon and the massive Z boson, and two oppositely charged massive particles which can be identified as $W^{\pm}$.
Furthermore, the *weak eigenstates* q' of the quarks differ from their mass eigenstates

meaning interaction via the weak interaction couples to those special eigenstates. They are defined as:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \mathbf{V} \cdot \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} \mathrm{V}_{ud} & \mathrm{V}_{us} & \mathrm{V}_{ub} \\ \mathrm{V}_{cd} & \mathrm{V}_{cs} & \mathrm{V}_{cb} \\ \mathrm{V}_{td} & \mathrm{V}_{ts} & \mathrm{V}_{tb} \end{pmatrix} \cdot \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{2.3}$$

The matrix V is called Cabbibo-Kobayashi-Maskawa (CKM) matrix [16] and represents the various couplings between different up and down type quarks even across generations. This matrix slightly differs from the unit matrix implying that quarks can change flavour and even generation when interacting via the charged weak force [9].

2.2.2 Shortcomings of the Standard Model

Although the SM was verified in various occasions and all particles it contains have been found in the over 40 years of its existence, it is nowhere able to provide a description of all physics processes known. There are many different arguments that suggest a more complex model is needed in the future to create a more complete view of physics:

- Although the conditions of the creation of our universe, the Big Bang, are still relatively unknown one can generally assume that at some point particles and antiparticles have been produced in pairs. This means that today where the average temperature in the universe is much lower, all those pairs should have recombined. This would also mean that there would be no matter at all. A small fraction of the matter visible can be explained by the CP violation of the weak interaction resulting from a complex phase factor in the CMK matrix [9], but this effect is not sufficient to explain the observable discrepancy.

- After the unification of the weak and electromagnetic interaction into the Glashow-Weinberg-Salam theory of electroweak interaction [3–5] there was a strong desire to also include the strong interaction to form a so called *Grand Unified Theory* (GUT) that covers all forces considered by the SM. However, current measurements suggest that the coupling constants of the different models do not converge at the predicted scale of approximately 10^{16} GeV (Fig. 2.3).

- Another argument is proposed by astronomical observations. The matter proposed by the SM can only explain a small part of the known energy in the universe. This fact becomes evident when looking at the rotation speed of galaxies that are not in accordance with a universe that only consists of particles observable at the moment, and a different form of so called *Dark Matter* is suggested to solve this problem [17]. The particles constituting dark matter are not consistent with any particle described by the SM
  On the other hand, the expansion speed of the universe suggests the existence of yet another form of energy, i.e. *Dark Energy*, to be consistent with the observations made [18].

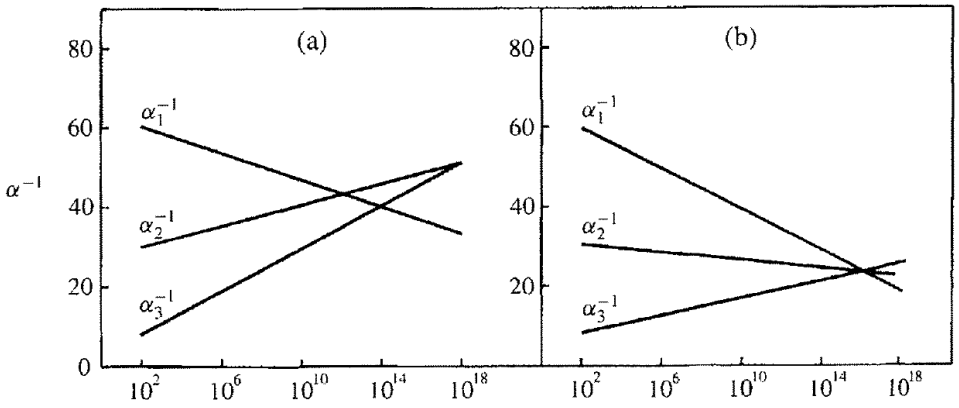


Figure 2.3: Convergence of the coupling constants at the GUT scale in (a) the Standard Model and (b) a supersymmetric model. Taken from [9].

Possible solutions for these problems can be found in extensions of the standard model introducing new particles or completely new models like the supersymmetry, that even in its most simple form, the *Minimal Supersymmetric Standard Model* (MSSM), predicts a variety of particles none of which has been observed yet. Also completely different theories using concepts like strings have been proposed, but no compelling evidence has been found supporting any of these theories.

2.3 The Top Quark

The top quark inherits a special role within the SM since its mass is substantially higher than that of any other quark $m_{top} = 173.29 \pm 0.23(\text{stat}) \pm 0.92(\text{syst})$ GeV [19], which is also the reason why the top quark was discovered as late as 1995 at

the Tevatron, namely the DØ and the CDF experiments [20, 21]. As a consequence it has very special properties. First of all top quarks do not form hadrons like the other quarks, but instead they almost immediately decay into bottom quarks via the weak interaction. This is due to the fact that the high mass also implies a very short average lifetime of the top quark in the range of $\tau_{top} \sim 5 \cdot 10^{-25}$s , whereas the hadronization, a strong process, typically takes $\sim 10^{-23}$s. Since the top quark decays almost instantly it can only be detected indirectly through its decay products, however this is the only case where the properties of a *free* quark can be measured. Also the top quark has a very special relation to the Higgs boson. In the Lagrangian of the SM quarks and leptons are assumed massless at first, but are expected to couple to the Higgs via *Yukawa-Couplings*. Those Yukawa-Couplings break up into two components when the symmetry is spontaneously broken by choosing the ground state of the Higgs potential one of which accounts for the mass of fermions as explained before. The couplings are proportional to the mass of the fermion and as such is strongest for the top-quark.

2.3.1 Production Mechanisms

Top-quarks can either be produced on their own via the weak interaction or in pairs of two via the strong interaction. This analysis focuses on the more frequent top-quark pair production.

Top-quark Pair Production

Top-quark pair production is the dominant production mechanism at modern high energy hadron colliders like the TEVATRON ($p\bar{p}$ collision) or the LHC (pp collision). Protons and antiprotons consist of quarks and gluons, thus production processes at hadron colliders are described by perturbative QCD. This approach assumes a hard scattering process between two hadrons as the interactions between its respective constituents, which are referred to as *partons*. The interaction can be divided into two parts: the short distance hard scattering process of two partons i, j interacting with a cross section σ^{ij} and the long distance part accounted for in form of parton distribution functions (PDF) $f_i(x_i, Q^2)$. These functions describe what fraction x_i of the hadron momentum falls upon the different partons of flavour i constituting the particle. This seperation is called factorization and is illustrated in Figure 2.4 [22].

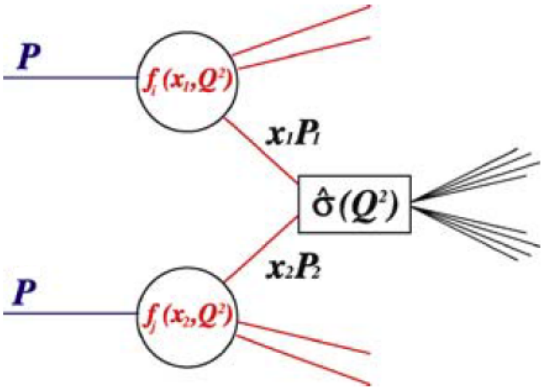


Figure 2.4: Parton model description of the hard scattering of a proton-proton collision using the factorization approach. Taken from [22].

Generally $t\bar{t}$ pairs can be produced via gluon-gluon fusion and quark-antiquark annihilation, the four lowest order diagrams are shown in Figure 2.5.

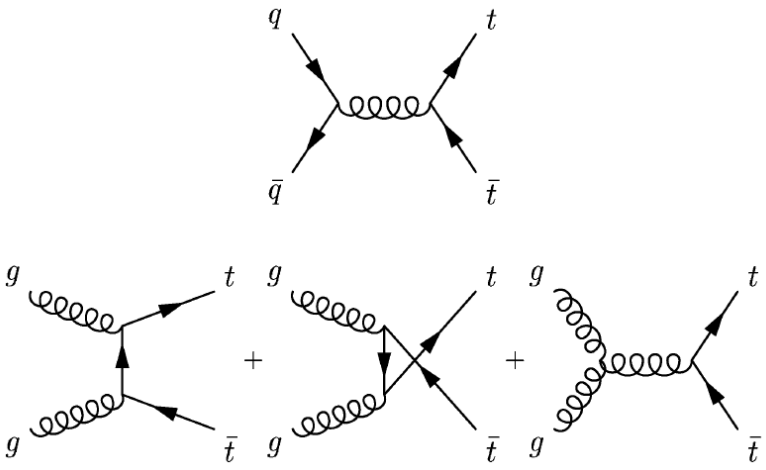


Figure 2.5: Leading order Feynman diagrams for $t\bar{t}$ production via the strong interaction at the LHC [22].

2.3.2 Associated Production with a Higgs Boson

This thesis focuses on $t\bar{t}$ production together with a Higgs boson. This production is predicted to exist by the SM theory. Its cross section is calculated to be

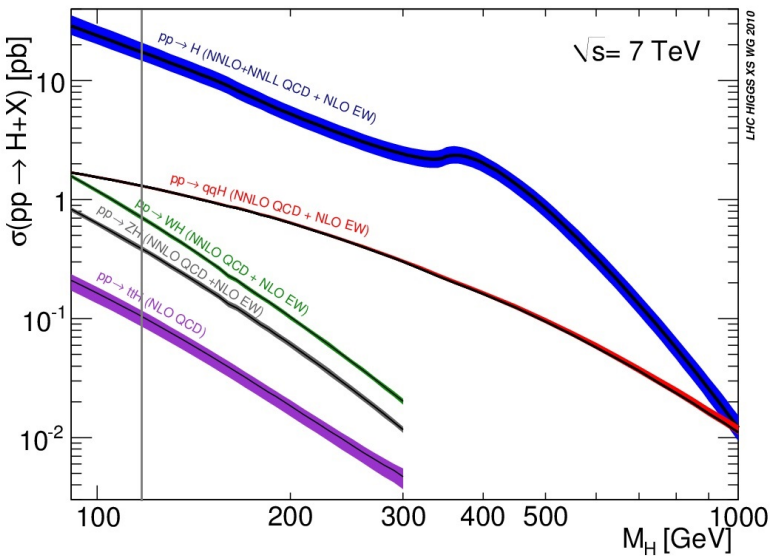


Figure 2.6: Theoretical production cross sections with error bands for a SM Higgs boson with different masses m_H. The mass 125 GeV is highlighted with a horizontal line [23].

$\sigma_{t\bar{t}H} = 0.086\,\text{pb}^{-1}$ at a Higgs mass of $m_H = 125$ GeV. In comparison to other Higgs production channels, e.g. the isolated Higgs production, the $t\bar{t}H$ cross section is rather small as can be seen in Figure 2.6. The associated production of top-quark pairs together with a Higgs boson is an interesting process, since a successful measurement of the cross section would also be sensitive to the Yukawa-Coupling between top quark and the Higgs. Three leading order Feynman diagrams of this production channel are shown in Figure 2.7. Until now, $t\bar{t}H$ production has not observed at the LHC yet. The best upper limit on the cross section times branching ratio for $t\bar{t}H(H \to b\bar{b})$ was measured by CMS [24] in the combined di-leptonic and semi-leptonic decay channel. The value is 3.8 times the SM expectation.

2.3.3 Decay Modes of Top-quark Pairs

In the beginning of this section, it was already explained that the top quark has a very short lifetime. This is due to the fact that its mass is more than twice the mass of the W-bosons leading to a significantly stronger weak interaction of the top. A interacting W-boson does not need to be virtual anymore meaning the top quark always decays into a down-type quark and a $W^{\pm}$. The down type quark will in almost every case be a b-quark since the corresponding entry in the CKM matrix

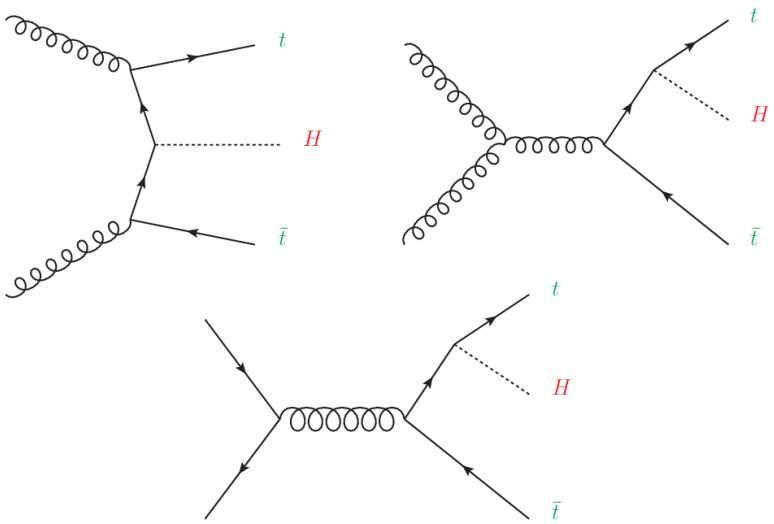


Figure 2.7: Leading order Feynman diagrams of $t\bar{t}H$ production at hadron colliders.

$|V_{tb}|$ is approximately 1. There are, in principle, three different decay modes for a system of $t\bar{t}$ or better of the two W bosons from the initial top decay. The $W^{\pm}$ can decay in a lepton and its neutrino partner or into two *light* quarks, one up-type and one down-type. It is clear that a decay back into a top quark is excluded since the mass of the W is not nearly high enough. This leads to three unique signatures for top-quark pair decays:

- **di-leptonic**($W^+W^- \to l^+l^-\nu_l\bar{\nu}_l$): both W bosons decay into opposite charged leptons and neutrinos
- **semi-leptonic**($WW \to l\nu_l q\bar{q}$): only one W boson decays leptonically, whereas the other decay hadronically
- **full-hadronic**($WW \to q\bar{q}q\bar{q}$): both W bosons decay in pairs of quarks

The individual branching ratios for W decays are close to the pure combinatorial number taking the three different colour possibilities for quarks into account. The final branching ratios for the top pair decay are visualized in Figure 2.8.
The analysis featured in this thesis focuses on the semi-leptonic $t\bar{t}$ decay, which has the added benefit that the missing transverse energy detected can be associated with the neutrino. The decay has a very clear signature that has a high chance to be detected since it features a high-energy isolated lepton suitable for trigger purposes and still has a relatively high combined branching ratio of $\frac{24}{81}$ for the electron and

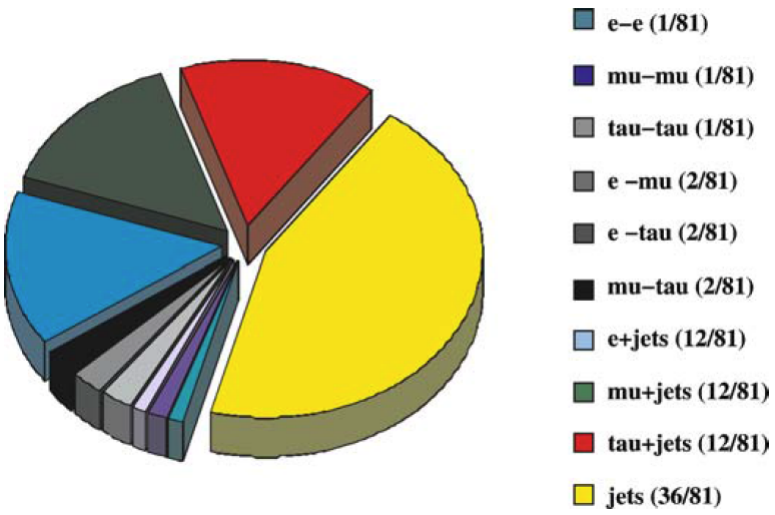


Figure 2.8: Pie chart showing the fraction of different final states of the top pair decay [22].

muon channel together.

Decay of the Higgs boson

The top pair decay is independent from the Higgs boson decay, which we assume to decay into two b-quarks in this analysis, which has also the highest predicted branching ratio for Higgs mass around $m_H = 125$ GeV of all possible processes as is shown in Figure 2.9. Although the branching ratio for a decay into two b-quarks is higher than any other the discovery of the Higgs was done in the combined decay channels $H \to \gamma\gamma, ZZ, WW$. This is due to the large background of multi-jet events at hadron colliders.

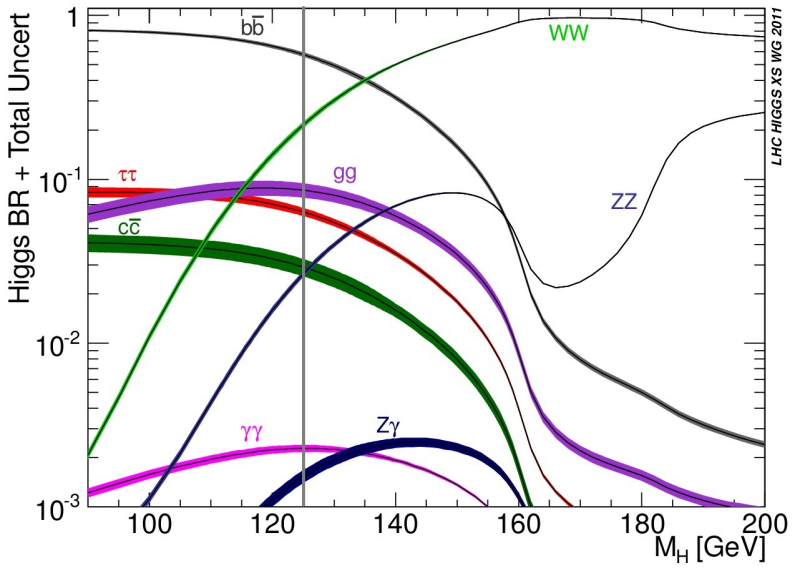


Figure 2.9: Theoretical branching ratios of the SM Higgs boson for different masses m_H. The mass 125 GeV is highlighted with a horizontal line [23].

3 Experimental Setup

The following chapter will give an overview over the Large Hadron Collider and the ATLAS multipurpose detector both located at CERN in Geneva. The different subdetector systems, the detector observables and the trigger mechanism will be described.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a double ring hadron collider that uses superconducting cavities to accelerate protons and lead ions. The design center of mass energy is $\sqrt{s} = 14$ TeV [25]. It is located in the tunnel of the former Large Electron-Positron Collider (LEP) [26] constructed between 1984 and 1989. It has a length of about 27 km, a diameter of 3.8 m and lies over 100 m under the surface of the earth shielding it from a great fraction of cosmic rays.
The LHC is filled with bunches of protons that are pre-accelerated by a number of smaller particle accelerators also located on the CERN compound. The initial protons are produced by ionization of hydrogen gas and are then sent to the Linear Particle Accelerator (LINAC 2) that raises their energy to 50 MeV. Afterwards they are transfered to the Proton Synchrotron Booster (PSB), which further accelerates them to an energy of 1.4 GeV. Following this step the protons are injected into the Proton Synchrotron, pushing the particle beam to 25 GeV. The last pre-accelerating step is the Super Proton Synchrotron (SPS) raising the energy to 450 GeV until they are finally injected into the LHC itself. The entire accelerator chain is shown in Figure 3.1.
The LHC consists of two separate beam pipes in which the particles travel in opposite direction. Also there are eight *interaction points*, where the two particle beams cross through each other allowing proton-proton interactions. On four of those interaction points experiments are located namely, ATLAS [27] and CMS [28], which are multipurpose detectors sensitive to various physic processes and two dedicated

experiments, LHCb [29], a detector concentrating on b quark physics, and ALICE [30] researching a possible quark-gluon plasma in heavy ion collisions.

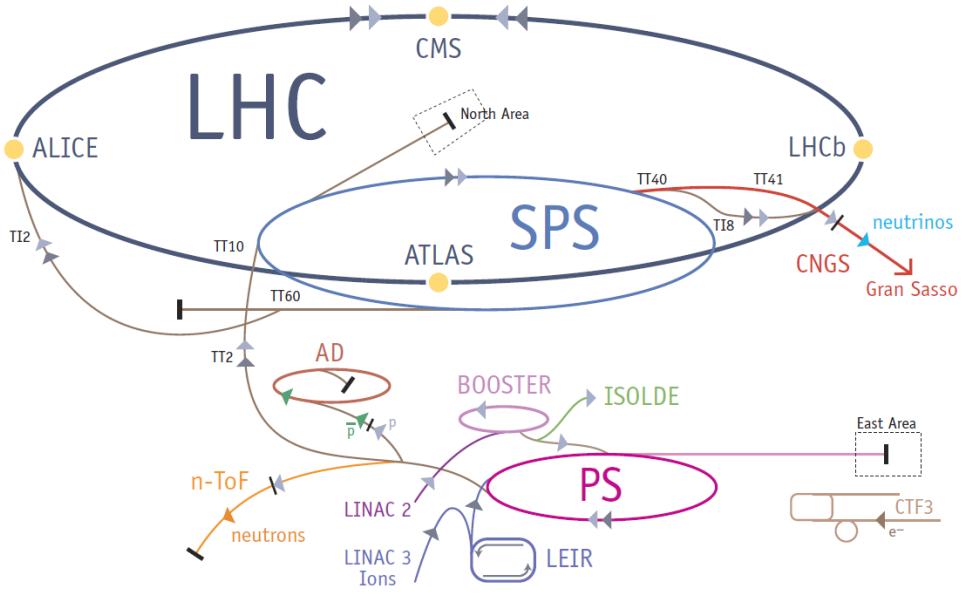


Figure 3.1: The LHC and its pre-accelerator chain as well as the location of the four main experiments [31].

3.2 The ATLAS Detector

ATLAS (A Toroidal LHC ApparatuS) is one of two multipurpose detectors located at CERN. It is composed of different subdetector systems each of which fulfills a specific task. The inner detector used to reconstruct the location of interaction vertices and reconstructing the tracks of charged particles, a calorimeter measuring the energy the different reaction products deposit inside it, a muon spectrometer to locate the hardly with other detector components interacting muons and a magnet system constituted of a solenoid and a toroidal system, all of which are described in more detail in the subsequent sections. The detector as a whole is over 40 m long, 25 m high and weighs approximately 7000 t. The ATLAS detector (see Fig. 3.2) is designed to be sensitive to a wide array of physics processes explaining its symmetric form and almost 4π angular coverage. The big difficulty is to distinguish the different reactions happening with every bunch crossing from each other, since on average 20 proton-proton reactions happen almost simultaneously [27].

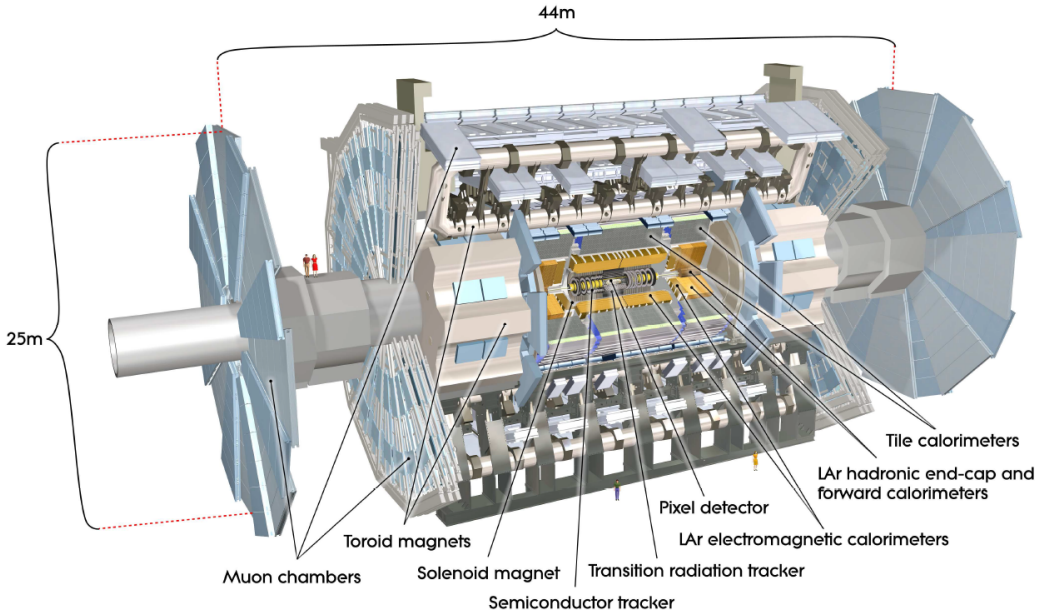


Figure 3.2: Overview over the entire ATLAS detector and its components [27].

Coordinate System

The coordinate system used to describe events recorded by the ATLAS detector is described in this section as it will be used throughout the following analysis.
The coordinate system used has its origin in the nominal interaction point at the center of the ATLAS detector. The x-axis points to the center of the LHC ring, the z-axis is parallel to the beam pipe leaving the last direction, the y-axis, to point upwards from the interaction point thus forming a right-handed coordinate system fulfilling $\vec{x} \times \vec{y} = \vec{z}$. These coordinates are often transformed into polar coordinates (r, ϕ, θ) instead, where the azimuthal angle θ is measured between $[-\pi, +\pi]$ lies in the x-y plane, the polar angle ϕ accordingly in the y-z plane and the distance r to the interaction point. Instead of the polar angle usually the *pseudorapidity* η defined as :

$$\eta = -\ln\left[\tan\frac{\theta}{2}\right], \tag{3.1}$$

is used since the difference between two pseudorapidity values $\Delta\eta$ is *Lorentz invariant* and also the particle flux in the detector is almost constant with respect to this $\Delta\eta$ quantity. Furthermore there is the distance ΔR in the pseudorapidity-azimuthal

angle space defined as

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \tag{3.2}$$

In many cases the transverse momentum p_T of particles is a variable of interest:

$$p_T = \sqrt{p_x^2 + p_y^2}. \tag{3.3}$$

3.2.1 Inner Detector

In order to distinguish the high number of simultaneous events happening with every bunch crossing a very precise reconstruction of the tracks corresponding to the charged interaction products has to be possible. By precise reconstruction of those tracks one can find interaction vertices for all the individual proton-proton collisions. The inner detector (ID) of the ATLAS experiment is able to ensure this. It consists of three subsystems: the *Pixel Detector*, the *Semiconductor Tracker* (SCT) and the *Transition Radiation Tracker* (TRT). The entire ID is immersed in a 2 T magnetic field (see Section 3.2.4) bending the tracks of charged particles allowing a measurement of their momentum.

Pixel Detector

The pixel detector is divided into a barrel part and an end-cap region covering the range of $|\eta| < 2.5$. The barrel region starts 5 cm away from the interaction point making it the detector subsystem closest to the beam pipe. It is formed by three concentric cylinders of doped silicon modules. In the end-cap region, each side constitutes three detector discs oriented perpendicular to the beam pipe. In total 1744 modules are used in the barrel and end-cap region combined, amounting to over 80,4 million separate readout channels. The individual modules consist of pixels with an area of at least $50 \times 400\ \mu\text{m}^2$ ensuring the very fine granularity needed to reconstruct the particle tracks accurate enough to distinguish interaction vertices, for example the fast decay of heavy quarks like the bottom quark, which results in a slightly displaced secondary vertex, used for *b-tagging*. The Pixel detector achieves a resolution of 10 µm and 115 µm in the $x - \phi$ and z-direction, respectively.

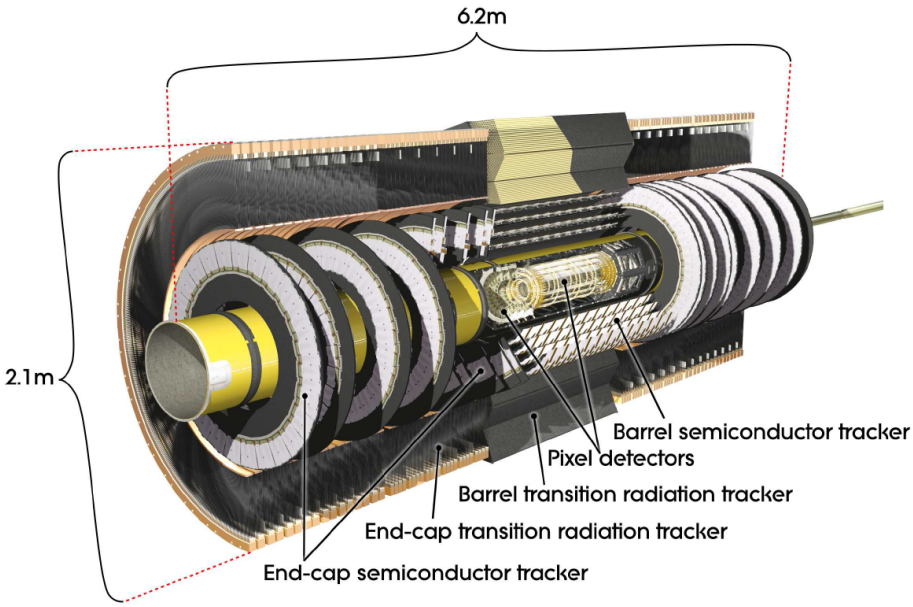


Figure 3.3: Cut-away view of the inner detector and its subcomponents [27].

Semiconductor Tracker

A similar concept is used for the Semiconductor Tracker (SCT). It consists of stereo silicon strips forming four concentric cylinders and nine end-cap discs on each side. In the stereo layers the strips are inclined by a small angle of 40 mrad allowing a simultaneous measurement of R and ϕ. The SCT consists of a total of almost 16000 strip sensors, each 6.4 cm long and having an 80 µm pitch leading to about 6.3 million individual read out channels with a resolution of 18 µm in the R-ϕ and 580 µm in the z-direction.

Transition Radiation Tracker

The last part of the inner detector, the TRT, is formed by about 350,000 drift tubes, referred to as *straws*, again divided into two different geometries for the end-cap and barrel region. The individual tubes have a diameter of 4 mm and are aligned in 73 planes parallel to the beam pipe in the barrel region, whereas the end-cap part is composed of 160 straw planes radially aligned to the beam pipe. A charged particle with $p_T > 0.5$ GeV and $|\eta| < 2.0$ will typically traverse 36 straws. The tubes are filled with a gas mixture of Xe, CO_2 and O_2, which is ionized by particles crossing them. The straws provide only R-ϕ information with an accuracy of 130 µm per straw. Also it is possible to distinguish different types of particles by the tube

resonance, since the masses of the particles directly influence the signal strength of individual tubes allowing for example discrimination between pions and electrons.

The entire inner detector is about 6.2 m long and 2.2 m in diameter. Considering all detector subsystems a total of 43 hits may be detected as illustrated in Figure 3.4.

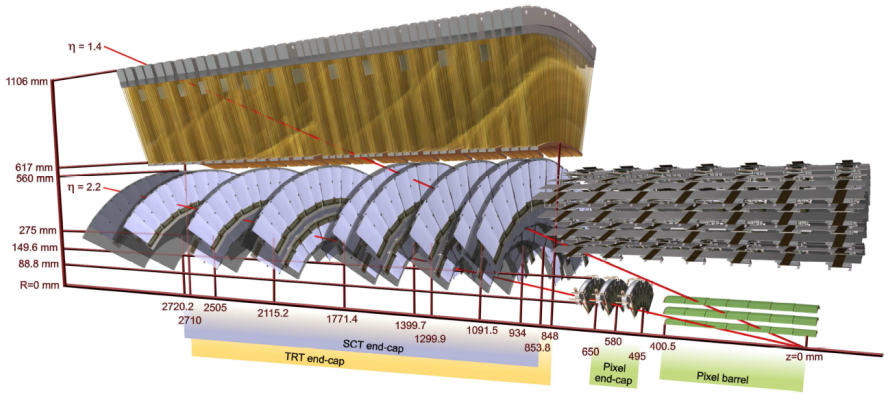


Figure 3.4: Two tracks with different η traversing the different elements of the inner detector. One track with $\eta = 1.4$ that traverses the barrel region of the pixel detector, four double layers of the SCT and about 40 straws of the TRT. The other track with higher $\eta = 2.2$ has no hits in the TRT and traverses the barrel and end-cap region of the pixel detector as well as four end cap SCT layers [27].

3.2.2 Calorimetry

The calorimeter system of the ATLAS detector is divided into 2 subsystems, the inner *electromagnetic calorimeter* (EM) and the outer *hadronic calorimeter*. As the names suggest they serve to detect different sorts of particles through their interaction with specific materials. Both are *sampling calorimeters* meaning they are composed of multiple layers, alternating between an absorber material, which induces particle reactions leading to so called *showers* and an easy ionizable active material to detect the particles emerging from the absorber material. An overview of the ATLAS calorimeters is shown in Figure 3.5. The almost full 4π angular coverage of the calorimeters ensures that almost no particles are left undetected, thus allowing the possibility to indirectly measure the missing transverse energy E_T^{miss}. Also the thickness of the combined calorimeter components provide at least

22 *radiation lengths*, ensuring a suppression of *punch-through* effects to the muon spectrometer.

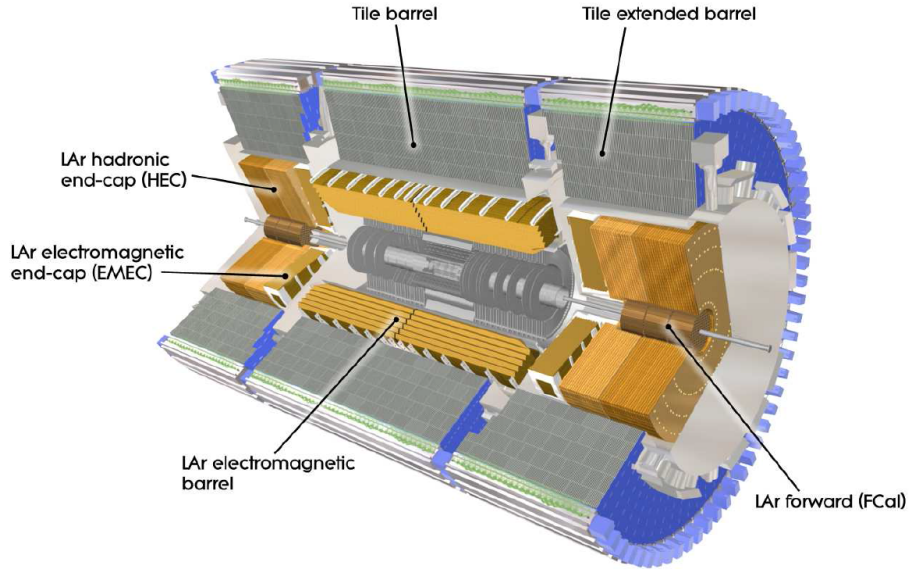


Figure 3.5: Overview over the calorimeter system of the ATLAS detector [27].

Electromagnetic Calorimeter

The EM calorimeter is again divided in a barrel part (EMB) centered around the z-axis and two end-cap wheels (EMEC), which combined cover the region $0 < |\eta| < 3.2$. It consists of alternating layers of lead and liquid Argon (LAr) aligned in a *accordion-shape*, providing a complete symmetry with respect to ϕ without any azimuthal gaps. The lead plates serving as absorber material have a variable thickness optimized for the calorimeters energy resolution, while in the LAr filled gaps kapton electrodes measure the ionization currents. The energy resolution of the EM calorimeter has been measured with electron test beams and has been verified with simulations. It is usually given in the form $\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus b$ and is given by [27]:

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}[\text{GeV}]} \oplus 0.7\%.$$

Hadronic Calorimeter

A tile calorimeter and two sample calorimeters in the end-cap region (HEC) combined are used as hadronic calorimeters. The tile calorimeter is placed directly outside of the EM calorimeter and is subdivided into a barrel region covering the region $|\eta| < 1.0$ and two extended barrels ranging from $0.7 < |\eta| < 1.7$. It is composed of steel as an absorber and scintillating tiles as active material.
The HEC covers the range between $1.5 < |\eta| < 3.2$ using cooper as active material and LAr as active material. The calorimeter achieves an energy resolution of [27]:

$$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E}[\text{GeV}]} \oplus 3.0\%.$$

Forward Calorimeter

To ensure an almost complete coverage over the entire pseudorapidity range the forward calorimeter (FCal) is integrated in the region $3.1 < |\eta| < 4.9$ and consists of three layers. The first serving as an additional EM calorimeter (FCal1) in the forward direction and the other two ensuring the full coverage of the hadronic calorimeter (FCal2+FCal3). The FCal1 module uses Copper as absorber material, whereas the other two use mainly Tungsten. In both cases LAr is used as active material.

3.2.3 Muon Spectrometer

The muon spectrometer detects the deflection of muon tracks in strong magnetic fields and is divided in two parts used for trigger purposes covering $|\eta| < 2.4$ and precision measurements covering $|\eta| < 2.7$. In the barrel region, a total of three cylindrical layers of muon chamber measure muon tracks, whereas in the end-cap region the chambers lie in planes perpendicular to the beam pipe, also in three layers. The muon spectrometer is shown in Figure 3.6
Most of the η-range is covered by *Monitored Drift Tubes* (MDT), which precisely measure the bended tracks of the deflected muons. In the range of higher pseudorapidities $2 < |\eta| < 2.7$ *Cathode Strip Chambers* are used for the precise measurement, since MDTs are not well suited for the possibly very high muon flux.
As will be explained in section 3.2.5, low response time trigger systems have to implemented into the detector to allow for preselection of interesting physics processes.

At ATLAS two systems are used, the *Resistive Plate Chambers* in the barrel region and the *Thin Gap Chambers* in the end cap region.

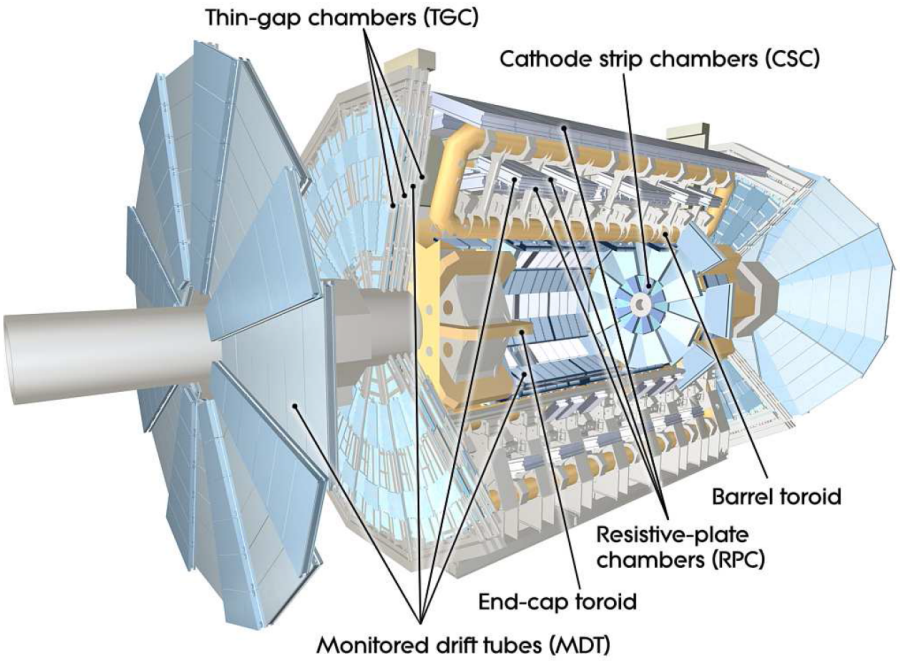


Figure 3.6: Overview over the muon spectrometer of the ATLAS detector [27].

3.2.4 Magnet System

The inner detector and the muon spectrometer require strong magnetic fields in order to bend the tracks of the energetic charged particles emerging the proton-proton interactions. These fields are produced by the ATLAS magnet system (Fig. 3.7) composing of a superconducting solenoid in between the inner detector and the calorimeter, while a number of superconducting coils in the barrel and end-cap region provide a toroidal magnetic field in the area of the muon spectrometer.
The solenoid is designed to be low weight and low density in order to minimize the energy loss of particles before they even reach the calorimeter system. It provides an axial magnetic field of 2 T in the region of the inner detector. The torroid magnet system is subdivided in two parts, a barrel toroid and end-cap toroids on each side providing a toroidal magnetic field of 0.5 T and 1.0 T inside the muon spectrometer in the central and end-cap regions, respectively [27].

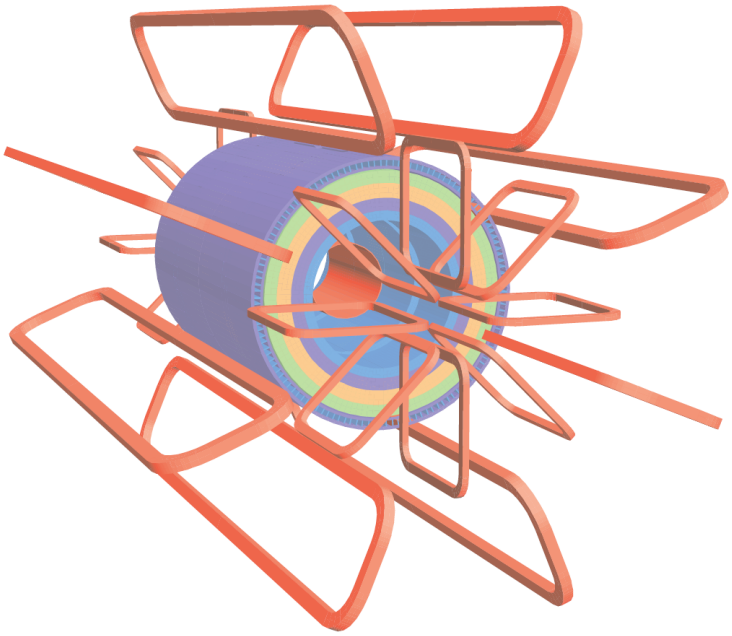


Figure 3.7: The solenoid and toroidal magnets providing the magnetic field in the ATLAS detector [27].

3.2.5 Trigger System

Since every single bunch crossing produces about 1000 individual particles almost at the same time amounting to a theoretical data rate of $60\frac{\mathrm{Tb}}{\mathrm{s}}$ a trigger system implemented in the ATLAS detector deciding whether the event at hand contains physics processes of interest for the analysis. It reduces the data rate using a three-level system.

Starting with a rate of proton-proton events of approximately 1 GHz the level one trigger (L1) searches for interesting physics objects, e.g. high transverse momentum leptons (indicating for example a $W^{\pm}$ decay). It is completely hardware based and identifies *Regions-of-Interest* (RoI) in the $\eta - \phi$ space in less than 2.5 µs. The RoIs constitute only about 2 % to the entire event data thus reducing the data rate to about 75 kHz. The second trigger level (L2) is software based, analyzing all detector subsystems in the RoIs defined by L1, further reducing the data rate to approximately 3.5 kHz in an average processing time of 40 ns. The last step, the L3, called event filter, cuts the data rate even further to a final rate of about 200 Hz, which is low enough to be recorded. This final process takes about 10 seconds and uses offline analysis procedures. The events passing the three trigger stages can

then be reconstructed into physics objects like leptons, jets or missing transverse energy.

4 Event Reconstruction

4.1 Kinematic Likelihood Fitter

In order to reconstruct the $t\bar{t}H$ events a kinematic fitting tool is used. It employs a likelihood approach which focuses on the final states object energy resolution. The KLFitter package is a C++ program that uses a maximum likelihood method in order to reconstruct general physics processes. In this study it is used to reconstruct $t\bar{t}H$ events, where the top pair decays semi-leptonically and the Higgs boson decays into two b-quarks. This decay results in a distinctive signature of particles in the final state, as discussed in Chapter 2. First of all, a high-p_T lepton together with missing transverse energy from the leptonic decay of a W are expected to be found. Also exactly six jets should be present if all are detected correctly and no initial/final state radiation occurred should be visible. Two light quark jets coming from the decay of the other W and four b quark jets from the decays of the top quarks and the Higgs boson are expected. These six jets add up to $6! = 720$ possible combinations of associating a reconstructed jet to a final state parton of the $t\bar{t}H$ decay. Since the two b-quarks from the Higgs boson decay and the light jets from the hadronically decaying W are indistinguishable only $\frac{6!}{2\times 2} = 180$ permutations remain.

4.2 Model Assumptions

The particles mentioned before are characterized by the following measured quantities:

- the energies $\tilde{E}_i$ and directions $\tilde{\Omega}_i$ of the six jets,
- the energy $\tilde{E}_l$ and direction $\tilde{\Omega}_l$ of the charged lepton,
- the missing transverse energy E_T^{miss} corresponding to the neutrino.

The direction of the jets and leptons are assumed to be measured with a negligible uncertainty. The energies of the jets (quarks) and leptons can only be measured with limited precision due to the corresponding uncertainty of angular and energy resolution of the detector systems and due the identification methods described in Section 5.2. These uncertainties are expressed by so called *transfer functions* $W(E_i, \tilde{E}_i)$, which vary in different η- and energy-regions due to the different uncertainties connected to the various regions in the detector. The transfer functions map the detector response of jets to the particles of the underlying hard scattering process. They denote the probability $W(x|y)$ to measure a certain response y given a truth value x. Using those transfer functions in the likelihood function one can not only map the various jets to the underlying events particles but also improve the resolution of measured quantities. The transfer functions will be described in more detail in Section 4.4.
Furthermore kinematic constraints are imposed to improve the reconstruction. The invariant masses of the light jets and the combination of the missing transverse energy and the lepton, which correspond to decays of W bosons are required to follow a Breit-Wigner (BW) distribution around the measured W mass of $m_W = 80.4$ GeV with a width of $\Gamma_W = 2.1$ GeV [32]. This is expressed by the normalized and relativistic probability density function of a mass m being BW distributed around a central mass of M:

$$BW(m|M) = \frac{2}{\pi} \frac{\Gamma M^2}{(m^2 - M^2)^2 + \Gamma^2 M^2}. \tag{4.1}$$

Additionally the invariant top quark masses consisting of either three jets in the case of hadronic decay and a jet, a lepton and missing transverse energy in the leptonic case are also required to follow a BW distribution around the top quark mass. The parameter can be left open resulting in an additional parameter during the kinematic fit, but in this analysis it is set to the value $m_{top} = 172.5$ GeV, which is in accordance with the mass used for simulating the data sample (see also Section 5), further the width is set to $\Gamma_{top} = 1.5$ GeV.
Another constraint is imposed on the invariant mass of the two b-quarks originating from the Higgs boson decay, again the mass should follow a BW distribution. There are two different options studied in this thesis, a baseline model where the central Higgs mass is left as a free parameter in the fit and another where a central Higgs

mass of $m_H = 125$ GeV is set. In both cases a narrow decay width of $\Gamma_H = 0.0035$ GeV is assumed. Each constraint decreases the degree of freedom k in the kinematic fit, which has to be at least one to enable the kinematic fit. The total degree of freedom k is given by the number of fit parameters M, the number of variables N and the number of constraints P:

$$k = N - M + P \tag{4.2}$$

4.3 Fit Parameters

Since the directions of the final state objects are assumed to be measured precisely the parameters varied in the maximization of the likelihood are reduced to:

- the energies E_i of the six jets amounting to 6 parameters,
- the energy E_l of the charged lepton adding 1 parameter,
- the three momentum $\vec{p}^{\nu}$ of the neutrino resulting in 3 parameters
- and 1 parameter for the optional free Higgs mass m_H .

The different parameters are varied during the kinematic fit, with individual ranges for each event depending on the measured energy of the different objects. For jets the energy is varied in the interval $[\min(m, \tilde{E} - 7 \cdot \sqrt{\tilde{E}}), \tilde{E} + 7 \cdot \sqrt{\tilde{E}}]$,where m denotes the mass of the model particle, i.e. light quarks and b quarks. For leptons the range is deduced in a similar fashion $[\min(0.001\text{GeV}, \tilde{\text{E}} - 2 \cdot \sqrt{\tilde{\text{E}}}), \tilde{\text{E}} + 2 \cdot \sqrt{\tilde{\text{E}}}]$. The much greater energy range for quarks can be explained by the bigger uncertainty of energy measurements in the hadronic calorimeter. For the missing transverse energy in the x- and y-direction a range of $[E_{x,y}^{miss} - 100\text{GeV}, \text{E}_{\text{x,y}}^{\text{miss}} + 100\text{GeV}]$ is set, while the z-component is constricted to a range of $[-1000\text{GeV}, +1000\text{GeV}]$. Those parameters are varied to maximize a likelihood function of the form

$$\begin{aligned}
\mathcal{L} = \; & W_b \left(\tilde{E}_{jet_1} | E_{b_{had}}\right) \times W_b \left(\tilde{E}_{jet_2} | E_{b_{lep}}\right) \times W_l q \left(\tilde{E}_{jet_3} | E_{q_1}\right) \times W_l q \left(\tilde{E}_{jet_4} | E_{q_2}\right) \times W_b \left(\tilde{E}_{jet_5} | E_b\right) \\
& \times \; W_b \left(\tilde{E}_{jet_6} | E_{\tilde{b}}\right) \times W_{miss} \left(\tilde{E}_x^{miss} | p_x^{\nu}\right) \times W_{miss} \left(\tilde{E}_y^{miss} | p_y^{\nu}\right) \times W_{lep} \left(\tilde{E}_{lep} | E_{lep}\right) \\
& \times \; BW \left[m_{q_1 q_2} | m_W, \Gamma_W\right] \times BW \left[m_{l\nu} | m_W, \Gamma_W\right] \\
& \times \; BW \left[m_{q_1 q_2 b_{had}} | m_t, \Gamma_t\right] \times BW \left[m_{l\nu b_{lep}} | m_t, \Gamma_t\right] \times BW \left[m_{q_5 q_6} | m_H, \Gamma_t\right],
\end{aligned} \tag{4.3}$$

for every possible permutation. The incorporation of the Higgs boson as a part of the KLFitter model is different from the analysis performed in the ATLAS environment until now, since the observation of the particle has only just been confirmed during the last two years. It is convenient to rather maximize the logarithm of this function, since for small individual likelihood components the numerical precision might suffer from the multiplication. The maximization process is done using the BAT package [33], that incorporates the numerical optimization algorithm Minuit. Starting point for the variation is always the set of measured values coming from the event data.

4.4 Transfer Functions

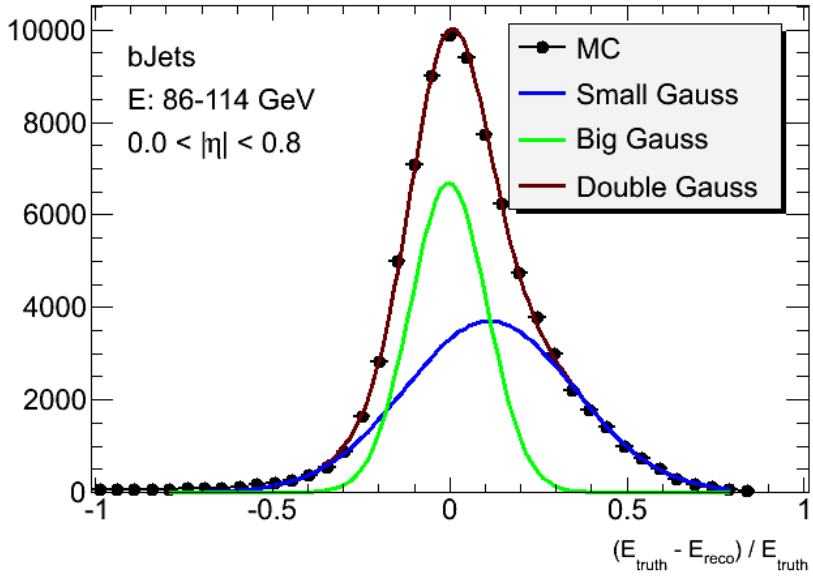


Figure 4.1: Parametrisation of the transfer function $0 < \eta < 0.8$ for a b quark with a double Gaussian function and the sub component single Gaussian functions.

Transfer functions are derived individually for different physics objects, i.e. leptons (electrons and muons), light quarks, b quarks and missing transverse energy depending on the η that region the objects are found in. The complete eta range is divided into regions corresponding to the detector geometry and taking care of the various uncertainties in the energy measurement and other effects like mis-identification of electrons as jets. The detector response is usually not symmetrically distributed

around the true value and as such they are derived by fitting double Gaussian distributions in $\Delta E = \frac{E_{truth} - E_{reco}}{E_{truth}}$ of the form:

$$W(\Delta E) = \frac{1}{\sqrt{2\pi} \cdot (p_2 + p_3 p_5)} \left[\exp\left(\frac{(\Delta E - p_1)^2}{2p_2^2} \right) + p_3 \exp\left(\frac{(\Delta E - p_4)^2}{2p_5^2} \right) \right], \quad (4.4)$$

to a simulated MC@NLO $t\bar{t}$ sample as indicated in Figure 4.1. This is done via a two dimensional likelihood, one dimension being δE and the other the energy of the reconstructed particle, where the bins are chosen such that the statistics in the individual bins are approximately the same. The missing transverse energy is modeled by a single Gaussian distribution. The different η-regions considered vary for the different physics objects:

- For jets five region are defined [0, 0.8, 1.37, 1.52, 2.5, 4.5], the splittings are the same for light and b jets.
- Electron transfer functions are divided in four regions [0, 0.8, 1.37, 1.52, 2.5], where the region $1.37 < |\eta| < 1.52$ is excluded, since events with electrons in this region are not considered for reconstruction, due to the uncertainties that arise in the measurements inside the transition region between barrel and end-cap.
- Muon transfer functions are divided in three regions [0, 1.11, 1.25, 2.5] once again reflecting the detector properties.

The example for the derivation of a transfer function shown in Figure 4.1 affirms the non-symmetric behavior of the distribution. The transfer function for the same η region but different E_{truth} is shown in Figure 4.2.

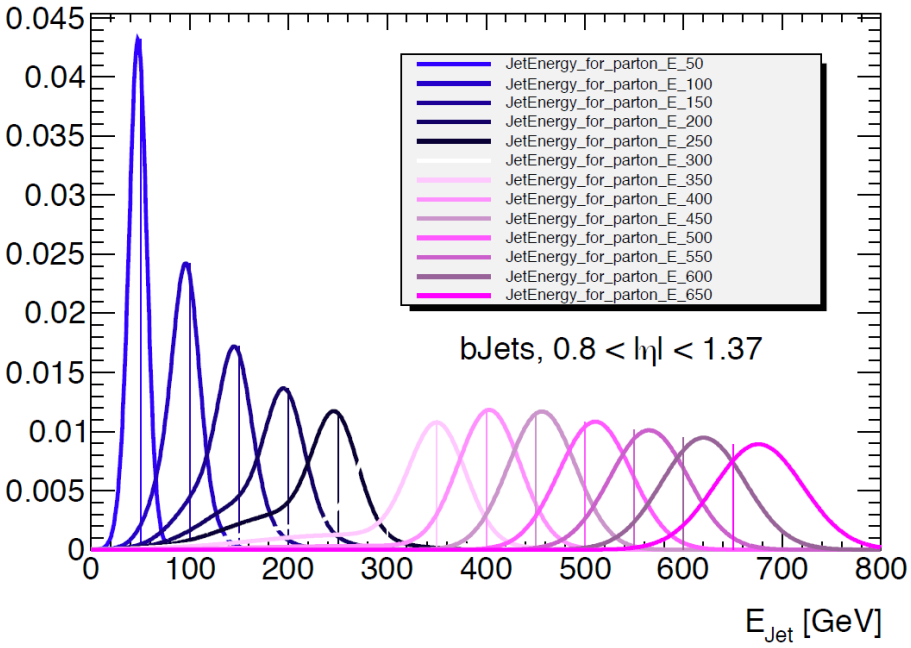


Figure 4.2: Transfer function for b-quark jets in detector range $0 < \eta < 0.8$ for different E_{truth} values.

4.5 Tuning of the Reconstruction

In order to improve the reconstruction a series of additional modifications are employed in the algorithm using different jet properties not yet considered by the purely kinematic approach described before. Investigating the reconstruction efficiencies using these additional options is the focus of this thesis.

4.5.1 B-Tag Weight

The b-tagging algorithm give every jet a so called *b-tag weight* w_{tag}, which reflects the chance of this jet actually being a jet originating from a b-quark. A cut on the b-tag wheight is applied, where all jets with $w_{tag} > w_{cut} \approx 0.6$ are assumed to be b-jets. This leads to an efficiency of about 70 % that a jet coming from the hadronization of a b quark is correctly tagged. The weight can be used to improve the reconstruction in two ways

- vetoing permutations, which feature a b tagged jet assigned to a light quark in the reconstruction and vice versa,
- assigning an additional probability term to permutations every time a jet is

incorrectly assigned depending on the efficiency mentioned before and a rejection factor that is correlated to the probability to incorrectly assign a light quark to a b jet.

4.5.2 Charge Information

If the charges of the final state partons are known an additional constraint on the reconstructed particles can be implemented. The $t\bar{t}H$ decay features a very distinctive signature, where the charge of the resulting particles is defined by whether the decay of the top or the antitop quark finally results in the lepton neutrino pair. The top quark for example decays into a b quark and a positively charged W^+. This W^+, if it decays leptonically, will always result in a positively charged lepton, thus also predefining the charge of all other particles of the decay, except for the interchangeable particles from the Higgs and hadronic decay of the second W-boson. One can now define conditions pairs of measured particles have to fulfill if they originate from a $t\bar{t}H$-decay:

- The product of the charge of the lepton and the b quark from the leptonic decay of on top quark should be < 0,
- the product of the charge of the lepton and the b quark from the hadronic decay of on top quark should be > 0,
- the product of the charge of the hadronic decay products of the $W^{\pm}$ should always reflect the charge of the W-boson,
- the product of the two b quarks from the decay of the Higgs should always be negative.

These assumptions are visualized in Figure 4.3.

The difficulty in this approach lays in the determination of the charge of the partons invoking a jet in the calorimeter. The lepton charge is easy to reconstruct since its trajectory bends in the strong magnetic field in the inner detector according to its charge allowing a precise measurement. Two different methods are considered for the assignment of a charge to the jets:

- The charge of the jet, and as such the charge of the particle causing it, is simply set to the charge of the highest p_T-track found when reconstructing the

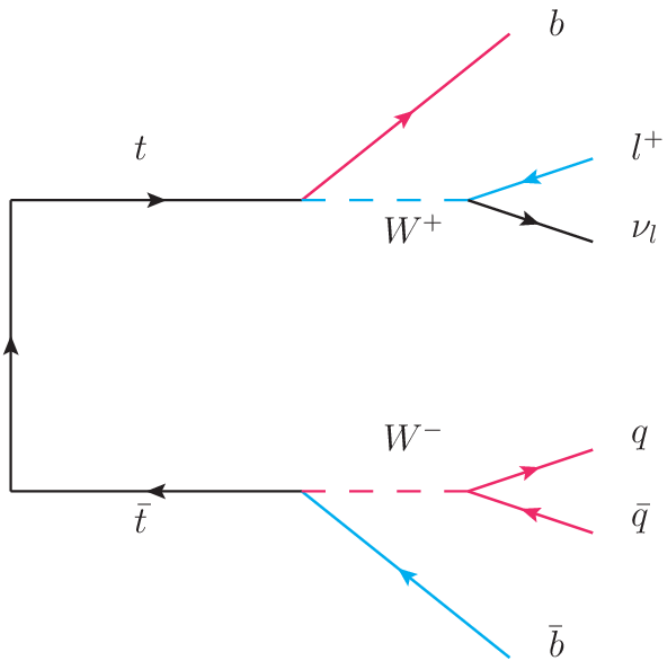


Figure 4.3: Charge of the different decay products of a $t\bar{t}$ decay, positive (negative) in blue (red).

jet. The charge is always set to ± 1 only depending on the sign of the selected highest p_T-track,

- a method that weights the charge of different tracks forming the jet according to their fraction of the total p_T. The charge of the jet is defined as $q_{jet} = \frac{\sum_i q_i |\vec{jt}_i|^k}{\sum_i |\vec{jt}_i|^k}$, where $\vec{jt}_i$ is the projection of the subtrack momentum on the direction of the jet, this method has successfully been used in studies of the top charge [34].

The lepton charge multiplied with the hadronic b jet charge calculated from the simulated $t\bar{t}H$ sample is shown in Figure 4.4 assuming the maximum track p_T method, as explained before. The distribution is not symmetric thus allowing for an effective reweighting of the significance of the events according to the proportion of correctly reconstructed charge products. In roughly 60 % of the events, the expected combination of jet charges is found while the remaining 40 % show the wrong sign. These values are now used to reweight permutations. For the leptonic b quark similar values are obtained. This additional reweighting happens after the maximization of the likelihood.

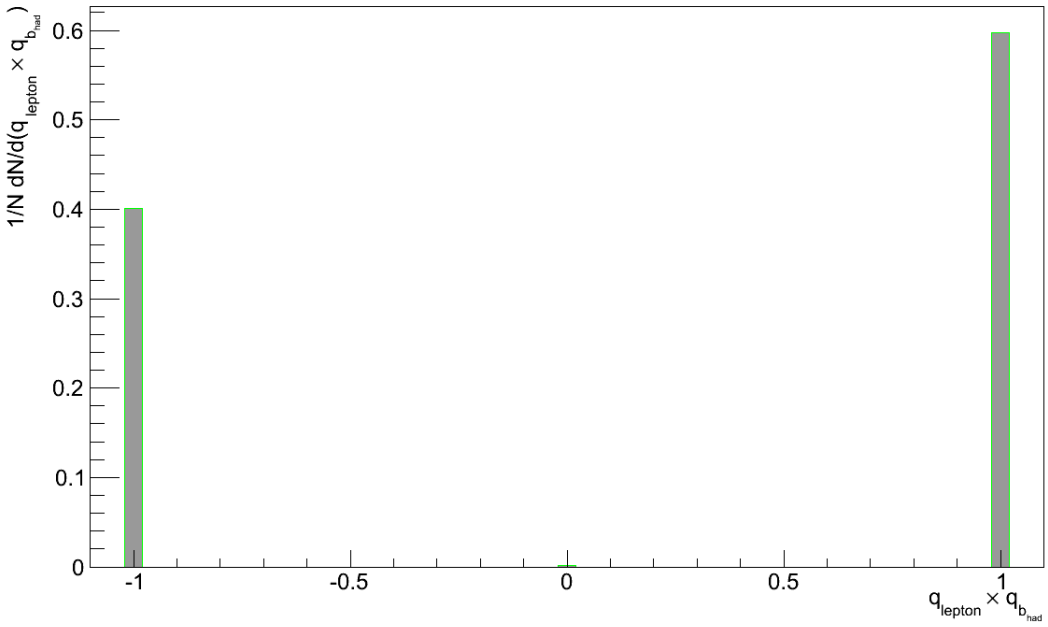


Figure 4.4: Charge of the lepton multiplied with the charge of the hadronic b jet.

4.5.3 Angular Distribution of Higgs Decay Products

In the rest system of the Higgs boson it should decay into two b quarks that leave in exactly 180 degree angle from each other. This becomes evident when assuming momentum conversion of the resting Higgs particle, the truth information from simulation indeed show a delta peak at a $\pi = 180°$.

This information might be useful to improve the reconstruction of the Higgs boson by imposing a constraint, that validates the angular direction of the b quarks comparing for each permutation, if they indeed decay 180 degrees away from each other. In this analysis this is done by adding an extra component to the likelihood of the form:

$$P_{angle}(\Delta\alpha_{b\bar{b}}) = \exp\left(\frac{(\Delta\alpha_{b\bar{b}} - \pi)^2}{\sigma_{angle}}\right), \tag{4.5}$$

where $\Delta\alpha_{b\bar{b}}$ is the angle between the two b quarks in the rest frame of the Higgs boson and σ_{angle} is a factor, that is chosen according to the resolution of the measurement of this angle.

5 Monte Carlo Sample

In order to evaluate the reconstruction efficiencies of the kinematic fit, the analysis is carried out with a sample of simulated signal data. This is sensible, since in normal data one can not determine the real underlying events, whereas in a Monte Carlo generated sample the truth information of the underlying process and the kinematic properties of the different hard process reaction products are also saved during the generation.

5.1 Monte Carlo Generator

The used sample consists of roughly 250,000 $t\bar{t}H \rightarrow l\nu_l b\bar{b}b\bar{b}q\bar{q}$, where the top quark pair decays semi-leptonically and the Higgs decays into two b quarks at a centre-of-mass energy $\sqrt{s} = 7$ TeV. It is generated using the leading-order (LO) generator Pythia 6.4.25 and the MRST LO** set of parton distribution functions (PDF) [35]. The Higgs mass is set to 125 GeV, while the top mass is set to 172.5 GeV.
The ATLAS detector simulation [36], based on Geant4 [37], is used to simulate the response of the detector and triggers. The events are reconstructed following the definitions given in the next section.

5.2 Object Definitions

The physics objects that are used in this search are electrons, muons, missing transverse energy and jets, both light and heavy flavor. The criteria used to reconstruct the events is summarized in the following section, the same criteria have been used in a dedicated search for $t\bar{t}H$ at ATLAS already [35].

Electrons

Candidates for electron events are reconstructed from energy deposits, so called *clusters*, in the electromagnetic calorimeter that can be associated with the according tracks reconstructed in the inner detector. Moreover they are required to have a transverse energy E_T of at least 25 GeV and have to be found within a detector range of $|\eta| < 2.47$. Clusters in the transition region from barrel to end-cap $1.37 < |\eta| < 1.52$ are excluded. To suppress the fraction of non-prompt electrons from hadron decays a 90 % isolation efficiency cut on the energy sum of the calorimeter cells in a cone of radius $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} = 0.2$ is required.

Muons

Muon candidates are reconstructed from track segments detected in the different layers of the muon spectrometer, that again have to be matched to the tracks reconstructed in the inner detector. Candidates fulfilling these requirements are then refitted using the information of both detector subsystems and should then also satisfy transverse momentum $p_T > 20$ GeV and $|\eta| < 2.5$. Furthermore muons are required to be separated from reconstructed jets by $\Delta R > 0.4$. Additionally the candidates should have a calorimeter isolation transverse energy within a cone of $\Delta R < 0.2$ of less than 4 GeV after excluding muon energy deposits in the calorimeter.

Missing Transverse Energy

This quantity is identified with the transverse momentum of the neutrino originating from the leptonic W decay of one of the top quarks in the $t\bar{t}H$ event. It is reconstructed from first reconstructing jets and leptons from energy clusters in the calorimeters, which are then calibrated for energy losses. The missing transverse energy E_T^{miss} is then calculated from the vector sum of the calibrated cluster momenta, together with the muon momenta from the muon spectrometer.

Light Jets

Jets are reconstructed with an anti-k$_t$ algorithm [38] with a cone radius of $R = 0.4$ from topological clusters built from energy deposits in the calorimeters. After an energy calibration the jets are required to have $p_T > 25$ GeV and $|\eta| < 2.5$. To exclude jets from secondary pp interactions at least 75 % of the sum of the p_T of

tracks with $p_T > 1$ GeV has to be compatible with originating from the identified hard -scatter primary vertex. If a jet lies within $\Delta R < 0.2$ of a reconstructed electron it is discarded to avoid double counting of electrons as jets.

Bottom Jets

A distinction is made between jets originating from light quarks (u, d, s, c) and bottom quarks. This is done via a b-tagging algorithm [39], which uses multivariate techniques based on, e.g. information of tracks coming from displaced secondary vertices.

5.3 Preselection

Using these object definitions an algorithm pre-selects suitable events with the following properties: An event that was accepted by the trigger, should have a reconstructed primary vertex with at least four associated tracks that is consistent with the nominal collision region. Events are supposed to have exactly one reconstructed lepton and at least six jets satisfying the requirements noted before. The reconstructed lepton has to be matched to the high-level trigger lepton within $\Delta R < 0.15$. Furthermore, a cut on E_T^{miss} as well as on the mass of a possible W boson defined as $m_T = \sqrt{2p_T^l E_T^{miss}(1 - \cos\Delta\phi)}$ is employed that would suppress the fraction of multijet background in the selection substantially. The cuts are different for the electron and muon channel $E_T^{miss} > 30$ GeV, $m_T > 30$ GeV and $E_T^{miss} > 20$GeV, $E_T^{miss} + m_T > 60$GeV, respectively. Also at least one b-tagged jet has to be present.

5.4 Truth Matching

In order to evaluate the quality of the reconstructed events it is necessary to compare the truth information of the simulated event with the association of jets with model particles during the kinematic fit. This is done by requiring a *truth matching* criterion that evaluates geometrical distances in the $\eta - \phi$ space. A jet and a truth particle are called matched if they fulfill:

$$\Delta R = \sqrt{(\phi_{reco} - \phi_{truth})^2 + (\eta_{reco} - \eta_{truth})^2} < 0.3. \tag{5.1}$$

For electrons, a tighter matching criterion of $\Delta R < 0.1$ is required since the direction of leptons can be measured with a higher precision. The entire event is called *matched* if every jet selected for the reconstruction with the KLFitter can be successfully matched with the truth particles from the hard scattering process. Furthermore, in a matched event every truth parton can only be assigned to exactly one jet and no two partons can be matched to more than one jet.
In this sample $N_{tot} = 17167$ events pass the preselection process, while another $N_M = 2949$ can be matched to the six partons of the hard scattering process as described above. This means a *matching efficiency* $\epsilon_M = \frac{N_M}{N_{tot}}$ of 17.18 % is achieved for this MC data sample. Allowing more jets to be used by KLFitter would obviously result in a higher matching efficiency since more jets would be considered in the matching process.

6 Search for $t\bar{t}H$ at ATLAS

A search for $t\bar{t}H$ has been carried out by the ATLAS collaboration using 4.7 fb^{-1} of data at $\sqrt{s} = 7$ TeV collected during 2011 [35]. The search focused on the semi-leptonic decay mode of the $t\bar{t}$ as described before and used the KLFitter to reconstruct events in the main signal channels. During the search the events were classified into nine different topologies depending on their jet multiplicity and the number of b tagged jets to improve the sensitivity of the search. No significant excess of events above the background expectation was found in this search, but 95% confidence-level upper limits on the observed and expected production cross section times branching ratio $\sigma(t\bar{t}H) \times BR(H \to b\bar{b})$, were derived for a standard model Higgs boson with a mass between 110 - 140 GeV.

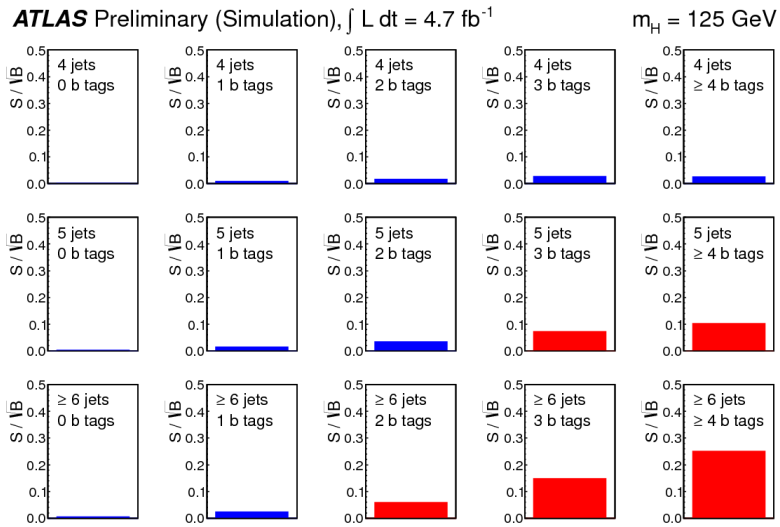


Figure 6.1: Signal-over-background ratio for different event topologies [35].

6.1 Event Topologies

All events that passed the preselection process similar to the one described in Section 5.3. Nine different topologies are considered in the search that are analyzed separately, divided into two groups: a signal region and a region that is used to constrain background and different uncertainties. The signal region features the topologies with (5,6) jets and (3,4) b tags which have a higher signal-to-background ratio than the remaining ones. The remaining five topologies are 4 jets with (0,1,2) b tags and also (5,6) jets with only 2 b tags. The signal-over-background ratio yields of different topologies is shown in figure 6.1.

6.2 Fitting of the main Signal Topologies

The KLFitter was used to reconstruct the main discriminant variable in this analysis, the reconstructed mass $m_{b\bar{b}}$ of a possible Higgs candidate, using an approach similar to the one described in section 4.1. The KLFitter model did not contain the BW constraint on the Higgs mass. Using a total of six jets into the reconstruction the focus was on reconstructing the $t\bar{t}$ system correctly and simply assuming the remaining jets that were not identified to come from the top-quark pair decay to originate from the Higgs boson decay. Simulation studies showed that with a probability of 26 % the correct b quarks were identified with the Higgs for events that featured six jets in total and four b tagged jets.

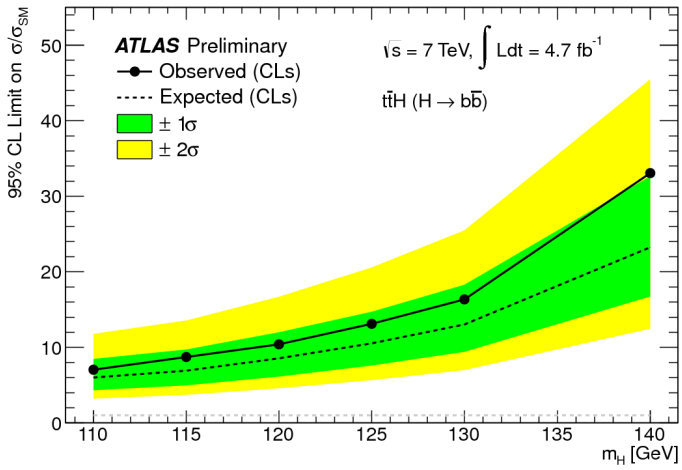


Figure 6.2: Observed and expected (median, for the background-only hypothesis) 95% CL upper limits on $\sigma(t\bar{t}H) \times BR(H \rightarrow b\bar{b})$ relative to the SM prediction, as functions of m_H. [35].

6.3 Results

The nine different event topologies are combined using a profile likelihood fit in order to exploit the high statistic background topologies. No significant excess in events could be found during the analysis as can be seen in figure 6.2. For a standard model Higgs boson with a mass of 126 GeV an observed (expected) 95% confidence-level upper limit on $\sigma(t\bar{t}H) \times BR(H \rightarrow b\bar{b})$ times the standard model cross section was found to be 13.1 (10.5). A possible improvement of this result could be realized by increasing the reconstruction efficiency of the Higgs boson, which is my ambition during my master thesis.

7 Reconstruction of $t\bar{t}H(H \to b\bar{b})$ with the KLFitter

The simulated data sample described in Section 5.1 is processed with the KLFitter to evaluate the reconstruction efficiencies for different settings of the KLFitter, i.e. the combination of the various additional constraints described in Chapter 4. The analysis is carried out using the different settings:

- the baseline model using the likelihood given in equation 4.3 with a free Higgs boson mass,
- the baseline model adding the inclusion of the charge information using the maximum charge approach as described in section 4.5.2,
- the baseline model with the addition of a constraint on the angular distribution of the two b quarks as described in section 4.5.3,
- the baseline model with a fixed Higgs mass of $m_H = 125$ GeV,
- and finally all the previous options combined.

Not only the reconstruction efficiency for matched and all events are evaluated but also the kinematic distributions of the reconstructed Higgs particle are studied in order to estimate the performance of the KLFitter and to see if the reconstructed distributions are positively influenced by choosing different settings.

7.1 Reconstruction Efficiencies

7.1.1 Baseline Model

The reconstruction efficiency ϵ_R is defined as the fraction of events, where the KLFitter suggested the correct permutation of jets matching the truth particles from

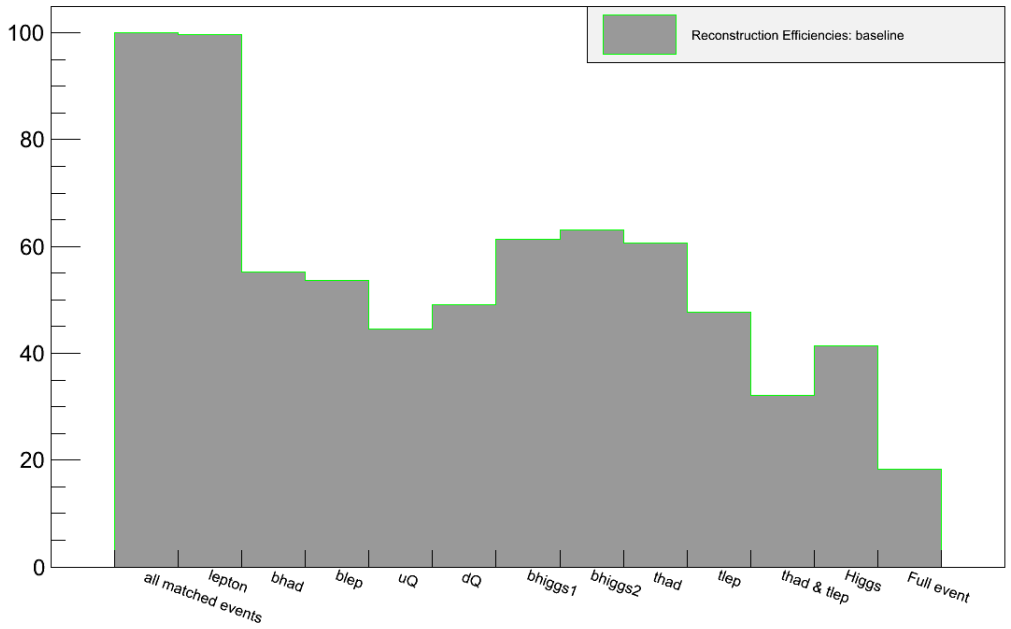


Figure 7.1: Reconstruction efficiencies of different decay end products of the $t\bar{t}H$ decay for the baseline model for matched events only.

the simulation. It is also interesting to take a look at the efficiencies not only for the correct reconstruction of all particles but individual particles, in particular the Higgs boson individually. When assuming only matched events, events fulfilling the condition explained in Section 5.4, one can compare the efficiencies to the pure combinatoric ones, which can be calculated assuming the 180 possible permutations and considering all possible combinations of jets that are invariant, e.g. for the Higgs boson this would yield $\epsilon_{H,comb} = \frac{2}{6 \cdot 5} = 6.67$ %. The calculated efficiencies are compared to the combinatoric ones in Table 7.1 and are shown for selected final state particles in Figure 7.1.

In comparison to the statistical values, the KLFitter achieves a massive improvement with the baseline model alone. This can be seen for all considered efficiencies. The correct permutation for all jets could be found in 18.21 % of the events, while the pure statistical probability is only $\frac{1}{180} = 0.56$ % meaning an improvement by a factor of about 32. The quantity that is most important for this study is the fraction of events where the Higgs boson is correctly reconstructed, which is the case for 41.37 % of the matched events compared to 6.67 % from pure combinatorics.

	Statistical Probability in %	Efficiencies in % matched events	unmatched events
$\epsilon(b_{had})$	16.67	55.02 ±1.37	35.43 ±0.45
$\epsilon(b_{lep})$	16.67	53.65 ±1.35	42.78 ±0.50
ϵ(Higgs)	6.67	41.37 ±1.19	21.05 ±0.35
ϵ_R	0.56	18.21 ±0.79	3.13 ±0.14

Table 7.1: Calculated reconstruction efficiencies for the baseline model compared to the statistical probabilities. Uncertainties are calculated assuming Poisson statistics.

In order to show that not only the Higgs boson is reconstructed correctly by the algorithm the reconstruction efficiencies for the hadronic and leptonic b-quark from the top quark decays is also investigated. Statistical probabilities for both partons are the same 16.67 %, while again the fitter gives much better values with 55.02 % and 53.65 % for the hadronic and leptonic b-quark, respectively.

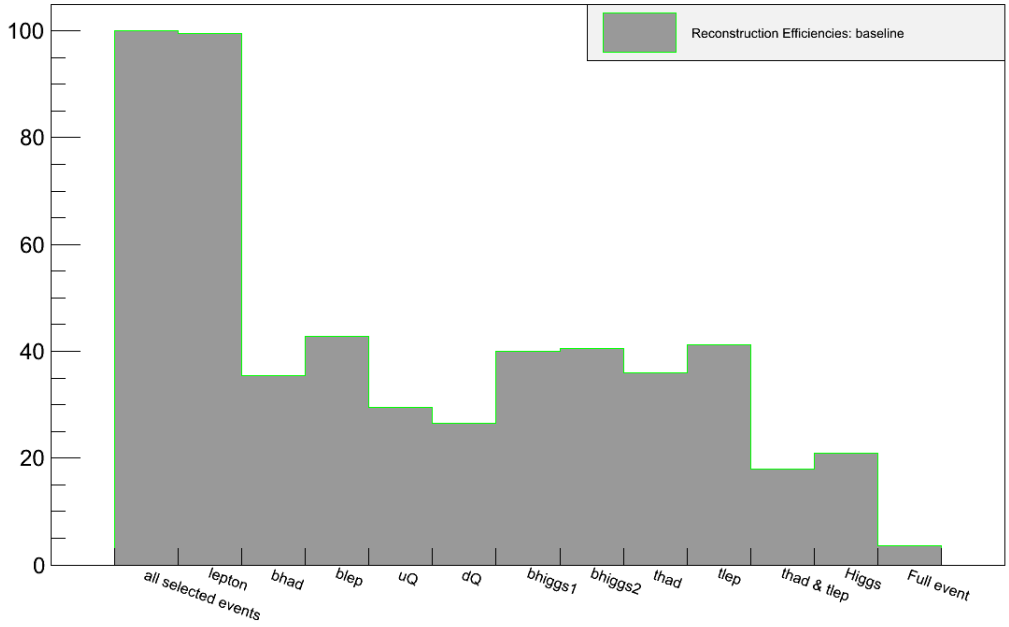


Figure 7.2: Reconstruction efficiencies of different decay end products of the $t\bar{t}H$ decay for the baseline model for all events in the MC sample.

These values were calculated only taking matched events into account, thus not

showing the reconstruction efficiencies that would be found in an actual event recorded by the ATLAS detector (since for this case of course no truth matching is possible). The total efficiency ϵ_{tot} is given by :

$$\epsilon_{tot} = \epsilon_M \cdot \epsilon_R, \tag{7.1}$$

with $\epsilon_M = 17.18$ % being the matching efficiency defined in section 5.4. This obviously yields a much smaller probability to reconstruct the entire event correctly, because in over 80 % of the events it is not possible to do so in the first place, since not all jets belonging to the $t\bar{t}H$ system are considered during the fit. The reconstruction efficiencies for all events rather than all matched events is shown in Figure 7.2. The theoretically calculated value of $\epsilon_{tot} \approx 3.1\%$ is reproduced and is shown in Figure 7.2. For unmatched events the reconstruction efficiency of the Higgs boson drops to 21.05 %.

7.1.2 Addition of Charge Information

Using the additional information of the jet-charge the analysis is carried out again, this time comparing the reconstruction efficiencies to the baseline model to see whether an improvement can be seen for the Higgs boson in particular. The reweighting of the individual permutations after the maximization of the likelihood fit is done as described in Section 4.5.2. The efficiencies are given in Table 7.2 compared to the baseline model and are also shown in Figure 7.3. Overall a slight increase can be seen throughout all channels, e.g. ϵ(Higgs) can be raised to 42.32 %.

	Efficiencies in %	
	baseline	charge information
$\epsilon(b_{had})$	55.02 ±1.37	56.56 ±1.38
$\epsilon(b_{lep})$	53.65 ±1.35	55.24 ±1.37
ϵ(Higgs)	41.37 ±1.19	42.32 ±1.20
ϵ_R	18.21 ±0.79	19.57 ±0.81

Table 7.2: Calculated reconstruction efficiencies for the baseline model compared model using charge information. Uncertainties are calculated assuming Poisson statistics.

The efficiencies for unmatched events show the same behavior. This seems rea-

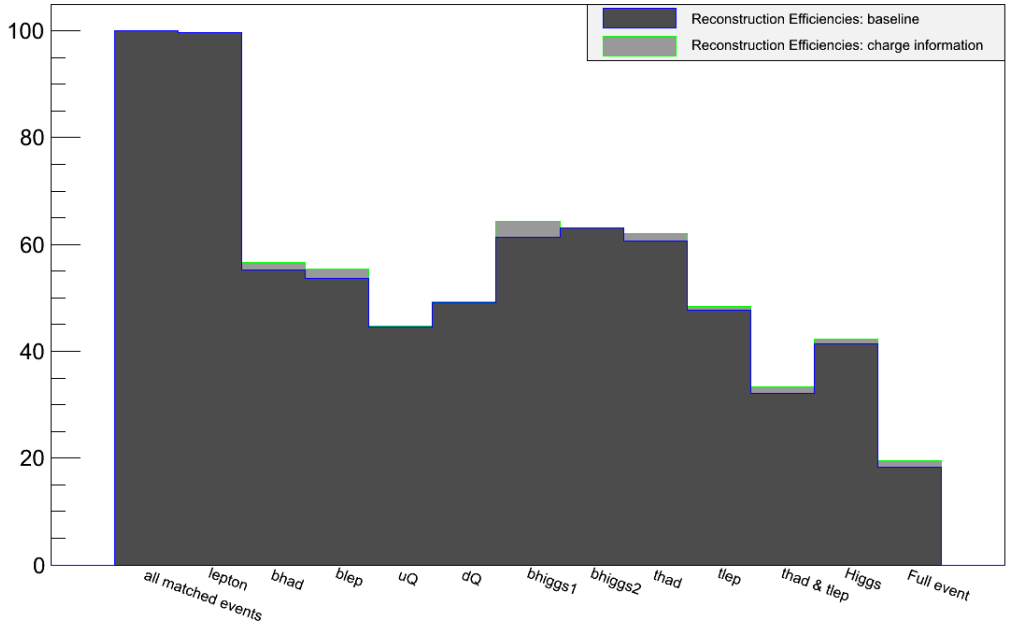


Figure 7.3: Reconstruction efficiencies of different decay end products of the $t\bar{t}H$ decay for the baseline model compared to the model containing charge information for matched events only.

sonable since the use of charge information should improve the reconstruction of the top quark pair depending on the quality of the charge determination. Assuming perfect charge information for b-jets or rather the jet inducing particles the b quarks from the $t\bar{t}$ system would never be interchanged thus eliminating a source of reconstruction mistakes. The charge information could further be exploited, e.g. by determining the sign of the charge of the b jets from the Higgs decay making sure their product fulfills the condition to be always negative.

7.1.3 Addition of Angular Information

Including a constraint on the angular distribution of the two b jets from the Higgs boson decay one finds that this approach does not improve the reconstruction at all. A comparison of the reconstructed angle between the two b-jets associated with the Higgs decay for the baseline model and the model using a constraint on the variable is shown in Figure 7.4. The distributions for the two different fitting setups show almost no difference, which is reasonable seeing that the reconstruction efficiency does not improve. The baseline model already gives a good prediction of the truth

case where the angle is always equal to π as expected for a two body decay in the rest frame of the parent particle. This might be due to the fact that the parameter σ_{angle} in equation 4.5 should have been chosen in a more strict way, but tightening this constraint could also lead to unwanted side effects that would suppress otherwise correctly reconstructed events only due to the fact that the jet direction can not a priori be assumed to be measured entirely accurate.

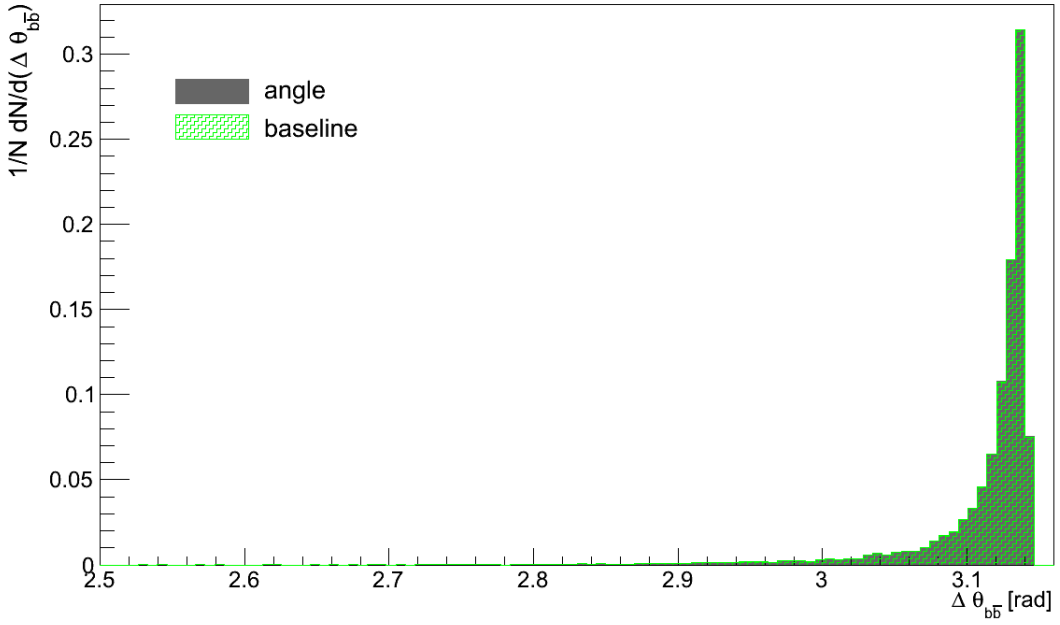


Figure 7.4: The angle between the b-quarks from the Higgs boson decay in the rest system of the Higgs particle for the baseline model and the model considering the angular information.

7.1.4 Addition of a Higgs Mass Constraint

Imposing a constraint on the Higgs mass should drastically improve the reconstruction since incorrectly assigned jets from other partons or initial and final state radiation jets are not likely to correctly form the Higgs mass. Again the reconstruction efficiencies for this setting are compared to the baseline model shown in Figure 7.7 and the values can be seen in Table 7.3.

As expected the efficiencies rise, especially for the Higgs reconstruction. Compared to the baseline model an improvement by a factor of ~ 1.5 from 41.37 % to 59.85

	Efficiencies in %	
	baseline	mass constraint
$\epsilon(b_{had})$	55.02 ±1.37	62.02 ±1.45.
$\epsilon(b_{lep})$	53.65 ±1.35	65.79 ±1.50
ϵ(Higgs)	41.37 ±1.19	59.85 ±1.42
ϵ_R	18.21 ±0.79	25.03 ±1.052

Table 7.3: Calculated reconstruction efficiencies for the baseline model compared to the model using a mass constrain with a fixed Higgs mass. Uncertainties are calculated assuming Poisson statistics.

% can be seen. Also a noticeable rise in $\epsilon(b_{had})$ and in particular $\epsilon(b_{lep})$ is visible, although of course not as much as for the Higgs boson. The leptonic b-quark efficiencies even surpass the hadronic one, which used to be higher for all KLFitter settings investigated before. This means that for the baseline model more jets from leptonic b-quarks must have been incorrectly assigned to b-jets from the Higgs decay, which is now suppressed by the strict mass constraint.

On the other hand the light quark assignment does not improve at all, which is consistent, due to the strong re-weighting of incorrect permutations where light jets are assigned to b jets. The probability to assign a light quark to a quark coming from the Higgs decay is relatively low to begin with thus leaving almost no room for improvement imposing a constraint only depending on the Higgs particle.

7.1.5 Combined Model

At last all options discussed so far are combined to see whether the individual positive effects prevail or if the individual components counteract each other. The results are compared to the baseline model and the fixed Higgs mass model discussed in the previous section, Table 7.4 shows the respective yields for the reconstruction efficiency again for matched events only.

As can be seen the combination results in the highest efficiencies for all KLFitter settings showing a final probability of 26.42 % to correctly assign all partons to their respective jets. The value of ϵ(Higgs) peaks at 60.73 % compared to the 41.37 % of the baseline model.

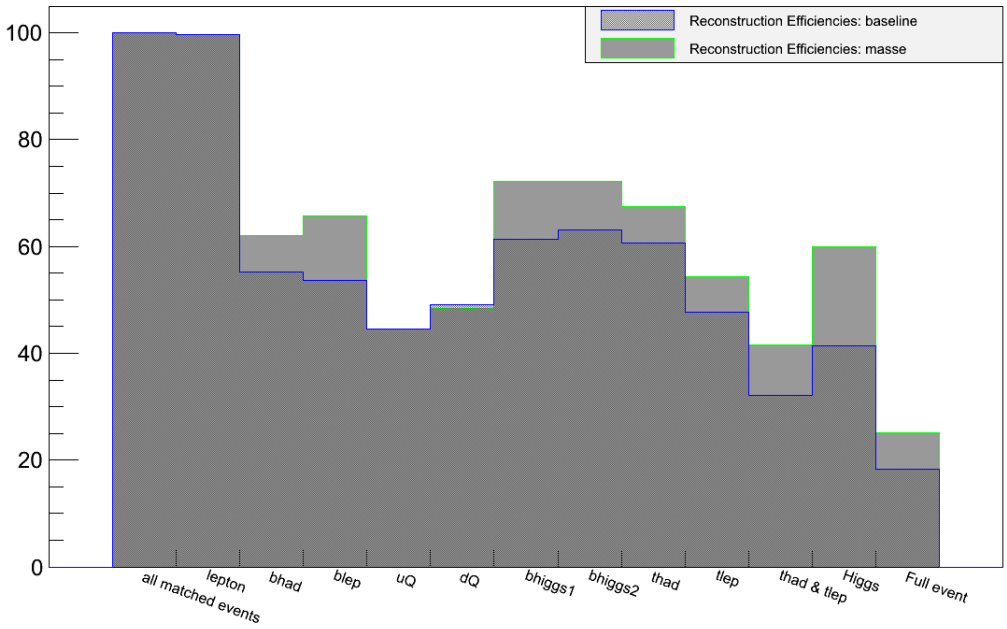


Figure 7.5: Reconstruction efficiencies of different decay end products of the $t\bar{t}H$ decay for the baseline model compared to the model with fixed Higgs mass for matched events only.

	Baseline Efficiencies in %		Combined Model Efficiencies in %	
	matched events	unmatched events	matched events	unmatched events
$\epsilon(b_{had})$	55.02 ±1.37	35.43 ±0.45	63.00 ±1.43	36.39 ±0.49
$\epsilon(b_{lep})$	53.65 ±1.35	42.78 ±0.50	67.18 ±1.51	46.15 ±0.55
ϵ(Higgs)	41.37 ±1.19	21.05 ±0.35	60.73 ±1.44	34.31 ±0.48
ϵ_R	18.21 ±0.79	3.13 ±0.14	26.42 ±1.055	5.18 ±0.19

Table 7.4: Calculated reconstruction efficiencies for the baseline model compared to the combined model. Uncertainties are calculated assuming Poisson statistics.

As already explained the reconstruction efficiencies for not matched events are what actually matters in an analysis of real data. The calculated values are shown in Table 7.4. The same tendency that could be seen for matched events is visible for complete MC sample. The Higgs reconstruction efficiency reaches a maximum value of 34.31 % in comparison to only 21.05 % for the baseline model.

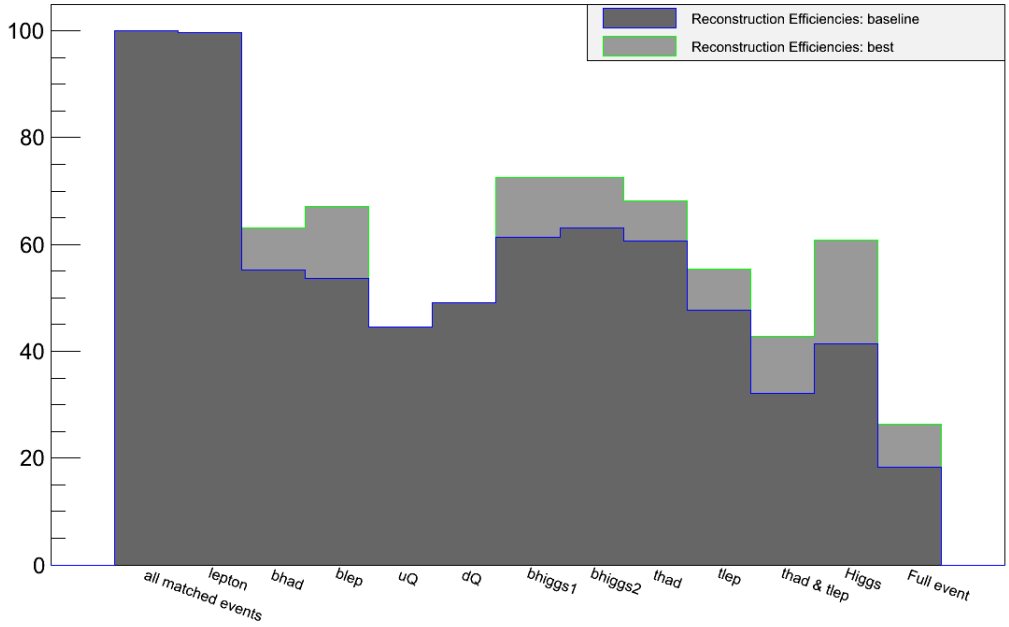


Figure 7.6: Reconstruction efficiencies of different decay end products of the $t\bar{t}H$ decay for the baseline model compared to the combined model for matched events only.

Dependency on the Transverse Momentum

Assuming matched events the reconstruction efficiencies are further studied as a function of the transverse momentum of the reconstructed Higgs boson. For both the baseline (Fig. 7.8 (a)) and combined model (Fig. 7.8 (b)) the reconstruction efficiencies are plotted over the transverse momentum of the reconstructed Higgs boson. For the baseline model the efficiencies are rather low at low transverse momentum with values around 35 %. From 150 GeV upwards the efficiencies start to rise yielding high values of 80 % for high p_T Higgs bosons. A similar behaviour can be seen in the plot for the combined model but with overall higher reconstruction values to begin with. For $p_{T,Higgs} > 400$ GeV the statistics are not sufficient enough to draw any conclusions.

Assuming matched events the sum of the combined transverse momentum of the jets should be zero in total, thus a high transverse momentum Higgs boson implies that the top pair system should also carry a relatively high p_T. This leads to b jets from top decays that will mostly face away from the Higgs boson b quarks making it

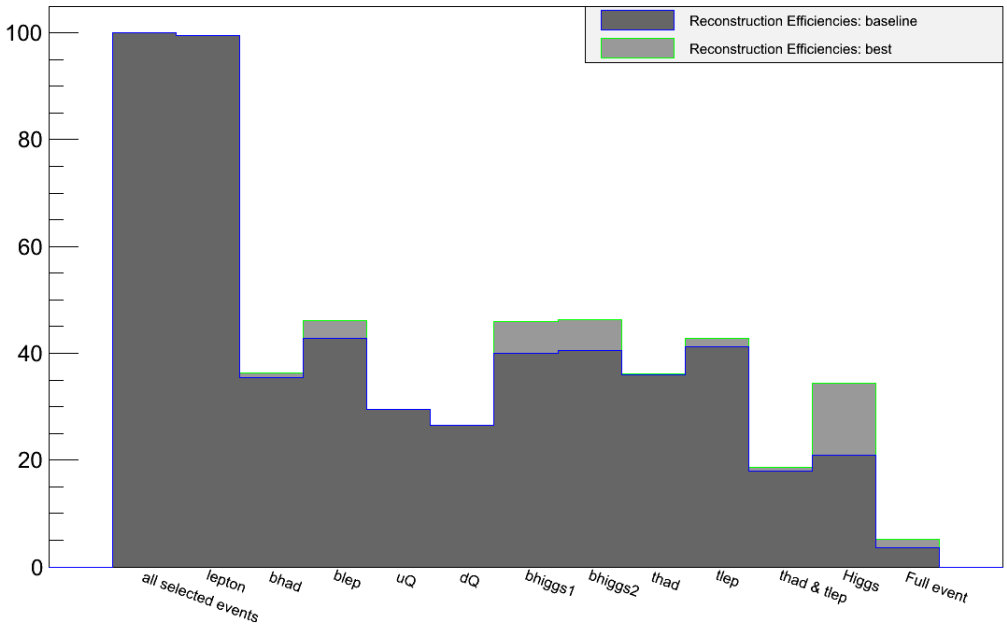
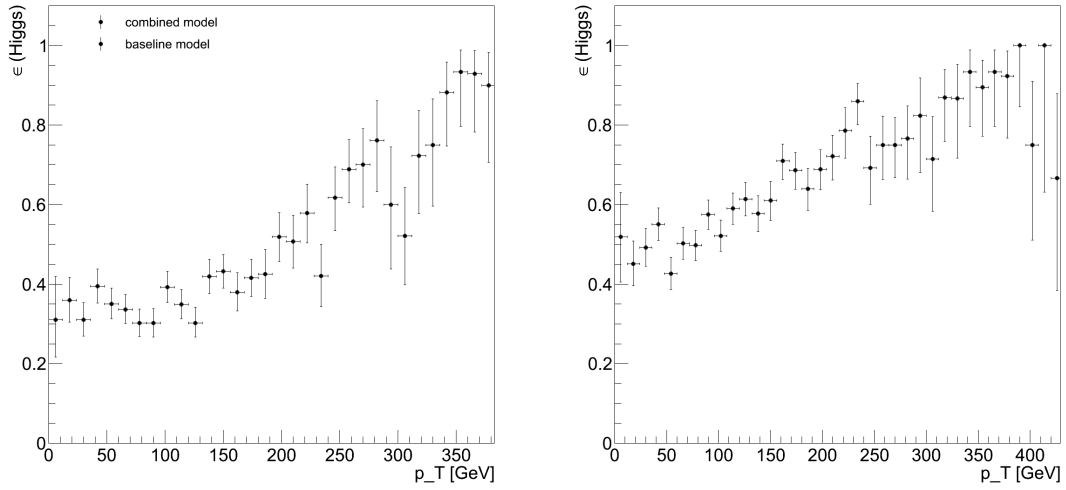


Figure 7.7: Reconstruction efficiencies of different decay end products of the $t\bar{t}H$ decay for the baseline model compared to the model containing charge information for all events in the MC sample.

more difficult to incorrectly assign the wrong b jets to the Higgs decay. Imposing the strict mass constraint makes it even less likely for a wrong assignment to happen.



(a) baseline model

(b) combined model

Figure 7.8: Reconstruction efficiency ϵ(Higgs) to correctly reconstruct the Higgs boson for matched events as a function of the transverse momentum p_T of the reconstructed Higgs particle for the combined model (right) and the baseline model (left).

7.2 Kinematic Distributions

In order to evaluate the performance of the kinematic fit a number of particle properties are compared to the truth distributions in order to verify that KLFitter indeed improves the estimation of those parameters, e.g. the energy. The focus is placed on the Higgs properties. For all considered distributions the truth values from the Monte Carlo sample are compared to the reconstructed values of the baseline and the combined model.

Higgs Boson direction

A good variable for evaluating the performance of the reconstruction of the Higgs boson is the distance of the reconstructed particle compared to the true direction expressed by ΔR. This is especially interesting if the different KLFitter settings are compared. First all events are considered meaning no matching is required. Figure 7.9 shows the normalized ΔR distribution of the Higgs boson for the baseline and combined model. To allow an easier comparison, the distributions show both the matched and unmatched fraction of the total events in different colours. Also a

logarithmic scale between -3 and 1 is chosen to allow higher resolution for low ΔR values. This range corresponds to $0.001 < \Delta R < 10$. Both cases feature a shoulder in the distribution at values around -1 ($\Delta R = 0.1$), which is much more pronounced for the reconstruction using the more sophisticated setup. This shoulder represents the correctly reconstructed fraction of Higgs bosons. For matched events both distributions feature a second, less pronounced peak at values around 0.5 ($\Delta R \approx \pi$) meaning that the Higgs boson is reconstructed in the opposite direction compared to the true one. This might be explained with one of the b quarks of the Higgs decay to be wrongly assigned to a top quark decay. This most likely will lead to a b jet from a top decay to be associated with the Higgs boson and maybe even inverting its reconstructed direction. For unmatched events the shoulder around -1 is almost not visible for the baseline model, whereas for combined model is still vaguely perceptible. Overall the distributions are dominated by combinatorial background arising from the selection of initial and final state radiation jet. When evaluating this plots one should keep in mind that due to the logarithmic scale the ΔR bins are not equidistant.

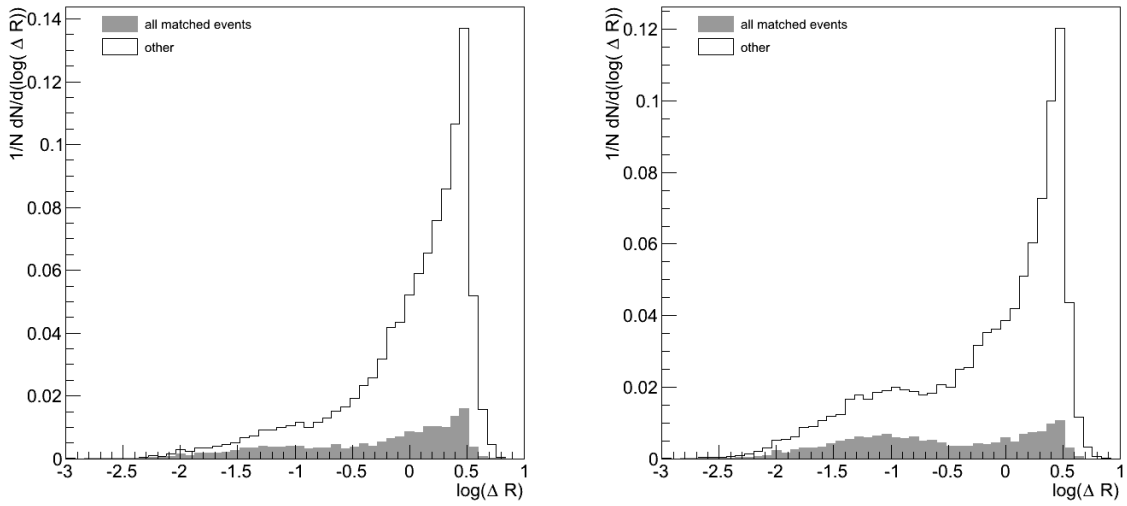


Figure 7.9: The ΔR distance of the reconstructed Higgs boson to the true ones for all events are shown in a stacked plot for the combined model (right) and the baseline model (left). It is distinguished between events that are matched (grey shaded) and events that are not matched. The x-axis is on a logarithmic scale, thus the bin size is not equidistant in ΔR.

The fraction of matched events is further assessed in Figure 7.10. Again the distribution is divided in two subcomponents. The Grey shaded area corresponds to the events where the two b jets from the Higgs decay could be successfully matched to their truth counterpart. As described before a clear peak around -1 is visible consistent with a great fraction of the events where reconstructed particles have almost no angular displacement to the truth particles. The remaining part again arises from incorrectly reconstructed events where the Higgs boson is formed from a jet belonging to the top quark decay.

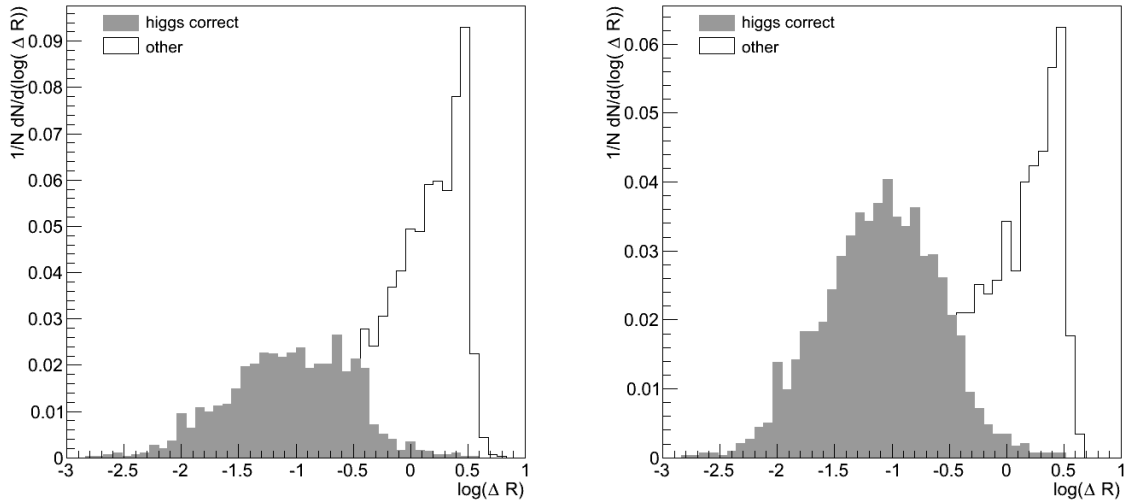


Figure 7.10: The ΔR distance of the reconstructed Higgs boson to the true ones for matched events are shown in a stacked plot for the combined model (right) and the baseline model (left). It is distinguished between events in which the b jets from the Higgs boson decay are correctly assigned (grey shaded) and vice versa. The x-axis is on a logarithmic scale, thus the bin size is not equidistant in ΔR.

Finally events where jets could be matched to the b quarks from the Higgs decay are considered. These events interestingly show a similar behavior, with slightly less distinct peaks (Fig. 7.11). This means that if a better selection of the events used for reconstruction with the KLFitter could be found very high efficiencies at least for the Higgs boson reconstruction can be achieved.

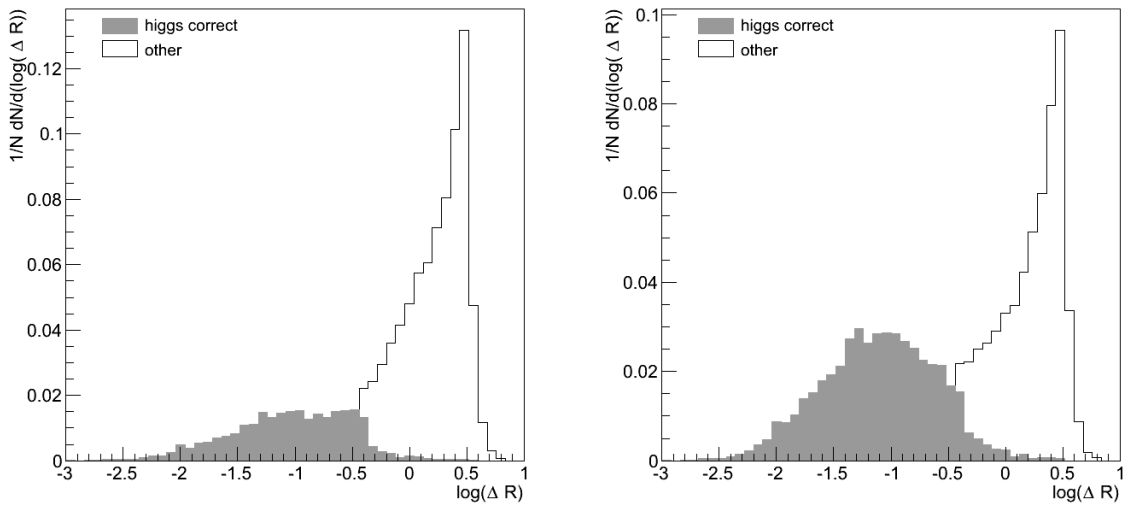


Figure 7.11: The ΔR distance of the reconstructed Higgs boson to the true ones for events where the two b jets from the Higgs decay are considered in the fit are shown in a stacked plot for the combined model (right) and the baseline model (left). It is distinguished between events in which the b jets from the Higgs boson decay are correctly assigned (grey shaded) and vice versa. The x-axis is on a logarithmic scale, thus the bin size is not equidistant in ΔR.

Higgs Energy

The energy of the Higgs particle is certainly an interesting quantity to investigate depending on the different reconstruction setups. The energy of the jets of each individual event are varied in order to find the best permutation of the event, but after the fitting is done and a permutation is chosen it is still possible to maintain the assignment of jets to partons. Rather than looking at the fitted energy values the energy of the particles as measured by the detector are used for comparison with the truth values. No matching is required. The distribution of the energy of the Higgs particle for the baseline and combined model together with the true energy is shown in Figure 7.12. Also the ratio of individual bin contents of the truth and the two reconstructed distributions are calculated.

The true distribution has a mean value of 256.2 GeV compared to 273.9 GeV and 321.0 GeV for the combined and baseline model respectively implying that not only the reconstruction efficiencies rise for the combined model, but also the correspond-

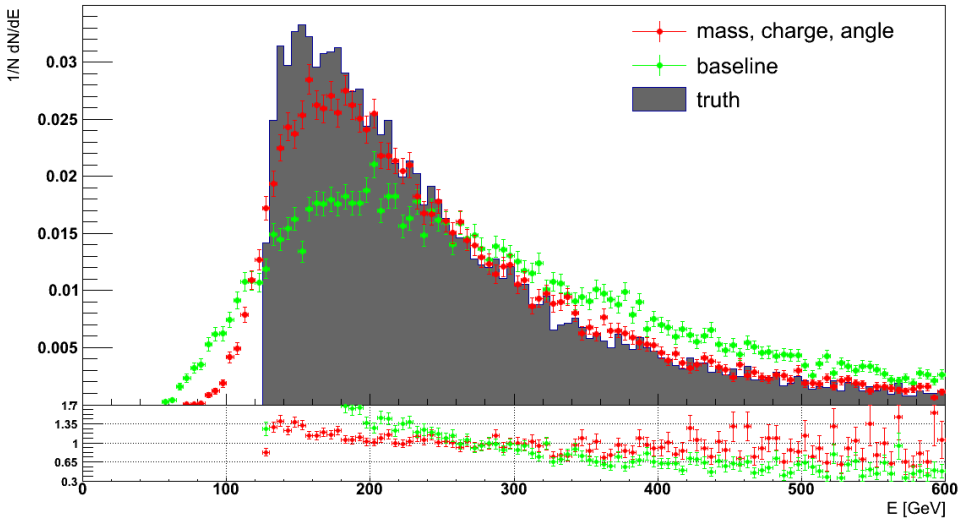


Figure 7.12: The reconstructed energy of the Higgs boson for the baseline(green) and combined(red) model is compared to the true (gray shaded) distribution. The ratio of reconstructed and true values is given in a subdivision of the plot.

ing energy distribution for the Higgs particle show a much better agreement to the true values. Not only the mean but also the rms value drastically improves from 182.2 GeV to 137.9 GeV compared to a truth value of 129.2 GeV. This further consolidates the impression of a more accurate reconstruction of the Higgs boson using a more sophisticated likelihood. Also the shape of the combined model is much closer to the true distribution. This can be extracted from the ratio plot that mostly stays inside a band of maximum ±30 % for the combined model below energies of 500 GeV, whereas the baseline model grossly underestimates the values up to 250 GeV and overestimating events that have a high reconstructed Higgs boson energy. This might be due to wrong assignment of higher energy b-jets from top quark decays or initial and final state radiation. For high energy values the statistics are suffering thus not allowing a reasonable evaluation in regions with $E_{Higgs} > 500$ GeV.

Higgs transverse momentum

The transverse momentum (Fig. 7.13) is reconstructed in good agreement to the truth distributions for both reconstruction setups. For values up to 100 GeV, both

KLFitter settings underestimate the truth yield. A good agreement is visible in the range between 100-300 GeV, whereas the contribution for higher p_T values is overestimated. Again for high transverse momentum the sample lacks statistics. This is also visible in the mean of the distribution, which is slightly higher for the reconstructed distributions with almost identical values of about 134 GeV compared to the truth mean of 127.8 GeV.

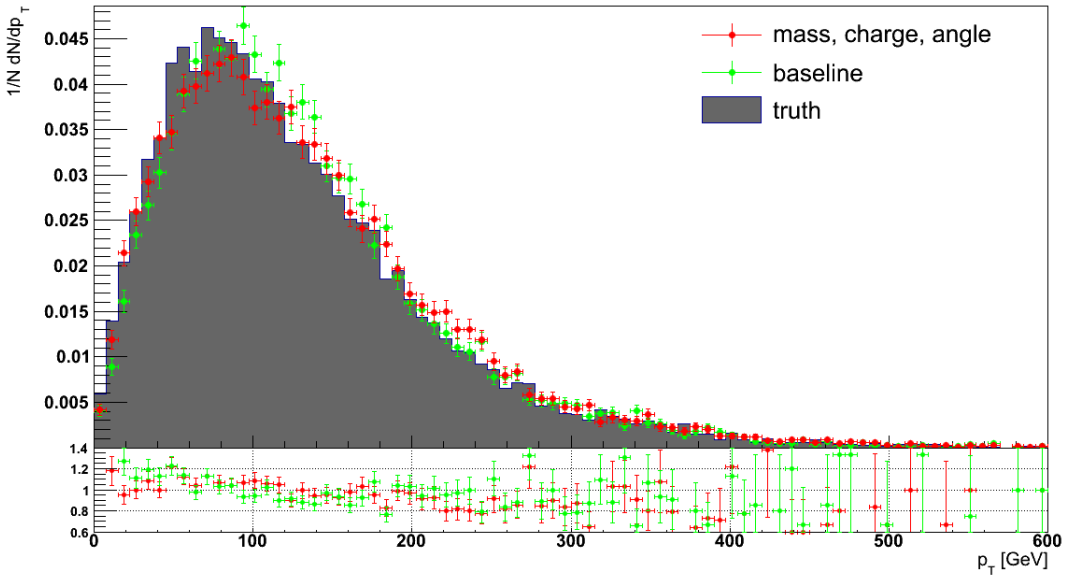


Figure 7.13: The reconstructed transverse momentum of the Higgs boson for the baseline(green) and combined(red) model is compared to the true (gray shaded) distribution. The ratio of reconstructed and true values is given in a subdivision of the plot.

Higgs Mass

Another important variable for the reconstruction is the mass of the invariant combination of the two jets assigned to the Higgs decay products. This is the main discriminant used when comparing background to signal events since a pronounced peak should be visible in the signal region while the background events should basically be distributed evenly. The true distribution corresponds to a very sharp peak around $m_H = 125$ GeV with a very narrow width. This is induced by the simulation process, where the mass and decay width are fixed according to Section 5.1. Again

the baseline and combined model are compared (Fig. 7.14). Both distributions peak around 125 GeV as expected. The combined model shows a much sharper, more localized peak, when compared to the baseline distribution, which is rather flat. Both distributions are asymmetric with a tail going to higher mass values that is most likely due to reconstruction of particles with too high energy that was already visible evaluating the respective distributions.

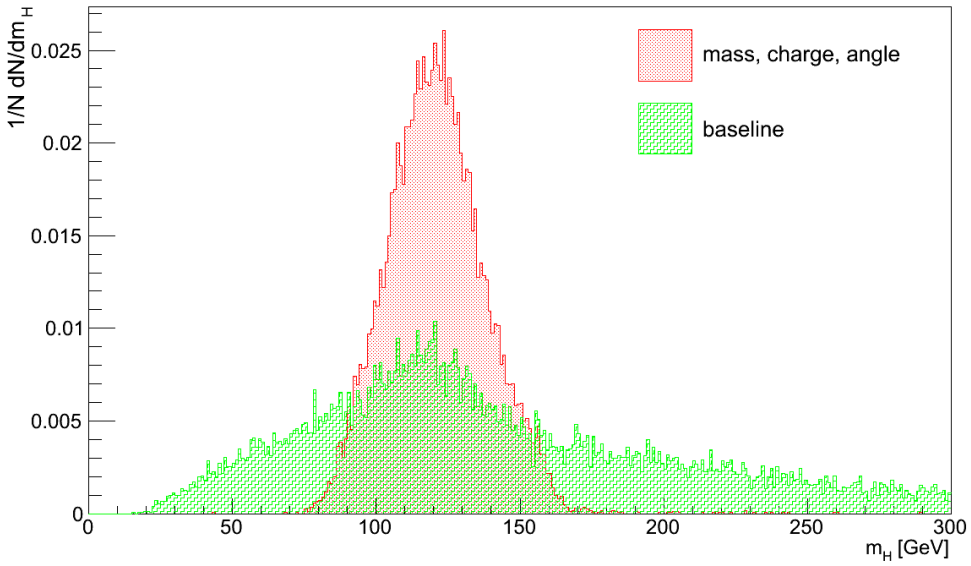


Figure 7.14: The reconstructed Higgs mass m_H for the baseline (green) and combined (red) model.

Higgs Angular Variables

For the sake of completeness both angular variables are shown as well. Again no matching is required thus all selected events are evaluated. The η distributions in Figure 7.15 show the truth and combined model outcome. In the central region between $|\eta| < 2$, reconstructed values are in good agreement with the truth values. In the central region relative fluctuations of ± 15 % are achieved. Going to higher values of η, the statistics do not allow meaningful evaluation of the distributions.

The ϕ distribution of the combined model is compared to the truth values in Figure 7.16. Again both distributions can be matched reasonably well. Over the entire ϕ-range the reconstructed values deviates between ± 15 % to the true distribution.

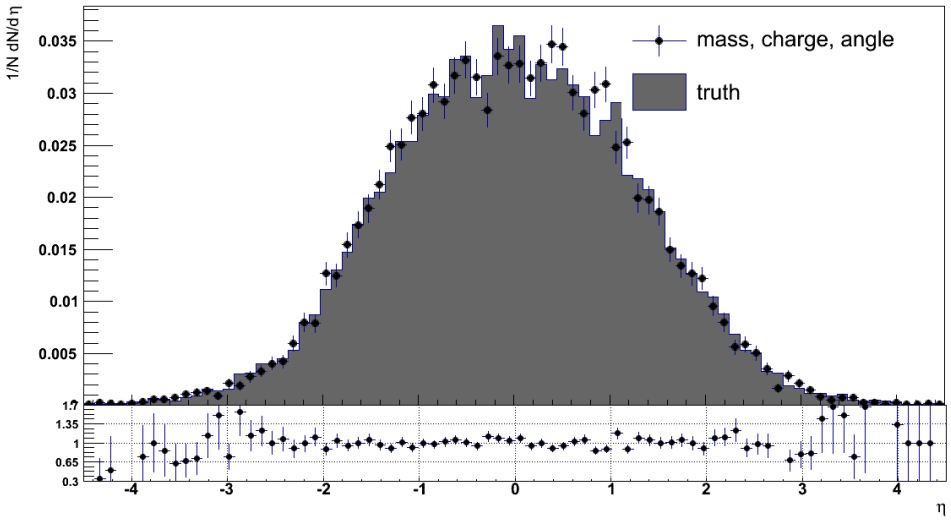


Figure 7.15: The reconstructed Higgs η for the combined model is compared to the true (gray shaded) distribution.

The Higgs boson is evenly distributed in ϕ.

7.2.1 b-Tag Dependency

The study presented in Chapter 6 considers different topologies depending on jet and b-tag multiplicity to extract a possible $t\bar{t}H$ signal. It is thus reasonable to determine the reconstruction efficiencies for different b-tag multiplicities. So far, at least one b-tagged jet was required by the KLFitter, but the output can be split into different subcategories with different numbers of b-tags. Again evaluating both, matched and all events the number of events depending on the b-tag multiplicity can be seen in Table 7.5.

	$N_{tag} = 1$	$N_{tag} = 2$	$N_{tag} = 3$	$N_{tag} \geq 4$
matched events	(1)	(7)	1585	1356
all events	65	1349	9341	4295

Table 7.5: Number of events with different b-tag multiplicity for matched events and events with no matching required. The Values in brackets are not considered due to low statistics.

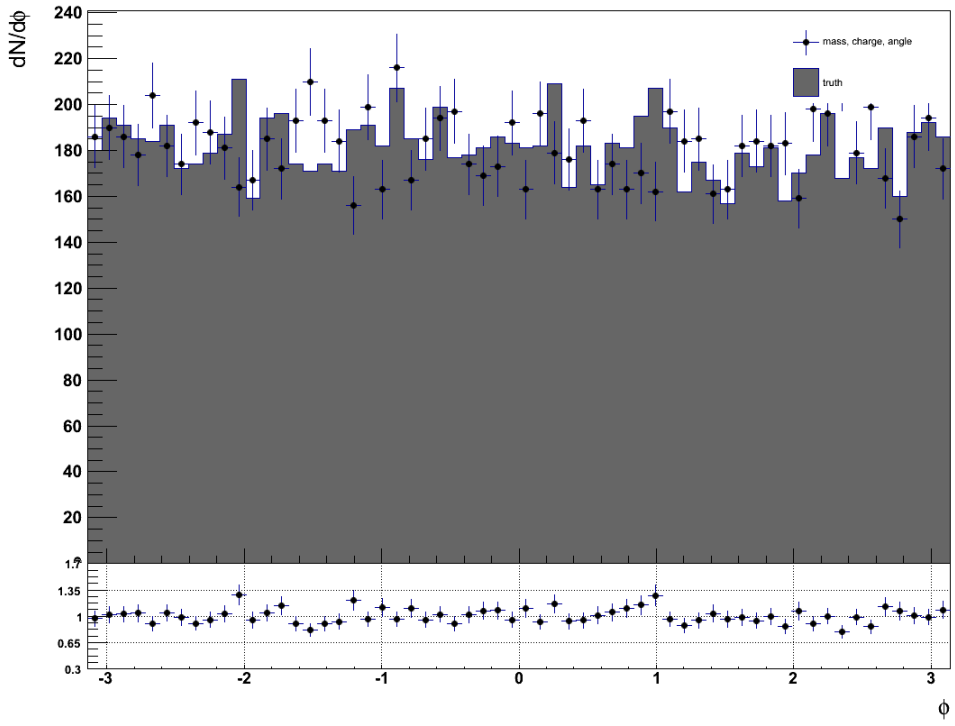


Figure 7.16: The reconstructed Higgs ϕ for the combined model is compared to the true (gray shaded) distribution.

These numbers alone already allow the conclusion that for an event to be successfully matched at least three b-tagged jets should be present. Since the count of matched events in the one and two b-tagged topology is very low, no efficiencies are derived in these cases.

	$N_{tag} = 1$		$N_{tag} = 2$		$N_{tag} = 3$		$N_{tag} \geq 4$	
	ϵ(Higgs)	ϵ_R	ϵ(Higgs)	ϵ_R	ϵ(Higgs)	ϵ_R	ϵ(Higgs)	ϵ_R
no match combined	9.23	0	18.16	0.15	32.65	4.25	43.38	8.85
no match baseline	4.62	0	10.82	0.07	20.16	3.01	25.96	5.94
match combined	—	—	—	—	58.93	25.05	62.83	28.02
match baseline	—	—	—	—	39.94	17.73	42.92	18.81

Table 7.6: Reconstruction efficiencies in percent for topologies with different b-tag multiplicity.

Looking at the reconstruction efficiencies in the case of no matching required the values are about twice as high going from one to two b-tags that hits a saturation at more than four b-tagged jets (Fig 7.17), which is consistent with the signature of the $t\bar{t}H$ decay that should contain exactly four jets from b-quark hadronization. A further increase in bottom quark jets would probably even lead to a decrease of the efficiencies due to one of the light jets of the hadronic W decay not being detected.

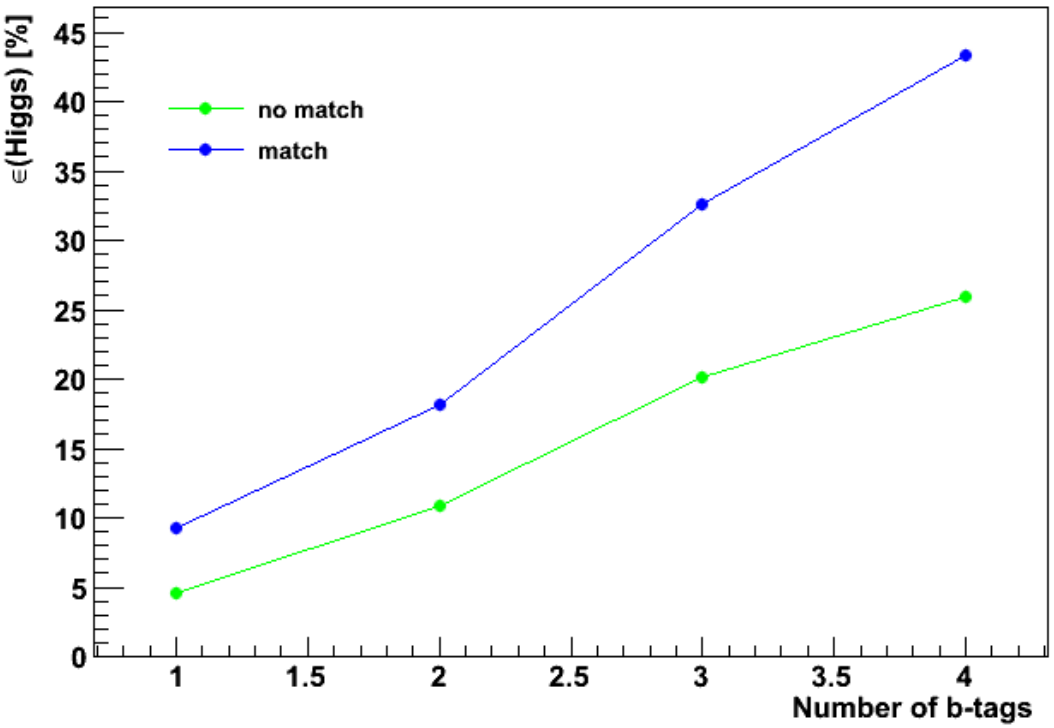


Figure 7.17: Higgs reconstruction efficiencies for topologies with different b-tag multiplicity for matched and unmatched events.

8 Conclusion and Outlook

In this thesis a kinematic fitting tool, the KLFitter package, was presented and used to reconstruct ATLAS Monte Carlo data to study the reconstruction of $t\bar{t}H(H \to b\bar{b})$ events in the semileptonic decay channel. The KLFitter uses a maximum likelihood method to reconstruct events produced in proton-proton collisions focusing on both, the correct assignment of detected jets to particles of certain decay signatures, but also the energy resolution of jets arising through hadronization of particles from the hard scattering process. A basic model used for $t\bar{t}H(H \to b\bar{b})$ searches at ATLAS was presented and various modifications have been tested to improve the reconstruction efficiencies with a strong focus on the reconstruction of the Higgs boson. The data sample is studied considering both all events in the sample and only those events that fulfill a simple truth-matching criterion in η-ϕ-space. In total, five different settings for the KLFitter are evaluated leading to an improvement of the reconstruction of the Higgs boson from 41.4 % to 60.7 % and 21.1 % to 34.3 % for truth matched events and all events, respectively. The total efficiencies could be raised from 18.2 % to 25.0 % for matched events compared to only 0.5 % if the particles are assigned completely random. Furthermore the kinematic properties of the reconstructed Higgs boson have been studied and a drastic improvement especially on the mass distribution could be found when going from the baseline model to the best setting. A distinctive peak in the simulation Higgs mass $m_H = 125$ GeV could be found.

Additionally, it was shown that the reconstruction efficiency of the Higgs boson is strongly depended on its transverse momentum p_T. For higher values of p_T the assignment of the jets to model particles improves drastically. Also the reconstruction of events which are required to feature the Higgs b-quarks during the kinematic fit have been investigated. Results show that for the best setting about 50 % of the Higgs particles could be successfully reconstructed, implying that both b-quarks from the Higgs decay are used during the kinematic fit, the probability is high that at least the Higgs boson will be reconstructed correctly regardless of the remaining

jets.
Finally the dependence of the reconstruction efficiencies relating to the number of b-tagged jets in the event have been studied. With increasing b-jet multiplicity the efficiencies improve showing almost exponential behaviour reaching a maximum at four or more b-tags. This is consistent with the $t\bar{t}H(H \to b\bar{b})$ decay signature containing exactly four bottom quarks.

In the future it would be interesting to examine the performance of the modified likelihood on a $t\bar{t}$ sample to cross check that the improvement is indeed limited to actual $t\bar{t}H$ signature and not biased to reconstruct Higgs particles that are not contained in the event. Also the extended likelihood can be used for data analysis of the data taken during 2013 with the ATLAS detector to derive tighter better limits on the observed cross section of $t\bar{t}H$ production. There is also the possibility to use the kinematic distributions reconstructed by the KLFitter to train Artificial Neural Networks to distinguish between signal and background events.

Bibliography

[1] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Rev. Lett.*, B716, 2012.

[2] CMS Collaboration. Observation of a new boson at a mass of 125 Gev with the CMS experiment at the LHC. *Phys. Rev. Lett.*, B716, 2012.

[3] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, 22:579, 1961.

[4] A. Salam and J. C.Ward. Electromagnetic andWeak Interactions. *Phys. Lett.*, 13:168, 1964.

[5] S. Weinberg. A Model of Leptons. *Phys. Rev. Lett*, 19:1264, 1967.

[6] H. D. Politzer. Reliable Perturbative Results for Strong Interactions. *Phys. Rev. Lett.*, 30:1346, 1973.

[7] D. J. Gross and F. Wilczek. Asymptotically Free Gauge Theories. I. *Phys. Rev. D*, 8:3633, 1973.

[8] H. D. Politzer. Asymptotic freedom: An approach to strong interactions. *Phys. Rev. Lett.*, 145:1156, 1966.

[9] David Griffith. *Introduction to Elementary Particles*. Wiley-VCH Verlag, second, revised edition edition, 2008.

[10] H. Weyl. *The Classical Groups: Their Invariants and Representations*. Westview Press, 1995.

[11] P. W. Higgs. Broken Symmetries, Massless Particles and Gauge Fields. *Phys. Rev. Lett.*, 12:132, 1964.

[12] P. W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508, 1964.

[13] P. W. Higgs. Spontaneous Symmetry Breakdown without Massless Bosons. *Phys. Rev. Lett.*, 145:1156, 1966.

[14] G. Arnison et al. Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540GeV$. *Phys. Lett. B*, 122: 103, 1983.

[15] G. Arnison et al. Experimental observation of lepton pairs of invariant mass around 95 GeV/c2 at the CERN SPS collider. *Phys. Lett. B*, 7126:398, 1983.

[16] M. Kobayashi and T. Maskawa. CP-Violation in the Renormalizable Theory of Weak Interaction. *Prog. Theor. Phys.*, 49:652, 1973.

[17] F. Zwicky. Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6:110, 1933.

[18] M. S. Turner and D. Huterer. Cosmic Acceleration, Dark Energy, and Fundamental Physics. *Journal of the Physical Society of Japan*, 76:111015, 2007.

[19] ATLAS Collaboration. Combination of ATLAS and CMS results on the mass of the top-quark using up to 4.9 fb^{-1} of $\sqrt{s} = 7$ TeV LHC data, 2013. ATLAS-CONF-2013-102.

[20] The CDF Collaboration, T. Aaltonen et al. Observation of Top Quark Production in $p\bar{p}$ Collisions. *Phys. Rev. Lett.*, 74:2626, 1995.

[21] The DØ Collaboration, S. Abachi et al. Observation of the Top Quark. *Phys. Rev. Lett.*, 74:2632, 1995.

[22] A. Quadt. Top Quark Physics At Hadron Colliders. *Eur. Phys. J. C*, 48:835, 2006.

[23] ATLAS Collaboration. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/HiggsTheoryPlots .

[24] CMS Collaboration. Search for Higgs boson production in association with a top quark pair in pp collisions. 2012. CMS PAS HIG-12-025.

[25] L. Evans, P. Bryant et al. LHC Machine. *JINST*, 3:S08001, 2008.

[26] LEP Design Report: Vol. 2. The LEP Main Ring. 1984. CERN-LEP-84-01.

[27] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider, 2008.

[28] CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3:S08004, 2008.

[29] LHCb Collaboration. The LHCb Detector at the LHC. *Journal of Instrumentation*, 3:S08005, 2008.

[30] ALICE Collaboration. The ALICE experiment at the CERN LHC. *Journal of Instrumentation*, 3:S08002, 2008.

[31] CERN Communication Group. CERN faq - LHC the guide, 2009. CERN-Brochure-2009-003-Eng.

[32] J. Beringer et al. Review of particle physics. *Phys. Rev. D*, 86:010001, 2012.

[33] A. Caldwell, D. Kollar and K. Kroeninger. BAT - The Bayesian Analysis Toolkit. arXiv:0808.2552v1.

[34] ATLAS Collaboration. Measurement of the top quark charge in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *JHEP*, 1311:031, 2013.

[35] ATLAS Collaboration. Search for a Higgs boson produced in association with a top-quark pair and decaying to $b\bar{b}$ in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector, 2012. ATLAS-CONF-2012-135 .

[36] ATLAS Collaboration. ATLAS Simulation Infrastructure. *Eur. Phys. J. C*, 70: 823, 2010.

[37] The GEANT4 Collaboration. GEANT4: A simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250, 2003.

[38] M. Cacciari, G. P. Salam and G. Soyez. The anti-k_t Jet Clustering Algorithm. *JHEP*, 0804:063, 2008.

[39] ATLAS Collaboration. Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data. 2011. ATLAS-CONF-2011-102.

Acknowledgements

First of all I would like to express my gratitude to Prof. Arnulf Quadt for the opportunity to draw up my master thesis at his department and for taking the time for being the first referee. My time at the second institute allowed me to get an insight in a scientific work-environment, which i am certainly very grateful for. I would also like to thank Prof. Ariane Frey for consenting to act as the second referee for this thesis.

Special thanks goes to Dr. Kevin Kröninger who supervised my work and provided me with inspiration for my studies. I am also thankful for his help during the write up of this thesis making time for proof reading on a very short time schedule. Furthermore I would like to thank Boris Lemmer for the technical insight he shared with me about the KLFitter and other issues arising during my thesis.

I would also like to thank my roommate and friend Ulf Stolzenberg, who not only shared a flat with me in Göttingen but also accompanied me during my entire time at the university sharing the difficult and the cheerful times throughout the years.

Lastly I am grateful for the support of my family that always supported me in every regard possible and especially for enabling me to take up my physic studies with their financial support.

Erklärung nach §18(8) der Prüfungsordnung für den Bachelor-Studiengang Physik und den Master-Studiengang Physik an der Universität Göttingen:

Hiermit erklare ich, dass ich diese Abschlussarbeit selbständig verfasst habe, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht habe.
Darüberhinaus erkläre ich, dass diese Abschlussarbeit nicht, auch nicht auszugsweise, im Rahmen einer nichtbestandenen Prüfung an dieser oder einer anderen Hochschule eingereicht wurde.

Göttingen, der 29. November 2013

(Veit Peter Dahlke)