Messung der Jetmassenverteilung in kollimierten Top-Quark Zerfällen

Measurement of the Jet Mass Distribution in Boosted Top Quark Decays

von
Dennis Schwarz
geboren am
28.07.1991

Masterarbeit im Studiengang Physik
Universität Hamburg

2017

1. Gutachter: Prof. Dr. Johannes Haller
2. Gutachter: Prof. Dr. Peter Schleper

Zusammenfassung

Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muss keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie „Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln.

Abstract

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Contents

Chapter 1

Introduction

Chapter 2

Theory

2.1 Standard model of particle physics

The standard models of particles is a quantum field theory describing elementary particles and their interaction. All particles contained in this theory have been discovered and experiments have confirmed predictions of the standard model at very high precision. In Fig. 2.1 all elementary particles and their basic properties are displayed. The particles can be ordered in groups by their properties. Firstly one distinguishes particles depending on their spin. Spin-$\frac{1}{2}$ particles are called fermions, Spin-1 and Spin-0 particles represent the bosons. Fermions are the building block of matter while bosons carry the fundamental forces included in the standard model. Quarks and leptons are sub groups of the fermions are divided by their possible interaction with forces. While the quarks are affected by the strong force, carried by the gluon, the leptons are not. Three fundamental forces are introduced in this theory. The strong force is carried by the gluon and affects quarks and the gluon itself. The charge of the strong interaction is called colour.

2.1.1 Top Quark

The top quark is an up type quark from the third generation and carries electromagnetic charge of $Q = +\frac{2}{3}e$. With this mass around 173 GeV the top quark is the heaviest particle in the standard model, therefore offers a large phase space for decays and thus has a life time of approximately 0.5×10^{-24} s [2]. Because of the short life time the top quark does not form bound hadronic states and thus measurement of the bare quark are possible. This provides a special access to parameters of the standard model. Especially the top quark mass is an essential parameter to check the standard model for consistency. The Higgs mechanism relates

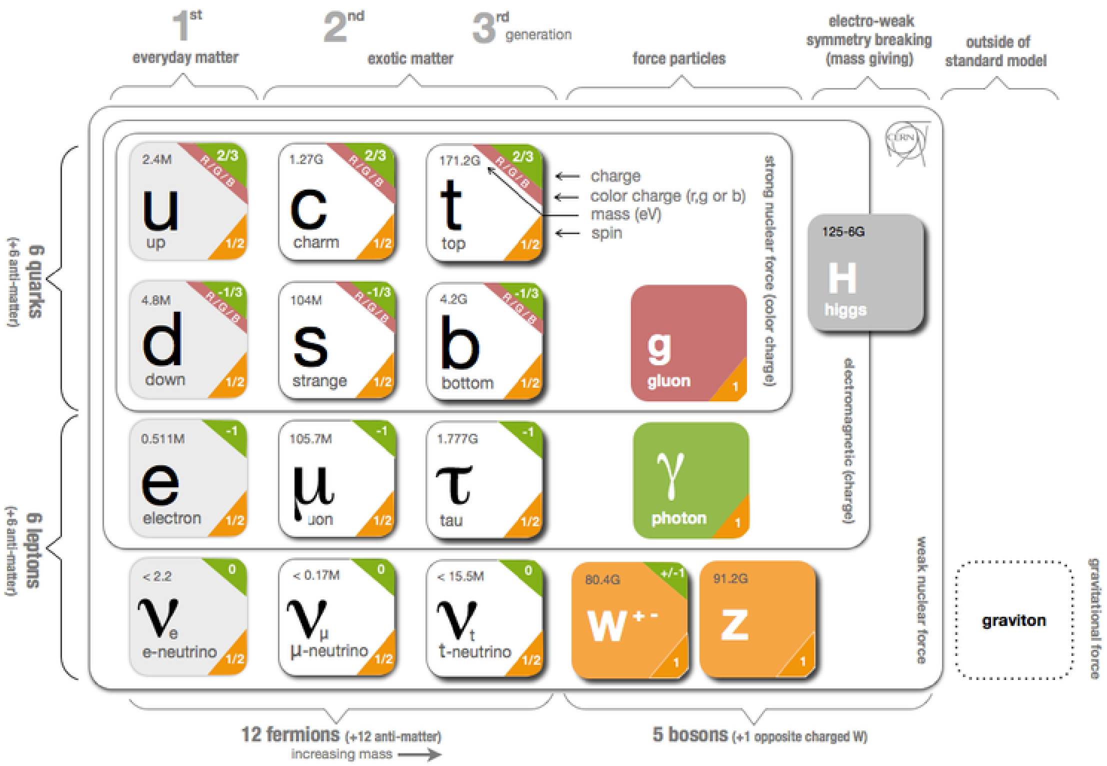

Figure 2.1: Particle content of the standard model [1]

the masses of the top quark with the masses of the W and Higgs boson. Measuring these masses with high precision gives the possibility to check the standard model for consistency. Additionally the top quark is important for searches of new physics since is is often part of the final state and/or a relevant background.

In hadron colliders, the production of a $t\bar{t}$ pair happens via $g\bar{q}$ annihilation or gluon fusion. At the centre-of-mass energy of 13 TeV from the LHC, gluon fusion is by far the dominant process. Top quarks can also be produced in single production, but has, being a electroweak process, a much smaller cross section. Hence, this analysis will focus on pair produced top quarks and will treat single top production as a background process.

Measuring the top quark mass is usually performed via a template fit. Therefore $t\bar{t}$ events are simulated for different masses m_{top}^{MC}. Then a selection is applied to data and simulation with the goal to select preferably $t\bar{t}$ events. Finally the different simulation with different m_{top}^{MC} are fitted to data. The most precise mass measurement of this kind is archived combining results of the CMS, ATLAS, CDF, and D0 Collaborations. The result of this world average is a value

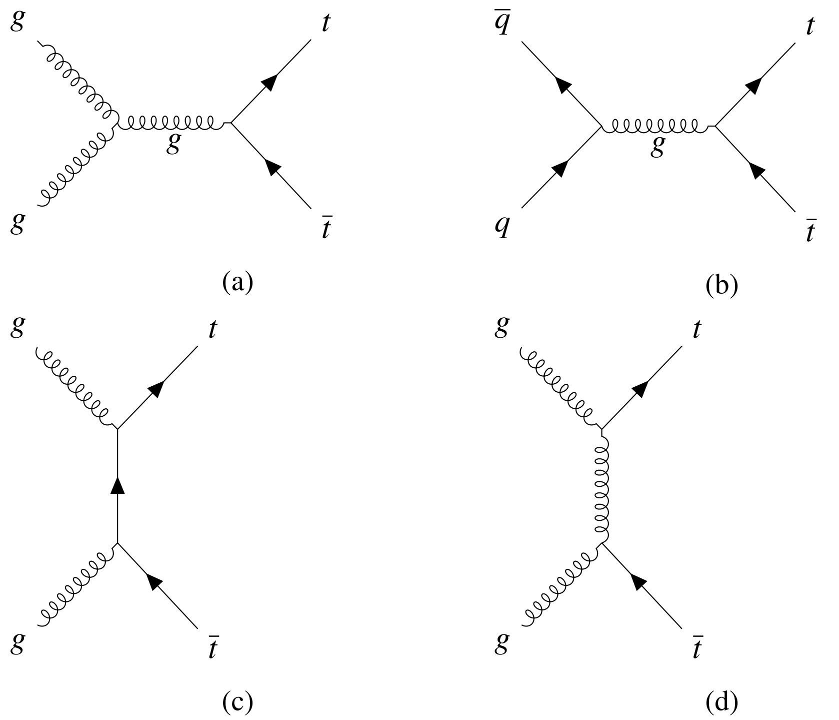

Figure 2.2: Feynman diagrams [4] showing the production of top quark pairs. Displayed are the gluon-gluon fusion (a), the quark anti-quark annihilation (b) and the t-channel (c+d).

of $173.34 \pm 0.27(\text{stat}) \pm 0.71(\text{sys})$ GeV [3].

The top quark decays via the weak interaction with a probability of almost 100% into a bottom quark and a W boson. While the bottom quark is seen as a jet in the detector, the W boson decays further into a quark anti-quark pair (see fig. 2.3a) or into a lepton and a neutrino. These two cases are called hadronically respectively hadronically top quark decay (see fig. **??**). Looking at the $t\bar{t}$ production, these two possible decays for each top quark corresponds to three channels for the $t\bar{t}$ process.

- both top quarks decay into quarks (full hadronic)
- one top quark decays hadronically, the other one leptonically (lep+jets)
- both top quarks decay leptonically (dilepton)

The full hadronic and lepton+jets channels are dominant and occur 45.7% respectively 43.8% of the time. 10.5% of all $t\bar{t}$ events result in two leptonically decaying top quarks [2]. This analysis will focus on the lepton+jets channel which is pictured in figure 2.4.

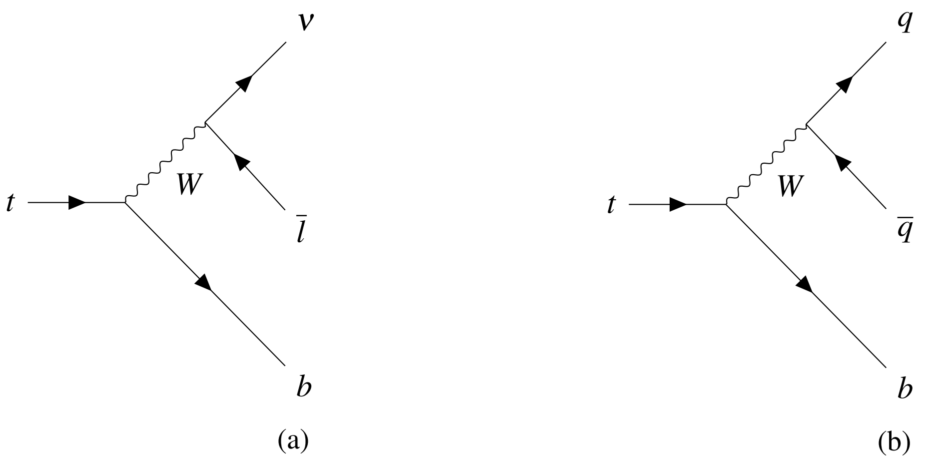

Figure 2.3: Feynman diagrams [4] of a leptonically (a) and hadronically (b) decaying top quark.

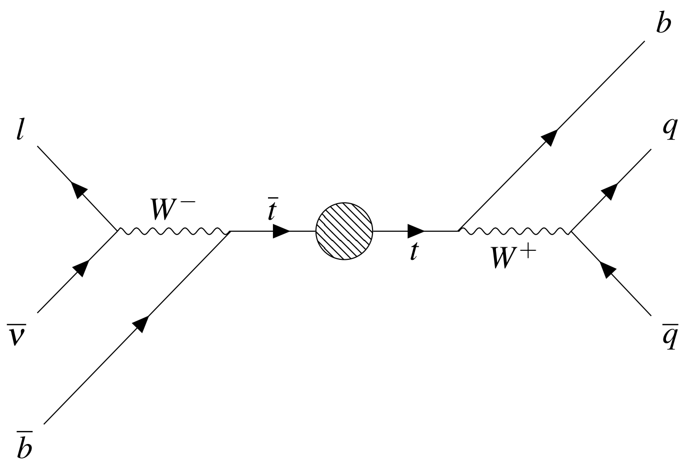

Figure 2.4: Feynman diagram [4] displaying an event from the lepton+jets channel. The centred circle indicates a mechanism to produce a $t\bar{t}$ pair.

2.2 Unfolding

Most analyses at LHC measure distributions of appropriate variables and then compare the obtained results in data with event simulations. In this method the MC samples also include detector effects. What one measures in this case is the real distribution on particle level folded with an unknown detector function. Studying the difference in MC between particle level and reconstruction level it is possible to calculate the probabilities that a measured value in a bin x_i is originating from bin y_i on particle level. The resulting matrix can then be applied to real data to obtain data on particle level. This can then be compared with theory calculations.

Chapter 3

Experiment

3.1 Large Hadron Collider

The Large Hadron Collider (LHC) is a circular particle collider with a circumference of 27 km.

Four Experiments are placed at each interaction point of LHC.

3.2 CMS Detector

The 'Compact Muon Solenoid' (CMS) experiment is one of the multi purpose detectors at LHC.

3.2.1 Tracker

The purpose of the tracking system is to measure the momentum of created particles. It is taken advantage of the Lorentz force which changes the momentum of charged particles in a magnetic field. Paths of those particles are therefore bended with a bending radius proportional to the momentum in a constant magnetic field. Therefore a solenoid provides a constant magnetic field of about 4 T inside the tracker. Per path one spatial point per layer is measured. These points are the input for a track finding algorithm.

3.2.2 Calorimeters

To define the type of particle one has to measure not only momentum but also energy. Therefore outside the tracker are two types of calorimeters installed.

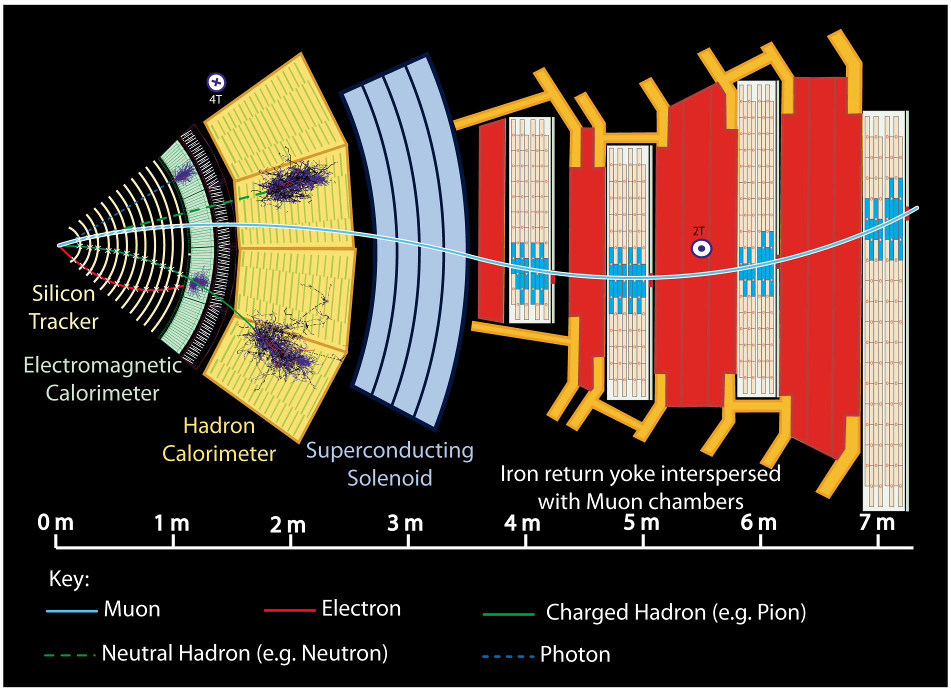

Figure 3.1: Slice through the CMS detector looking in direction of the beam pipe [5]. Tracks of different particles on their way through the layers of the detector are displayed. From left to right, you can see the tracker, calorimeters, the superconducting solenoid and the muon system. These detector components, their purpose and how they work are described in the following sections 3.2.1 - 3.2.4.

Electromagnetic Calorimeter

Hadronic Calorimeter

3.2.3 Solenoid

Outside the hadronic calorimeter a superconducting solenoid is installed and provides a magnetic field of about 4 T inside the tracking system. The solenoid has a length of 13 m and a diameter of 6 m. The purpose of the magnetic field is to bend paths of charged particles due to the Lorentz force. The bending radius then is directly connected to the momentum of the particle.

3.2.4 Muon System

Because muons are not absorbed in the calorimeters, it is possible to use an additional tracking detector for muons at the very outside of the detector. Muon chambers are embedded in the iron return yoke of the magnet. Thus a magnetic field of about 2 T is present. In cooperation with the inner tracker it is possible to reconstruct muon momenta with high precision.

3.2.5 Trigger

At the interaction points, proton bunches are brought to collision every 25 ns. Per crossing about ???? interactions take place. This adds up to about ??? interactions per second. It is impossible to store data with this rate and the total required storage capacity would is not feasible either. Additionally it would need a tremendous number of computing cores to analyse this data. Therefore a trigger system is installed to select only the interesting events to reduce the amount of data. For the computation system it is required to reduce the amount from ???? to ????. To decide which events are worth storing and which will be neglected different criteria are defined.

Chapter 4

Reconstruction of Objects

From all kinds of information provided by the detector systems has to be interpreted as physical objects in order to analyse the recorded data. Therefore algorithms are run to link the information from the detector to usable objects like muons, electrons or jets. In this chapter these objects and how they are defined is described.

4.1 Coordinate System

The coordinate system used in the CMS experiment is based on cartesian and right-handed coordinates. The origin is set in the center of the CMS detector. To define the direction of the axes other fix points are set. The x-axis point in the direction of the center of the LHC ring, the y-axis points up and the z-axis is defined parallel to the beam axis. Important variables used in analysis of CMS data are the angles ϕ and θ. ϕ is defined as the angle in the x-y-plane measured from the x-axis and θ describes the angle from a given point to the beam axis. Because the LHC is a hadron collider and physical events are therefore not symmetric in θ it is useful to construct the Lorentz invariant variable η:

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \tag{4.1}$$

The distance ΔR between two objects i and j is calculated using the differences $\Delta\phi = \phi_i - \phi_j$ and $\Delta\eta = \eta_i - \eta_j$:

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} \tag{4.2}$$

An important quantity used in this analysis is the transversal momentum p_T which is constructed out of the x and y-components (p_x and p_y) of the total momentum of an object:

$$p_T = \sqrt{p_x^2 + p_y^2} \tag{4.3}$$

It is practical to not consider the z component in a hadron collider because it depends on the initial state of interacting partons which is unknown. The p_T sum of all objects is expected to be 0 in every event. If it is not the p_T may be reconstructed wrong for some objects or objects left the CMS experiment undetected.

4.2 Particle Flow

CMS uses a special algorithm called particle flow [6] to combine information from tracker and calorimeters and thus reconstruct objects to a high precision. Tracks from the inner tracker are extrapolated into the calorimeters. If a shower fits to the track, information from these two subdetectors are combined. Since only charged particles will interact with the tracking system, showers from neutral hadrons and photons cannot be associated with a track.

4.3 Electrons

Electrons are reconstructed in the tracker and electromagnetic calorimeter. Due to the magnetic field, electrons will take a curved path through the tracking system.

4.4 Muons

Since the energy of muons cannot be measured in the calorimeters, a muon is identified by hits in the muon chambers. Combining informations from the muon system and the inner tracker leads to a very precise measurement of muons in the CMS detector. Therefore muons are very reliable objects in CMS data analyses. An object is announced a muon if

4.5 Jets

Jets are objects, used to reconstruct quarks. Because of confinement quarks cannot exist isolated but hadronize. This results in a particle shower consisting of hadrons. To reconstruct the initial quark one needs to sum up all particles from the final shower. To combine the tracks in a well defined way jet algorithms are used.

4.5.1 Anti-k_T Jet Algorithm

In CMS the common way to cluster jets from the detected particles is to use an iterative jet algorithm, especially the Anti-k_T (AK) algorithm [7].

Here d_{ij} describes a effective distance between two objects i and j, d_{iB} is a distance measure from object to beam axis.

$$d_{ij} = min(k_{T,i}^{-2}, k_{T,j}^{-2}) \frac{\Delta R_{ij}^2}{R^2} \tag{4.4}$$

$$d_{iB} = k_{T,i}^{-2} \tag{4.5}$$

4.5.2 HOTVR Jet Algorithm

Another approach to cluster jets is the 'heavy object tagger with variable R' (HOTVR) [8]. This algorithm does not use a constant radius parameter R but a p_T dependent effective R_{eff} (see eq. 4.6). Thus the R_{eff} decreases with increasing p_T leading to smaller jets when the decay products are expected to be more close because of the Lorentz boost. The p_T dependence is scaled with a parameter ρ with a default value of 600 GeV. Additionally upper and lower boundaries for the jet radius can be set. The default values are $R_{\text{min}} = 0.1$ and $R_{\text{max}} = 1.5$.

$$R_{\text{eff}} = \begin{cases} R_{\text{min}} & \text{for } \rho/p_T < R_{\text{min}} \\ R_{\text{max}} & \text{for } \rho/p_T > R_{\text{max}} \\ \rho/p_T & \text{else} \end{cases} \tag{4.6}$$

This effective R is then used with the equations of the Anti-k_T algorithm described earlier (see eq. 4.4 - 4.5).

4.5.3 XCone Jet Algorithm

XCone [9] is an exclusive jet algorithm, returning conical jets. It is well suited for analysis where the final state and therefore the expected number of jets is known since it returns a fixed number of jets. Thus, a physical final state has direct influence on the jet finding.

Starting with a fixed number of jet axes N the algorithm calculates the direction of these axes by minimizing the N-jettiness variable. N-jettiness is a measure for how N-jet-like an event looks. The definition is shown in Eq. 4.7.

$$\tilde{\tau}_N = \sum_i \min\{\rho_{\text{jet}}(p_i, n_1), \ldots, \rho_{\text{jet}}(p_i, n_N), \rho_{\text{beam}}(p_i)\} \tag{4.7}$$

Once the minimizing process converges, all particles inside a radius R from a jet axis are added to one jet.

Since the N-jettiness variable is often used to du theory calculations of particle physics events, the XCone algorithm is easier to include in these calculations.

4.5.4 b-Jets

In this analysis jets originating from a bottom quark are identified to reduce background. To identify a jet as an b-jet the "Combined Secondary Vertex" (CSV) algorithm is used. Since b-hadrons have a large lifetime of 1.5 ps they travel about 450 µm in the detector before decaying. This leads to a secondary vertex at the spatial point where the hadron decays which can be reconstructed in the tracking system. Additionally the composition of hadrons in a b-jet is different from other jets. These properties are taken advantage of in the CSV algorithm.

4.5.5 $\not{E}_T$ and S_T

With the information of mentioned objects two important variables are defined. The missing transverse energy $\not{E}_T$ is defined to estimate the energy carried away by particles which leave the experiment undetected, like neutrinos. Summing up all transverse momenta of each particle in the final state returns the p_T of the system in the initial state which is $p_T = 0$. Thus the transverse energy of all undetected particles is defined as the absolute value of the negative sum over all transverse momenta of detected objects (Eq. 4.8). Since every object, independent from it's η, has to be taken into account for this variable, $\not{E}_T$ is independent from the selection applied.

$$\not{E}_T = \left| - \sum_{\text{leptons, jets}} \vec{p}_T \right| \tag{4.8}$$

Another important variable to describe an event is S_T. S_T is defined to be the scalar sum of all transverse momenta of reconstructed objects plus the missing transverse energy mentioned above (Eq. 4.9). Different from $\not{E}_T$ only objects surviving the selection are considered.

$$S_T = \left(\sum_{\text{leptons, jets}} |\vec{p}_T| \right) + \not{E}_T \tag{4.9}$$

Additionally, a definition for S_T which just takes leptonic activity into account is used in this analysis. It is similarly defined and referred to as S_T^{lep} (see Eq. 4.10).

$$S_T^{\text{lep}} = \left(\sum_{\text{leptons}} |\vec{p}_T| \right) + \not{E}_T \tag{4.10}$$

Chapter 5

Analysis

5.1 Data and MC Simulation

This Thesis analyses data recorded by the CMS detector in the years 2015 and 2016 at a centre-of-mass energy of 13 TeV. The size of the data set corresponds to an integrated luminosity of ???. Additionally MC samples listed in table 5.1 are processed. The most important simulation for this analysis is of course the $t\bar{t}$ sample. The main background are W+jets and single-top.

Process	Sample	Cross Section [pb]	MC Generator	Number of Events
$t\bar{t}$	$0 < M_{t\bar{t}} < 700$	831.76		
$t\bar{t}$	$700 < M_{t\bar{t}} < 1000$	76.605		
$t\bar{t}$	$1000 < M_{t\bar{t}} < \infty$	20.578		
Single Top	t-channel			
Single Top	t-channel (anti top)			
Single Top	tW-channel			
Single Top	tW-channel (anti top)			
Single Top	s-channel			
W+jets	$100 < S_T < 200$			
W+jets	$200 < S_T < 400$			
W+jets	$400 < S_T < 600$			
W+jets	$600 < S_T < 800$			
W+jets	$800 < S_T < 1200$			
W+jets	$1200 < S_T < 2500$			
W+jets	$2500 < S_T < \infty$			
Z+jets				
QCD (muon enriched)				
QCD (EM enriched)				
WW				
WZ				
ZZ				

Table 5.1: Summary of data sets and MC samples used in this analysis. Assumed cross section, MC generator and number of events are displayed for each sample.

5.2 Jet Studies

For this analysis it is crucial to choose a suitable jet algorithm and cone size. A similar analysis has been performed by CMS on LHC's 2012 data set with a center-of-mass energy of 8 TeV with Cambridge-Aachen jets with a radius parameter of $R = 1.2$. This large radius was chosen to compensate for low statistics in the boosted $t\bar{t}$ regime. Since the cross-section of $t\bar{t}$ production is much higher at a center-of-mass energy of 13 TeV it has to be studied, how the jet radius parameter R influences the measured distribution and what influence different jet algorithms have.
The jet studies are performed with a $t\bar{t}$ simulation using the information of generator particles. Only events are considered where one top quark decays leptonically while the hadronically decaying top quark carries a transverse momentum greater 300 GeV.

5.2.1 XCone Strategy

The XCone jet algorithm described in section 4.5.3 has already been tested resolving $t\bar{t}$ decays. Studies for hadronically decaying top quark pairs are presented in a paper from Thaler and Wilkason [10]. Here, the XCone algorithm is tuned to the $t\bar{t}$ final state, expecting six jets. Using the information that it is expected to find three jets from each top quark, a promising approach to reconstruct the top quark decays was made with a strategy using two clustering steps. Firstly, the event is divided in two parts. This is done by require the XCone algorithm to find exactly two jets with a radius parameter $R = \infty$. Thus, every particle in the event is clustered into one of the jets. The goal of this first step is to separate the two top quarks into independent jets. Now, a second clustering step finding three jets is run where separate lists of particles from each fat jet are used as an input. Thus, in each fat jet, three smaller jets with $R = 0.4$ are found. These three small jets are then combined and used as the final top jets.

In this analysis the lepton+jets channel is used. Therefore the jet originating from the hadronically decaying top quark is identified while calculating the distance between each jet and the lepton. Additionally studies with a method where the fat jet originating from the leptonically decaying top quark is divided into two instead of three jets is presented. This approach was made because the neutrino from the W decay cannot be seen in the detector. Nevertheless it proved to be handy to use the $3+3$ method for data and therefore also for MC samples after detector simulation.

5.2.2 Comparing Jet Algorithms

5.3 Selection

To obtain a data set consisting of mostly $t\bar{t}$ events in the lepton+jets channel, a selection is applied to simulation and data. The selection can be divided into three main steps. Firstly, a Pre-Selection with very loose cuts is used to sort out non-relevant events and therefore reduce the size of the data set (see section 5.3.1). Secondly, a selection is made to suppress background processes (see section 5.3.2). Thirdly, the final phase space is defined (see section 5.3.3).

5.3.1 Pre-Selection

In the lepton+jets channel of the $t\bar{t}$ process one expects to find exactly one muon or electron, two small jets from the hadronically decaying W boson, two b-jets and missing transverse energy since the neutrino cannot be detected. This Pre-Selection is designed to remove events from the data set that are surely uninteresting for this analysis. Thus, very loose cuts are applied. For an event to be considered for this analysis, it has to contain

- 1 or more leptons (electrons or muons),
- missing transverse energy of more than 20 GeV,
- 2 or more AK4 jets with a p_T greater than 50 GeV.

5.3.2 Suppress Background

In this selection step cuts are used to reduce background processes to obtain a clean $t\bar{t}$ data set.

- exactly one muon or electron with a veto on additional leptons
- $S_T^{\text{lep}} > XX$
- B-TAG
- 2D Cut
- $\dots$

5.3.3 Measurement Phase Space

- $p_T^{\text{1st topjet}} > 400\,\text{GeV}$
- $M^{\text{1st topjet}} > M^{\text{2nd topjet}}$

5.4 Unfolding

5.5 Results

Chapter 6

Summary and Outlook

Bibliography

[1]CERN, *The Standard Model of Particle Physics*, https://webfest.web.cern.ch/sites/webfest.web.cern.ch/files/styles/large/public/Standard_model_infographic.png?itok=sZoMRgso, [Online; accessed January 31, 2017].

[2]C. Patrignani et al., “Review of Particle Physics”, Chin. Phys. **C40**, 100001 (2016).

[3]T. ATLAS, CDF, CMS, and D. Collaborations, *First combination of tevatron and lhc measurements of the top-quark mass*, 2014.

[4]J. Ellis, “TikZ-Feynman: Feynman diagrams with TikZ”, (2016) 10.1016/j.cpc.2016.08.019.

[5]D. Barney, *CMS slice*, https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/RetrieveFile?docid=5581&filename=PictureforPoint5_oct04_allp.jpg&version=1, [Online; accessed January 30, 2017].

[6]F. Beaudette, “The CMS Particle Flow Algorithm”, (2014).

[7]M. Cacciari, G. P. Salam, and G. Soyez, “The anti-kt jet clustering algorithm”, (2008) 10.1088/1126-6708/2008/04/063.

[8]T. Lapsien, R. Kogler, and J. Haller, “A new tagger for hadronically decaying heavy particles at the LHC”, (2016) 10.1140/epjc/s10052-016-4443-8.

[9]I. W. Stewart, F. J. Tackmann, J. Thaler, C. K. Vermilion, and T. F. Wilkason, “XCone: N-jettiness as an Exclusive Cone Jet Algorithm”, (2015) 10.1007/JHEP11(2015)072.

[10]J. Thaler and T. F. Wilkason, “Resolving Boosted Jets with XCone”, (2015) 10.1007/JHEP12(2015)051.

Hiermit bestätige ich, dass die vorliegende Arbeit von mir selbstst"andig verfasst wurde und ich keine anderen als die angegebenen Hilfsmittel – insbesondere keine im Quellenverzeichnis nicht benannten Internet-Quellen – benutzt habe und die Arbeit von mir vorher nicht einem anderen Prüfungsverfahren eingereicht wurde. Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium. Ich bin damit einverstanden, dass die Masterarbeit veröffentlicht wird.

Ort, Datum Unterschrift

Danksagung