

Messung der Jetmassenverteilung in kollinierten 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

1	Introduction	1
2	Experiment	2
2.1	Large Hadron Collider	2
2.2	CMS Detector	4
2.2.1	Tracker	4
2.2.2	Calorimeters	4
2.2.3	Solenoid	5
2.2.4	Muon System	6
2.2.5	Trigger	6
3	Reconstruction of Objects	7
3.1	Coordinate System	7
3.2	Particle Flow	8
3.3	Electrons	8
3.4	Muons	8
3.5	Jets	8
3.5.1	Anti- k_T Jet Algorithm	9
3.5.2	HOTVR Jet Algorithm	9
3.5.3	XCone Jet Algorithm	9
3.5.4	b-Jets	10
3.6	\cancel{E}_T and S_T	10
4	Analysis	12
4.1	Analysis Strategy	12
4.2	Previous Results	12
4.3	Data and MC Simulation	12
4.4	Jet Studies on Particle Level	14

4.4.1	Selection on particle level	14
4.4.2	XCone Strategy	14
4.4.3	Comparing Jet Algorithms	15
4.5	Selection on Reconstruction Level	15
4.5.1	Baseline Selection	15
4.5.2	Measurement Phase Space	16
4.6	Unfolding	16
4.7	Results	16
5	Summary and Outlook	17

Chapter 1

Introduction

Chapter 2

Experiment

2.1 Large Hadron Collider

The Large Hadron Collider (LHC) is a circular particle collider with a circumference of 27 km. At the LHC protons are accelerated to an energy of 6.5 TeV and then brought to collision resulting in a center-of-mass energy of $\sqrt{s} = 13$ TeV. Because protons are composite particles, the actual collision involves quarks or gluons carrying only a fraction of this energy.

Protons are not injected directly into the LHC ring but run through several acceleration stages. The preaccelerators used for this purpose are old accelerators that can now be reused.

Four Experiments are placed at each interaction point of LHC.

A display of the LHC complex with its four experiments and preaccelerators is shown in fig. ??.

A very important parameter of a particle collider is its luminosity. It gives an estimate how many collisions take place per area and second. It is calculated via Eq. 2.1. To get an estimate of the number of events produced by a given process one has to multiply the integrated luminosity (see Eq. 2.2) with the cross section of this particular process (see Eq. 2.3)

$$L = \quad (2.1)$$

$$L_{\text{int}} = \int L dt \quad (2.2)$$

$$N = L_{\text{int}} \sigma \quad (2.3)$$

CERN's Accelerator Complex

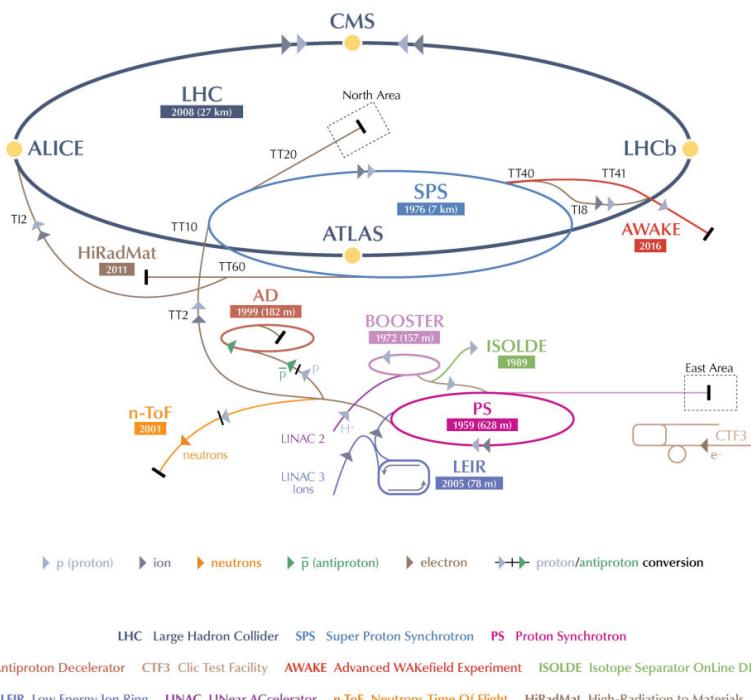


Figure 2.1: Display the LHC complex with the LHC ring itself and every accelerator used to bring protons to the required energy for injection in the LHC ring [fig:lhc]

2.2 CMS Detector

The 'Compact Muon Solenoid' (CMS) experiment is a multi purpose detectors at LHC. It is designed to measure momentum and energy of particles produced in proton-proton interactions. The CMS detector is built in layers of subdetectors with different purposes described in following sections 2.2.1 - 2.2.4.

A very important part besides the detector systems is the Trigger, described in section 2.2.5, providing fast decisions if an event is stored or discarded.

2.2.1 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. The largest uncertainties of the measured momenta rise from the spatial resolution of the tracker. Therefore a pixel structure is used which can measure the position of a passing charged particle within a resolution of ??. All measured spatial points are then put into a track finding algorithm returning track and momentum of every recorded object.

2.2.2 Calorimeters

To define the type of particle one has to measure not only momentum but also energy. Therefore outside the tracker a calorimetry system with two types of calorimeters is installed. The underlying principle of this subdetector is to absorb all of the energy a particle carries. This energy is then transformed into a light signal. The intensity of the light signal is proportional to the energy of the incoming particle. For this approach one needs to use scintillating materials. It also has to be dense enough to stop a incoming particle to transform all its energy into the light signal. At the end the light is collected and measured with photo multipliers.

Electromagnetic Calorimeter

The inner part of the calorimetry system is supposed to measure the energy of electrons and photons. The electormagnetic calorimeter of the CMS detector is designed to deliver a very

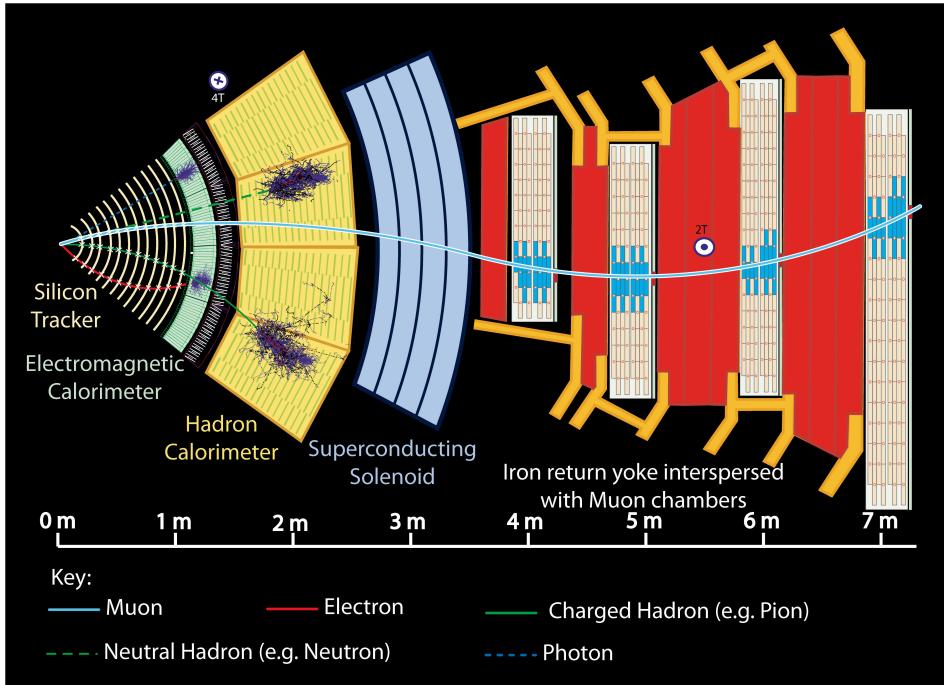


Figure 2.2: Slice through the CMS detector looking in direction of the beam pipe [2]. Tracks of different particles on their way through the layers of the detector are displayed. From left to right, you can see the tracker, calorimeters, the superconducting solenoid and the muon system. These detector components, their purpose and how they work are described in the following sections 2.2.1 - 2.2.4.

good spatial resolution. With this, photons which are not seen in the tracker can be reconstructed with a high precision.

Since hadrons have a larger absorption length they mostly pass the electromagnetic calorimeter and are measured in the hadronic calorimeter.

Hadronic Calorimeter

The hadronic calorimeter uses two alternating layers of different material. One is to absorb the energy of an incoming particle. The other one is made out of a scintillator.

2.2.3 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.

2.2.4 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.

2.2.5 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 be 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 3

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.

3.1 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 :

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right] \quad (3.1)$$

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

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} \quad (3.2)$$

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:

$$p_T = \sqrt{p_x^2 + p_y^2} \quad (3.3)$$

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.

3.2 Particle Flow

CMS uses a special algorithm called particle flow [3] 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.

3.3 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.

3.4 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

3.5 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.

3.5.1 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 [4].

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

$$d_{ij} = \min(k_{T,i}^{-2}, k_{T,j}^{-2}) \frac{\Delta R_{ij}^2}{R^2} \quad (3.4)$$

$$d_{iB} = k_{T,i}^{-2} \quad (3.5)$$

3.5.2 HOTVR Jet Algorithm

Another approach to cluster jets is the 'heavy object tagger with variable R' (HOTVR) [5]. This algorithm does not use a constant radius parameter R but a p_T dependent effective R_{eff} (see eq. 3.6). 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 $R_{\text{max}} = 1.5$.

$$R_{\text{eff}} = \begin{cases} R_{\text{min}} & \text{for } \rho/p_T < R_{\text{min}} \\ R_{\text{max}} & \text{for } \rho/p_T > R_{\text{max}} \\ \rho/p_T & \text{else} \end{cases} \quad (3.6)$$

This effective R is then used with the equations of the Anti- k_T algorithm described earlier (see eq. 3.4 - 3.5).

3.5.3 XCone Jet Algorithm

XCone [6] 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. 3.7.

$$\tilde{\tau}_N = \sum_i \min\{\rho_{\text{jet}}(p_i, n_1), \dots, \rho_{\text{jet}}(p_i, n_N), \rho_{\text{beam}}(p_i)\} \quad (3.7)$$

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.

3.5.4 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.

3.6 E_T and S_T

With the information of mentioned objects two important variables are defined. The missing transverse energy \cancel{E}_T is defined to estimate the energy carried away by particles which leave the experiment undetected. 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 $p_T = 0$ at LHC. 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. 3.8). Since every object, independent from it's η , has to be taken into account for this variable, \cancel{E}_T is independent from the selection applied. The missing energy is due to neutrinos which can not be detected with CMS because of their low probability to interact with the detector material or a new physics state.

$$\cancel{E}_T = \left| - \sum_{\text{leptons, jets}} \vec{p}_T \right| \quad (3.8)$$

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. 3.9). Different from \cancel{E}_T only objects surviving the selection are considered.

$$S_T = \left(\sum_{\text{leptons, jets}} |\vec{p}_T| \right) + E_T \quad (3.9)$$

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. 3.10).

$$S_T^{\text{lep}} = \left(\sum_{\text{leptons}} |\vec{p}_T| \right) + E_T \quad (3.10)$$

Chapter 4

Analysis

4.1 Analysis Strategy

This analysis aims for boosted $t\bar{t}$ where all decay products from the top decays merge into a single jet. To measure in this phase space a selection of events is applied. The detailed selection is presented in section 4.5. In the required phase space the distribution of the jet mass of a top quark decaying into quarks ($t \rightarrow W^+ b \rightarrow b q \bar{q}'$) is measured. Following a unfolding is performed using the TUnfold [7] software package. The goal is to compare data unfolded to particle level with first principle calculations.

4.2 Previous Results

A measurement of the top quark mass in highly boosted $t\bar{t}$ events has already been performed by CMS with the 8 TeV dataset corresponding to an integrated luminosity of 19.7 fb^{-1} [8].

4.3 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 4.1 are processed. The most important simulation for this analysis is of course the $t\bar{t}$ sample. The main background processes are $W+\text{jets}$ and single-top production.

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

Table 4.1: 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.

4.4 Jet Studies on Particle Level

For this analysis it is crucial to choose a suitable jet algorithm and cone size. The previous mentioned analysis [8] uses Cambridge-Aachen jets with the radius parameter set to $R = 1.2$ for its measurement. 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 goal is to select a jet algorithm which returns jets in which all decay products of a hadronically decaying top quark are merged. In this case, the jet mass M_{jet} is sensitive to the top quark mass M_{top} . The jet studies are performed with a $t\bar{t}$ simulation using the information of MC simulations at particle level. The detailed selection is described in Section 4.4.1.

4.4.1 Selection on particle level

Only events are considered where one top quark decays leptonically into $t \rightarrow W^+ b \rightarrow \mu^+ \nu_\mu b$ or $t \rightarrow W^+ b \rightarrow e^+ \nu_e b$ while the hadronically decaying top quark carries a transverse momentum greater 300 GeV. On this $t\bar{t}$ sample, a selection is applied to select boosted top quarks:

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

Where the first jet refers to the leading jet in p_T . The first jet is expected to be originating from the hadronically decaying top quark. When using jets clustered with the XCone jet algorithm the first jet already refers to the jet identified as originating from the hadronically decaying top quark because a distance requirement to the lepton is used in the jet finding (see section 4.4.2).

4.4.2 XCone Strategy

The XCone jet algorithm described in section 3.5.3 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 [9]. 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('2 + 3'). 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.

4.4.3 Comparing Jet Algorithms

4.5 Selection on Reconstruction Level

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 two steps. Firstly, a baseline selection is used to suppress background processes (see section 4.5.1). Secondly, the final phase space is defined (see section 4.5.2) to select $t\bar{t}$ events with boosted top quarks. This is crucial for this analysis because the goal is to reconstruct the top quark with one jet. This can only be done if all of its decay products merge into one jet.

4.5.1 Baseline 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 baseline selection is designed to remove non- $t\bar{t}$ events. After applying this selection the remaining sample consists of about 80% $t\bar{t}$ events. The main remaining backgrounds are W +jets and Single-Top production.

- trigger
- missing transverse energy of more than 50 GeV,
- 2 or more AK4 jets with a p_T greater than 50 GeV.

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

4.5.2 Measurement Phase Space

This analysis focuses on boosted top quarks. Therefore the leading jet, which is supposed to contain all decay products of the hadronically decaying top quark, is required to carry a transverse momentum greater than 400 GeV. Additionally only events are selected where the leading jet has a greater mass than the second jet. This last selection step prefers events with merged jets because here the jet from the leptonic top quark will only contain the lepton and a jet from the bottom quark since neutrinos cannot be detected. Therefore the mass of the hadronic jet is expected to be larger.

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

4.6 Unfolding

4.7 Results

Chapter 5

Summary and Outlook

Bibliography

- ¹F. Marcastel, “CERN’s Accelerator Complex. La chaîne des accélérateurs du CERN”, OPEN-PHO-CHART-2013-001 (2013).
- ²D. 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].
- ³F. Beaudette, “The CMS Particle Flow Algorithm”, (2014).
- ⁴M. Cacciari, G. P. Salam, and G. Soyez, “The anti- kt jet clustering algorithm”, (2008) 10.1088/1126-6708/2008/04/063.
- ⁵T. 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.
- ⁶I. 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.
- ⁷S. Schmitt, “Tunfold: an algorithm for correcting migration effects in high energy physics”, (2012) 10.1088/1748-0221/7/10/T10003.
- ⁸C. Collaboration, *Measurement of the jet mass in highly boosted t - $t\bar{\nu}$ events from pp collisions at $\sqrt{s} = 8 \text{ tev}$* , 2017.
- ⁹J. 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