

UNIVERSITÄT HAMBURG

MASTERARBEIT

IM STUDIENGANG PHYSIK

---

# Measurement of the Jet Mass Distribution in Boosted Top Quark Decays

---

*Autor:*

DENNIS SCHWARZ

*Gutachter:*

PROF. DR. JOHANNES HALLER

PROF. DR. PETER SCHLEPER



2017



# **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

	Page
<b>1 Introduction</b>	<b>1</b>
<b>2 Theory</b>	<b>2</b>
2.1 Standard model of particle physics . . . . .	2
2.2 Top Quark . . . . .	6
2.3 Unfolding . . . . .	8
<b>3 Experiment</b>	<b>9</b>
3.1 Large Hadron Collider . . . . .	9
3.2 CMS Detector . . . . .	11
3.2.1 Tracker . . . . .	11
3.2.2 Calorimeters . . . . .	11
3.2.3 Solenoid . . . . .	12
3.2.4 Muon System . . . . .	13
3.2.5 Trigger . . . . .	13
<b>4 Reconstruction of Objects</b>	<b>14</b>
4.1 Coordinate System . . . . .	14
4.2 Particle Flow . . . . .	15
4.3 Electrons . . . . .	15
4.4 Muons . . . . .	15
4.5 Jets . . . . .	15
4.5.1 Anti- $k_T$ Jet Algorithm . . . . .	15
4.5.2 HOTVR Jet Algorithm . . . . .	16
4.5.3 XCone Jet Algorithm . . . . .	16
4.5.4 b-Jets . . . . .	17
4.6 $E_T$ and $S_T$ . . . . .	17
<b>5 Data and Monte-Carlo Simulation</b>	<b>19</b>
5.1 Data . . . . .	19

## CONTENTS

---

5.2	Event Generator and Parton Shower . . . . .	19
5.3	Monte-Carlo samples . . . . .	19
<b>6</b>	<b>Analysis</b>	<b>21</b>
6.1	Analysis Strategy . . . . .	21
6.2	Previous Results . . . . .	21
6.3	Jet Studies on Particle Level . . . . .	21
6.3.1	Selection on particle level . . . . .	22
6.3.2	XCone Strategy . . . . .	22
6.3.3	Comparing Jet Algorithms . . . . .	23
6.4	Selection on Reconstruction Level . . . . .	23
6.4.1	Baseline Selection . . . . .	23
6.4.2	Measurement Phase Space . . . . .	24
6.5	Unfolding . . . . .	24
6.6	Results . . . . .	24
<b>7</b>	<b>Summary and Outlook</b>	<b>25</b>



# 1 | Introduction

# 2 | Theory

The goal of particle physics is to explain the universe we observe with basic building blocks which are not further divisible. Starting with quantum mechanics in the early 20th century, physicists have developed a theory answering many questions about the development of the universe. Up to now the resulting standard model of particle physics is the most precisely tested physical theory. In the following the basics of this theory will be described. A special focus is set to the elementary particle which plays the most important role in this analysis - the top quark.

## 2.1 Standard model of particle physics

The standard model of particle physics is a quantum field theory describing elementary particles, their interaction and provides the affiliated equations of motion. All particles contained in this theory have been discovered and various experiments have confirmed predictions of the standard model at very high precision. All known elementary particles of the standard model and their properties - mass, electromagnetic and colour charge and their spin - are displayed in Fig. 2.1. They are ordered in two fundamental groups by their spin.  $\text{spin-}\frac{1}{2}$  particles are called fermions, particles with integer spin are named bosons.

Fermions are further divided into quarks and leptons. While quarks are affected by the strong force, leptons are not. Both subgroups consist of three generations with two particles in each group. Each generation of quarks contain one up-type and one down-type quark distinguished by their electromagnetic charge. For leptons a generation is built from a particle with charge of  $1e$  and a neutral neutrino. The first generation of quarks and leptons then are the building blocks of atoms. All other particles have a larger mass and therefore a finite lifetime. Additionally every fermion has a anti-partner which has the same mass but opposite charge.

Bosons are the carriers of the three fundamental forces included in the standard model. Every interaction of two particles is described by a boson emitted from one and absorbed by the other particle. Thus a boson changes energy and momentum of an absorbing or emitting particle. For each force a charge is introduced. Only particles carrying the charge connected to a force can interact with the corresponding boson. Thus, the amount of charge carried by a particle is proportional to the probability of emitting or absorbing a boson.

### Quantum Electro Dynamics (QED)

In the history of the standard model the QED was the first mechanism to describe interactions. The charge of QED is the electromagnetic charge and usually stated in units of elementary charge  $e$ . To construct a quantum field theory ,the basic idea is to take the Lagrangian for particles described by Dirac spinors and require local gauge invariance. Equation 2.1 shows the Lagrangian of the Dirac equation.

$$\mathcal{L} = \underbrace{\bar{\psi} i \gamma^\mu \partial_\mu \psi}_{\text{kinetic term}} - \underbrace{m \bar{\psi} \psi}_{\text{mass term}} \quad (2.1)$$

Here,  $\psi$  denotes the Dirac spinors and  $m$  the mass corresponding to the Dirac spinor. While the mass term  $m \bar{\psi} \psi$  is invariant under a transformation  $\psi \rightarrow \psi'$ , the derivative  $\partial_\mu \psi$  is not. In order to construct a symmetric Lagrangian one introduces a new vector field  $A_\mu(x)$  and defines a covariant derivative  $D_\mu$ :

$$D_\mu = \partial_\mu + iqA_\mu(x) \quad (2.2)$$

where the parameter  $q$  will later be identified with a coupling strength. With the new introduced covariant derivative the Lagrangian reads:

$$\mathcal{L} = \underbrace{\bar{\psi} i \gamma^\mu \partial_\mu \psi}_{\text{kinetic term}} - \underbrace{m \bar{\psi} \psi}_{\text{mass term}} - \underbrace{q \bar{\psi} \gamma^\mu \psi A_\mu(x)}_{\text{interaction term}}. \quad (2.3)$$

Thus, by requiring the Lagrangian to be symmetric under gauge transformations a new vector field is predicted. This vector field can be associated with the photon. Additionally the coupling between fermions  $\psi$  and photon  $A_\mu$  are included.

### Quantum Chromo Dynamics (QCD)

The boson of the strong force is a massless gluon. It couples to particles carrying a so-called colour charge. Colour charge exists in three states for particles: red, green and blue. Additionally every colour has its anti-colour carried by anti-particles. The gluon itself carries one colour and one anti-colour simultaneously. Thus, quarks and the gluon itself are the only candidates affected by the strong force. The theory behind the strong force, QCD, is constructed very similarly to QED.

### Weak Interaction

The third force described in the standard model is the weak force. It is carried by the  $W^\pm$  and  $Z$  bosons. Because all weak bosons carry a mass, the interaction only takes place on short

distances since the bosons decay very quickly.

### **Electroweak Unification**

To get a better understanding of the forces in the SM, it would be a large success to lead the different forces back to one fundamental mechanism. A first step towards a unification of all three interactions can be archived by combining QED with the weak force.

### **Higgs mechanism**

The remaining boson, the Higgs boson, was discovered the most recent in 2012. In the standard model without a Higgs mechanism, all bosons are predicted to be massless. In contradiction to this, the mass of the  $W$  and  $Z$  bosons are measured to values of 80 GeV and 91 GeV, respectively [2]. Therefore a theory was developed which solves this problem via spontaneous symmetry breaking. A field is introduced which where the lagrangian for local variations is not symmetric while the lagrangian of the field itself is. This leads to mass terms for electro-weak bosons and a new boson - the Higgs boson - in the standard model lagrangian. Additionally fermions get their mass through Yukawa coupling to the Higgs field proportional to their masses.

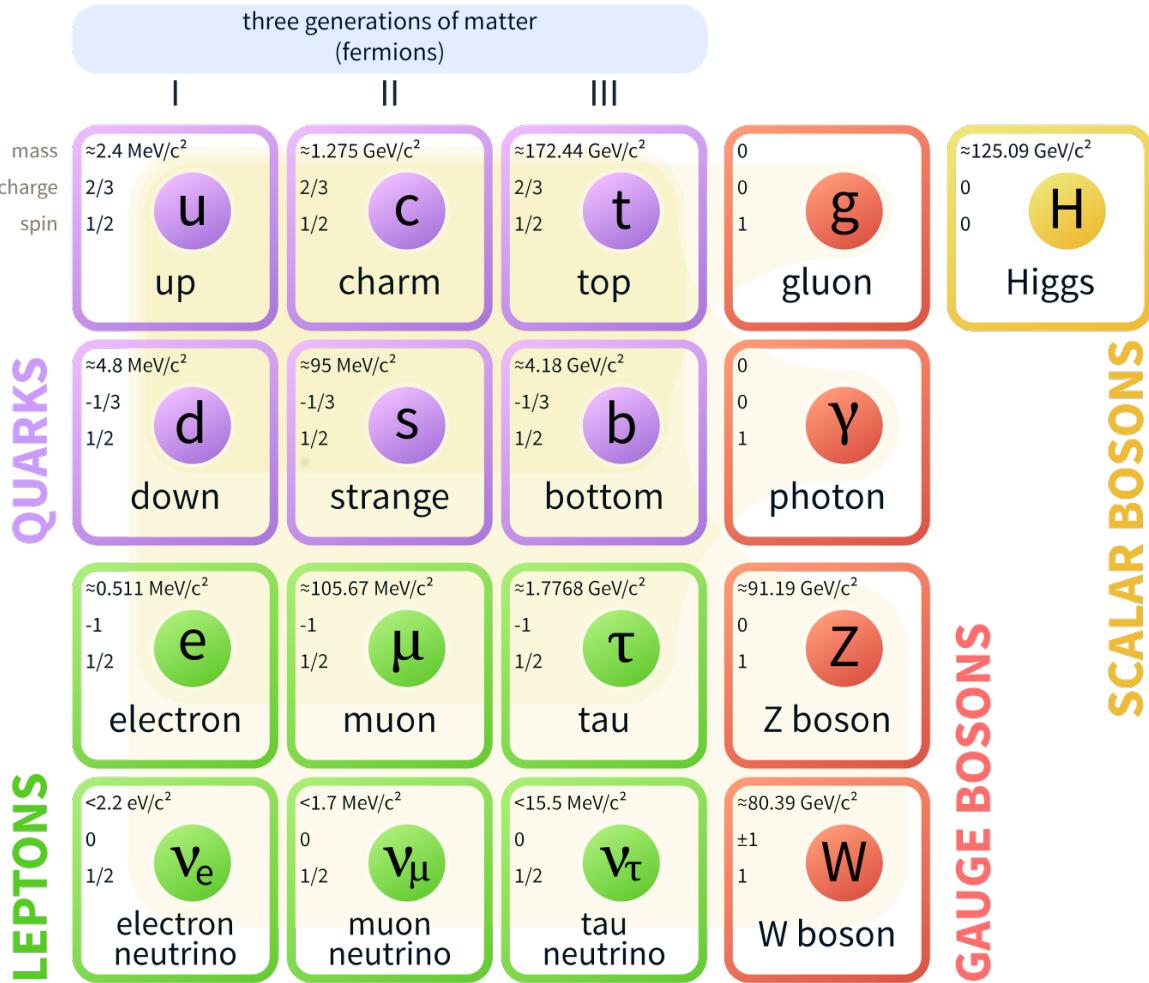


Figure 2.1: Display of the particle content of the standard model divided into groups of fermions and their three generations and bosons. For every particle mass, electromagnetic charge and spin is shown. Although neutrinos are supposed to be massless in the SM, experiments show that they carry mass. Thus, the upper limit of the neutrino masses are specified. [1]

## 2.2 Top Quark

The top quark is an up type quark belonging to the third generation and carrying electromagnetic charge of  $Q = +\frac{2}{3}e$  [2]. With its mass of about 173 GeV the top quark is the heaviest particle in the standard model. Therefore it offers a large phase space for decays and has a short life time of approximately  $0.5 \times 10^{-24}$  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 the masses of the top quark with the masses of the W and Higgs boson. This could be done, because the top quark plays an extraordinary role due to its high mass – and therefore large coupling to the Higgs boson – in this mechanism. 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 dominant 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.

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 (see fig. 2.3). Looking at the  $t\bar{t}$  production, these two possible final states 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 (lepton+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 leptons in the final state [2]. This analysis will focus on the lepton+jets channel which is pictured in figure 2.4. It is suitable because the lepton makes it easier to distinguish  $t\bar{t}$  events from background events but also includes a fully hadronically decaying top quark which will be the target for the presented measurement.

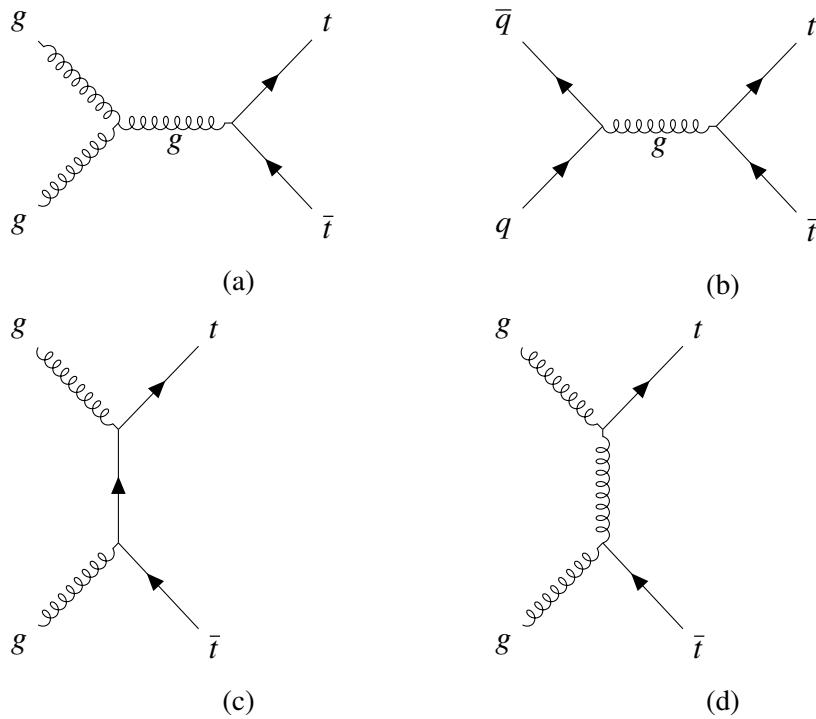


Figure 2.2: Feynman diagrams [3] showing the production of top quark pairs. Displayed are the gluon-gluon fusion (a), the quark anti-quark annihilation (b) and the t-channel (c+d).

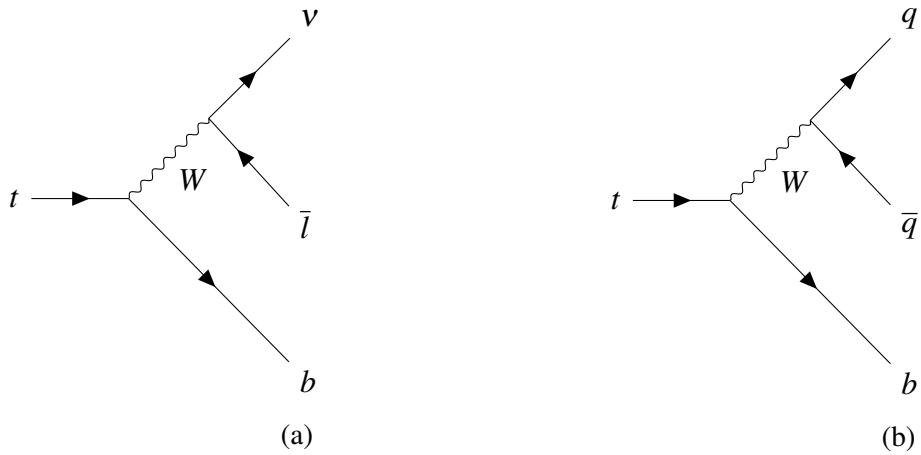


Figure 2.3: Feynman diagrams [3] of a leptonically (a) and hadronically (b) decaying top quark.

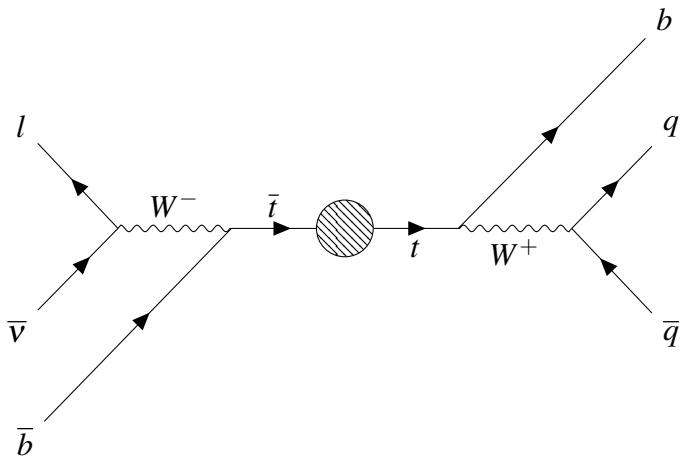


Figure 2.4: Feynman diagram [3] displaying an event from the lepton+jets channel. The centred circle indicates a mechanism to produce a  $t\bar{t}$  pair.

Measuring the top quark mass is usually performed via a template fit. Therefore  $t\bar{t}$  events are simulated for different masses  $m_{\text{top}}^{MC}$ . 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  $m_{\text{top}}^{MC}$  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 of  $173.34 \pm 0.27(\text{stat}) \pm 0.71(\text{sys})$  GeV [4]. Doing these kinds of measurements one relies on a correct simulation of  $t\bar{t}$  events. Especially the mass parameter in Monte-Carlo generators can not be related to a mass one would calculate in an Lagrangian.

## 2.3 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. The obtained distribution is corrected for detector effects by the unfolding and can then be compared with theory calculations. Additionally, measurements on particle level do not depend on simulation ambiguities like the mass parameter put into the Monte-Carlo generator.

# 3 | Experiment

For this analysis a data from a collision experiment is used. Protons are brought to collision in a circular particle accelerator. The outcome of the collisions is then measured with several detection systems. A description of the accelerator and the detector is presented in this section.

## 3.1 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. 3.1.

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. 3.1.

$$L = \frac{nN_1N_2f}{4\pi\sigma_x\sigma_y} \quad (3.1)$$

In this formula  $n$  denotes the number of bunches in the accelerator,  $N_1$  and  $N_2$  are the number of protons in the two colliding bunches and  $f$  is the collision frequency. The denominator gives the cross sectional area where  $\sigma_x$  and  $\sigma_y$  describes the spread of the proton beam in  $x$  and  $y$  direction, respectively. To get an estimate of the number of events produced by a given process one has to multiply the integrated luminosity (see Eq. 3.2) with the cross section of this particular process (see Eq. 3.3)

$$L_{\text{int}} = \int L dt \quad (3.2)$$

$$N = L_{\text{int}}\sigma \quad (3.3)$$

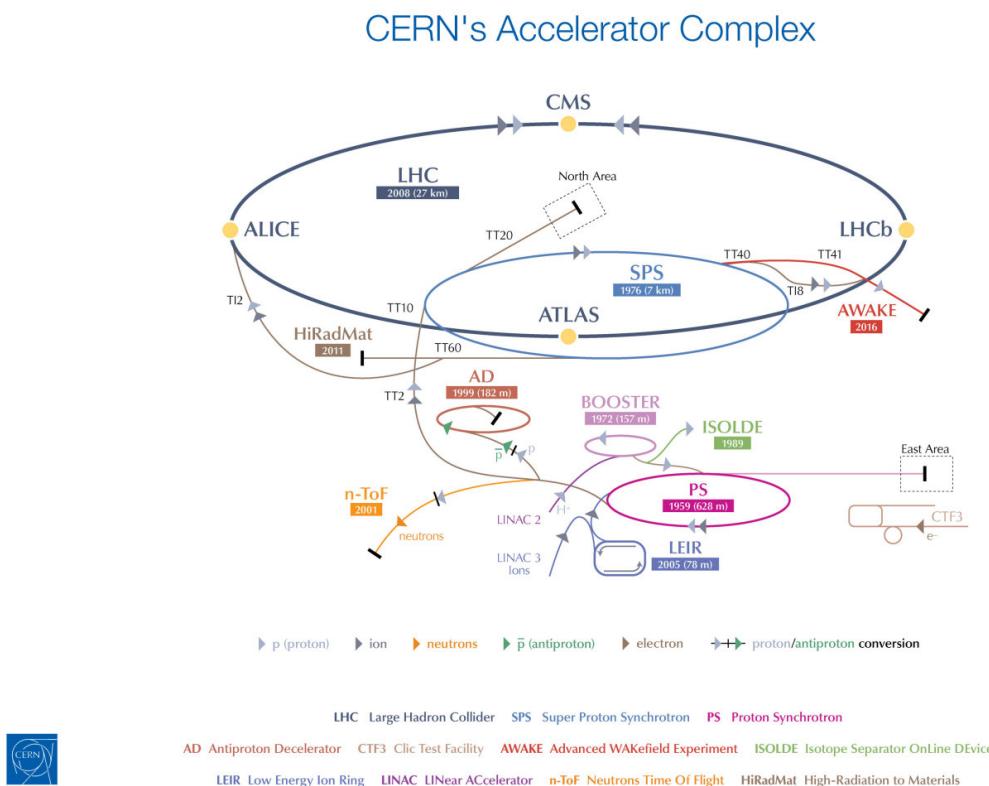


Figure 3.1: Display the LHC complex with the LHC ring itself and every accelerator used to bring protons to the required energy for injection in the LHC ring [5]

## 3.2 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 3.2.1 - 3.2.4.

A very important part besides the detector systems is the Trigger, described in section 3.2.5, providing fast decisions if an event is discarded or interesting enough to be stored.

### 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. 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.

### 3.2.2 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

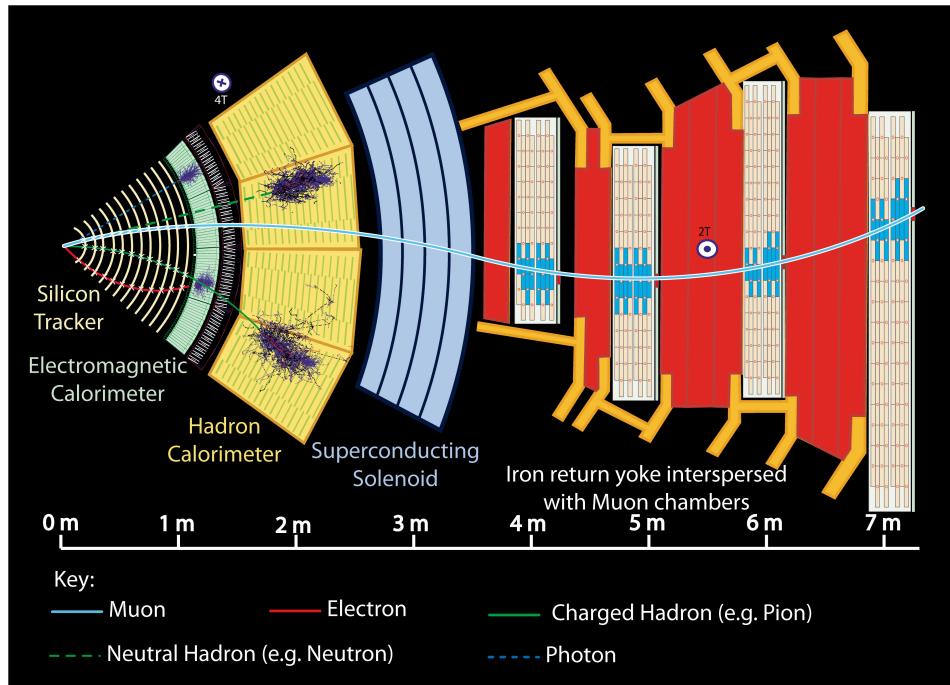


Figure 3.2: Slice through the CMS detector looking in direction of the beam pipe [6]. Tracks of different particles on their way through the layers of the detector are displayed. From left to right, you can see the tracker, calorimeters, the superconducting solenoid and the muon system. These detector components, their purpose and how they work are described in the following sections 3.2.1 - 3.2.4.

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.

#### 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 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.

# 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  $\phi$  and  $\theta$ .  $\phi$  is defined as the angle in the  $x$ - $y$ -plane measured from the  $x$ -axis and  $\theta$  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  $\theta$  it is useful to construct the Lorentz invariant variable  $\eta$ :

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

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

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} \quad (4.2)$$

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:

$$p_T = \sqrt{p_x^2 + p_y^2} \quad (4.3)$$

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 [7] 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. Two important requirements for an jet algorithm are to be infrared and collinear safe. All presented jet algorithms in this thesis fulfil these requirements.

### 4.5.1 Anti- $k_T$ Jet Algorithm

In CMS the common way to cluster jets from the detected particles is to use iterative jet algorithms. Thus particles are clustered step by step until an abort criterion is reached. The most common and proven to be useful iterative jet algorithm is the Anti- $k_T$  (AK) algorithm [8]. As an input a list of objects, reconstructed in the detector is given. The AK algorithm then

calculates two quantities  $d_{ij}$  (Eq. 4.4) and  $d_{iB}$  (Eq. 4.5) for each pair of objects  $i$  and  $j$ :

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

$$d_{iB} = k_{T,i}^{-2}. \quad (4.5)$$

Here  $d_{ij}$  describes an effective distance between two objects  $i$  and  $j$ ,  $d_{iB}$  is a distance measure from object to beam axis.  $k_T$  is the transverse momentum of an object,  $\Delta R_{ij}$  denotes the distance between objects  $i$  and  $j$  and  $R$  is a constant parameter that defines the radius of the resulting jet. When  $d_{ij}$  is smaller than  $d_{iB}$  both objects  $i$  and  $j$  are combined and both quantities are calculated again. At some point,  $d_{iB}$  will be smaller than any  $d_{ij}$ , then object  $i$  is called a jet and is removed from the list of objects. This procedure is repeated until the list of objects is empty.

### 4.5.2 HOTVR Jet Algorithm

Another approach to cluster jets is the 'heavy object tagger with variable R' (HOTVR) [9]. This algorithm does not use a constant radius parameter  $R$  but a  $p_T$  dependent effective  $R_{\text{eff}}$  (see eq. 4.6). Thus the  $R_{\text{eff}}$  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  $\rho$  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  $R_{\text{max}} = 1.5$ .

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

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

### 4.5.3 XCone Jet Algorithm

XCone [10] 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.

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

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 do 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  $\mu\text{m}$  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.

## 4.6 $\cancel{E}_T$ 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. 4.8). Since every object, independent from its  $\eta$ , 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.

$$\cancel{E}_T = \left| - \sum_{\text{leptons, jets}} \vec{p}_T \right| \quad (4.8)$$

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  $\cancel{E}_T$  only objects surviving the selection are considered.

$$S_T = \left( \sum_{\text{leptons, jets}} |\vec{p}_T| \right) + \cancel{E}_T \quad (4.9)$$

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  $S_T^{\text{lep}}$  (see Eq. 4.10).

$$S_T^{\text{lep}} = \left( \sum_{\text{leptons}} |\vec{p}_T| \right) + \cancel{E}_T \quad (4.10)$$

# 5 | Data and Monte-Carlo Simulation

In almost every particle physics analysis data is compared to simulations. Physicists have to rely on simulations because it is not possible to purely calculate the outcome one sees in the detector. In this section the used data set is described as well as the production of a Monte-Carlo Simulation of a given process.

## 5.1 Data

This Thesis analyses data recorded by the CMS detector in 2016 at a centre-of-mass energy of 13 TeV. The size of the data set corresponds to an integrated luminosity of ???.

## 5.2 Event Generator and Parton Shower

## 5.3 Monte-Carlo samples

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 processes are  $W+jets$  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		
$t\bar{t}$	$700 < M_{t\bar{t}} < 1000$	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.

# 6 | Analysis

This section will cover the analysis performed for this thesis. A basic idea of the goals and strategy of this analysis is given in section 6.1. Previous results, this analysis refers to, are presented in the following. Then a detailed look into event selections, studies on particle level, differences between jet algorithms and finally the unfolding process and its results follows in sections 6.3 to ??.

## 6.1 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 6.4. 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 [11] software package. The goal is to compare data unfolded to particle level with first principle calculations.

## 6.2 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  $19.7 \text{ fb}^{-1}$  [12].

## 6.3 Jet Studies on Particle Level

For this analysis it is crucial to choose a suitable jet algorithm and cone size. The previous mentioned analysis [12] 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  $M_{\text{jet}}$  is sensitive to the top quark mass  $M_{\text{top}}$ .

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 6.3.1.

### 6.3.1 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:

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

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 6.3.2).

### 6.3.2 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 [13]. 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.

### 6.3.3 Comparing Jet Algorithms

## 6.4 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 6.4.1). Secondly, the final phase space is defined (see section 6.4.2) 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.

### 6.4.1 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
- ...

### 6.4.2 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.

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

## 6.5 Unfolding

## 6.6 Results

## **7 | Summary and Outlook**

# Bibliography

- [1] Wikimedia Commons, *The standard model of elementary particles*, [https://commons.wikimedia.org/wiki/File:Standard\\_Model\\_of\\_Elementary\\_Particles.svg](https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg), [Online; accessed April 28, 2017].
- [2] C. Patrignani *et al.*, “Review of particle physics”, *Chin. Phys.*, vol. C40, no. 10, 2016. DOI: 10.1088/1674-1137/40/10/100001.
- [3] J. Ellis, “Tikz-feynman: feynman diagrams with tikz”, 2016. DOI: 10.1016/j.cpc.2016.08.019. eprint: arXiv:1601.05437.
- [4] The ATLAS, CDF, CMS, and D Collaborations, *First combination of tevatron and lhc measurements of the top-quark mass*, 2014. eprint: arXiv:1403.4427.
- [5] F. Marcastel, “Cern’s accelerator complex. la cha ne des acc l rateurs du cern”, 2013, OPEN-PHO-CHART-2013-001. [Online]. Available: <https://cds.cern.ch/record/1621583>.
- [6] D. Barney, *Cms slice*, [https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/RetrieveFile?docid=5581&filename=PictureforPoint5\\_oct04\\_allp.jpg&version=1](https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/RetrieveFile?docid=5581&filename=PictureforPoint5_oct04_allp.jpg&version=1), [Online; accessed January 30, 2017].
- [7] F. Beaudette, “The cms particle flow algorithm”, 2014. eprint: arXiv:1401.8155.
- [8] M. Cacciari, G. P. Salam, and G. Soyez, “The anti-kt jet clustering algorithm”, 2008. DOI: 10.1088/1126-6708/2008/04/063. eprint: arXiv:0802.1189.
- [9] T. Lapsien, R. Kogler, and J. Haller, “A new tagger for hadronically decaying heavy particles at the lhc”, 2016. DOI: 10.1140/epjc/s10052-016-4443-8. eprint: arXiv:1606.04961.
- [10] 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. DOI: 10.1007/JHEP11(2015)072. eprint: arXiv:1508.01516.
- [11] S. Schmitt, “Tunfold: An algorithm for correcting migration effects in high energy physics”, 2012. DOI: 10.1088/1748-0221/7/10/T10003. eprint: arXiv:1205.6201.
- [12] CMS Collaboration, *Measurement of the jet mass in highly boosted t-tbar events from pp collisions at sqrt(s) = 8 tev*, 2017. eprint: arXiv:1703.06330.

- [13] J. Thaler and T. F. Wilkason, “Resolving boosted jets with xcone”, 2015. doi: 10.1007/JHEP12(2015)051. eprint: arXiv:1508.01518.



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**