Contents

Chapter 1

The standard model and top quark physics

High energy physics (HEP) is the branch of physics in which the interactions between the fundamental particles are being investigated. For these studies physicists collide particles at energy scales of several TeV. To produce these collisions large particle accelerators are constructed, of which the Large Hadron Collider (LHC) is currently the most powerful example. Among the physics program at the LHC one can find the study of the standard model of particle physics.

1.1 The standard model

The standard model (SM) of particle physics aims to describe nature at the most fundamental level. A full theoretical description of the standard model will not be presented in this document since this is not in the scope of this study, but can be found in References [1, 2, 3]. At this moment physicists are able to describe and probe nature up to a depth of 10^{-18} m, which is for now considered the elementary scale at which all particles in the standard model behave as point-like entities. Given this small scale, to study the phenomena of these interactions one will need to conduct research at high energies.

With the aim of describing the interactions between the fundamental particles which lie at the origin of all processes in nature, physicists came up with a model consisting of a set of particles and physics laws for interactions between them. A summary of the particles in the standard model can be found in Figure 1.1. The standard model particles can be divided into two types: fermions and bosons. Fermions are the fundamental building blocks of all known matter. Bosons on the other hand are responsible for the interaction between these matter particles, with the exception of the Brout-Englert-Higgs boson providing mass to the standard

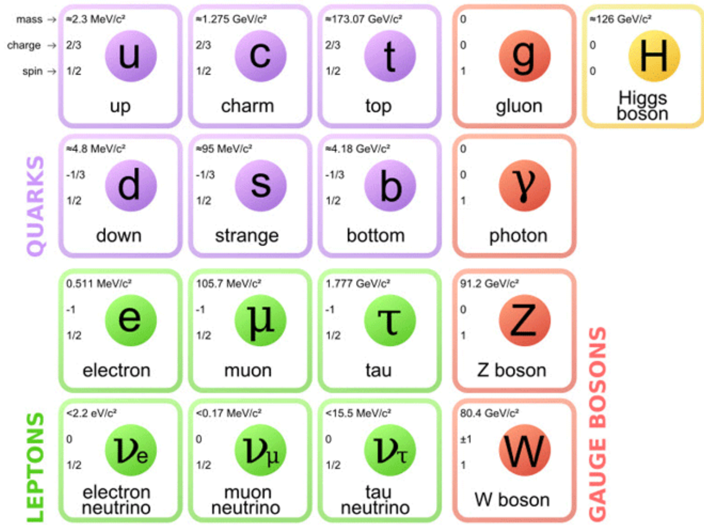

Figure 1.1: The particles in the Standard Model [4]

model particles. Although the standard model has been tested extensively over the past years, it is not yet complete since it is not fully able to describe all phenomena in nature.

1.1.1 Fermions

These fermions are the three columns on the left shown in Figure 1.1. All these particles are characterised by their half-integer spin ($\pm\frac{1}{2}$) and build up all matter in the visible universe. The fermions come in three generations. The particles in the 2nd and 3rd generation are identical copies of those in the first generation except for their rest mass. Within the fermions there are two families of particles, the quarks and the leptons.

Leptons are shown in the two bottom lines in Figure 1.1. Among these particles the most familiar are the electron (e^-) and electron-neutrino (ν_e). The electron is the lightest charged lepton, followed by the muon (μ) and tau (τ). These three particles all have an integer electric charge of -1. The neutrinos also come in three generations (ν_e, ν_μ, ν_τ), but in contrast to the charged leptons the lepton-neutrinos are electrically neutral and do not have a mass in the current formulation of the standard model. The lack of mass and charge results in a set of particles that interacts very little with the remainder of the standard model particles. However it has been observed that neutrinos have a very small mass and they do not necessarily follow the same mass hierarchy as the charged leptons.

The top two rows in Figure 1.1 are known as the quarks and are the particles which construct the hadronic matter. Also quarks come in three different generations and each generation consists of an up and down type quark which differ in electric charge. The positively charged particles are the up (u), charm (c) and top (t) quarks and carry a charge of $+\frac{2}{3}$. The negatively charged flavours are the down (d), strange (s) and bottom (b) quarks, with an electric charge of $-\frac{1}{3}$. Although there are six distinct particles in the quark family one can not observe these quarks as free particles. Due to the theory of the strong force and confinement these particles will always bind together in colour neutral configurations. This binding is mediated through gluons (g), the particle carrying the strong force.
Particles interacting through the strong force carry a so-called colour charge for which there are three possibilities: red, green and blue. To make a neutral colour configuration two kinds of combinations are possible. Either the quark binds with an anti-particle which then carries the corresponding anti-colour to form mesons. Otherwise three quarks with a different colour charge bind together to form a colour-neutral baryon. These configurations build up all known matter such as the nuclei of the atoms.

1.1.2 Bosons

To support the interactions between fermions, nature uses particles with spin 1 (bosons). For each interaction a different set of bosons is responsible for carrying the force. The electromagnetic force, probably the most commonly known interaction, is carried by photons (γ). The photon is massless. At a certain interaction energy ($\sim$ 90 GeV) the electromagnetic force unifies with the weak force. The weak force is carried by two different bosons. The first particle is the W boson, with an electric charge of ± 1. The W boson has a mass of 80.385 ± 0.015 GeV [5]. The second boson responsible for carrying the weak interaction is the Z boson, with a mass of 91.1876 ± 0.0021 GeV [5] and is electrically neutral. Therefore when collisions happen at high enough energy to produce these Z bosons, the weak and (EM) interactions will become indistinguishable.
The third interaction described by the standard model at elementary particle level is sustained by the so-called gluons. A set of 8 gluons is responsible for carrying the strong interaction which is restricted to particles carrying colour charge (quarks and gluons). The strong interaction is not (yet) unified with the other forces at energy scales currently reached at colliders. In the scope of interaction energies exceeding the TeV scale, physicists hope to get an idea of the behaviour and possible unification of the forces beyond the current limits of the standard model.
A last boson, slightly different from the other interactions, is the Brout-Englert-Higgs particle or more generally the scalar boson. This spin-0 particle with a cur-

rent measured mass of approximately 125.9 ± 0.4 GeV [5] is responsible for the electroweak symmetry breaking. This breaking and corresponding particle was first described by Brout, Englert and Higgs about 50 years ago and allows elementary particles to obtain their rest mass. The theoretical description of the particle and its recent discovery at CERN [6, 7] has lead to the Physics Noble prize 2013 for the surviving inventors of the electroweak symmetry breaking mechanism and the related scalar boson, Englert and Higgs.

1.1.3 Antimatter

In the SM every matter particle (fermion) has an antiparticle. These anti-particles construct the so-called antimatter but in the case of antiquarks they also bind with matter quarks to construct mesons. The antifermions ($\bar{f}$) differ from the matter fermions (f) only by the sign of their electric charge. A down (d) quark with an electric charge of -$\frac{1}{3}$, has a corresponding antidown quark ($\bar{d}$) which has an electric charge of +$\frac{1}{3}$. However since neutrinos do not carry electric charge their anti-particles are somewhat differently characterised. For a full description of how to describe particles and anti-particles in the standard model one can consult Reference [2].

1.1.4 Open questions in the standard model

Although the standard model itself provides an elegant description of nature at the fundamental level and has so far been tested with success, it is not yet able to describe all of natures phenomena. One of the remaining questions in the standard model is the lack of a coherent description of all forces including gravity. At this moment only approximately 4.9 % of the universe is built up by known matter [8]. The remaining 95.1 % is divided in dark matter (26.8 %) and dark energy (84.5 %). Dark energy cannot be probed in any way. Dark matter however does interact through gravitation and can thus be detected and roughly measured. One of the physics topics at the LHC is the search for a dark matter candidate particle. One of the most popular models which describes physics beyond the standard model is supersymmetry [9]. Current searches for new physics largely aim on discovering (hints of) supersymmetric particles, although non have been successful this far.
Another field the searches at LHC cover is the study of Quantum Chromodynamics (QCD). This theory describes the strong force but is not perturbative for low energies. Therefore the studies performed on QCD phenomena use physics models which are tested by comparing the measured data from the experiment to the output of simulations of these physics models.

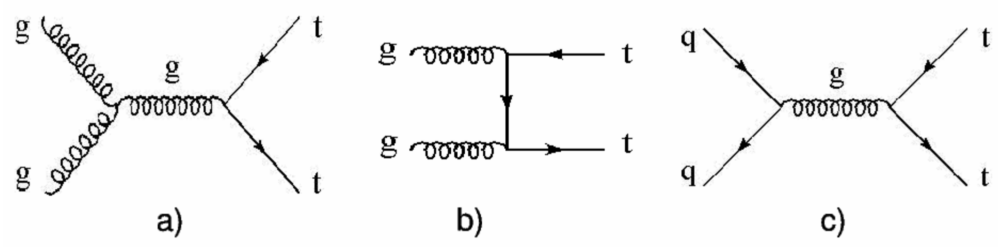

Figure 1.2: Top quark pair production mechanisms at the LHC. From left to right: gluon fusion (gg), two gluons exchanging a top quark and $q\bar{q}$ annihilation [16]

1.2 Top quark physics

The heaviest of the quarks and currently the heaviest known particle in the SM is the top quark (t), which was first discovered in 1995 [10, 11] by the CDF [12] and DØ [13] experiments at the Tevatron collider [14]. With a mass of 173.20 ± 0.87 GeV [8] measured to 0.6 % level accuracy it is by far the most accurately measured quark mass [15]. Top quark processes are an important background for many searches at the LHC. With the interaction energies at the LHC the quark events are produced abundantly. Therefore detailed studies of top quark physics are possible.

1.2.1 Top quark production

The top quark can be produced in two ways at the LHC, either in pairs ($t\bar{t}$) through the strong interaction or singly through the electroweak interaction. Single top quark events can be produced through three production processes: the single top t-channel, s-channel and W-associated channel. In this study we focus on $t\bar{t}$ events and refer to these as signal events. The single top quark processes and all processes which result in similar final states as the signal events, will be called background.

When using proton (pp) collisions three possibilities can occur in the collisions. Since a proton consists of quarks and gluons either a quark pair, gluon pair or quark and a gluon interact. As shown in Figure 1.2, there are three possible diagrams to produce top quark pair events. The top quark pairs are either produced through $q\bar{q}$ annihilation, gluon fusion (gg) or by two gluons exchanging a top quark.

1.2.2 Top quark decay

Top quarks will almost exclusively decay into a W boson and b quark. Due to the short lifetime of the top quark, this decay takes place before the hadronization process starts. The b quark from the decayed top quark will give rise to a b-jet due to the hadronization. Hadronization results in a shower of particles in the direction of the original quark or gluon produced in the interaction or decay of particles.
The W boson can decay in two ways. The first one happening in about $\frac{2}{3}$ of the decays is the hadronic decay where the boson decays to a $q\bar{q}$ pair. Each of these quarks will undergo the hadronization process and give rise to a jet. Secondly, the W boson may decay to leptons in about $\frac{1}{3}$ of the decays, giving rise to a lepton neutrino pair ($l\nu_l$). Since top quark pair events have both a top and anti top quark, two b-jets and two W bosons will be created. Therefore three different decay channels are possible, depending on the subsequent decay of the W boson. When both W bosons decay hadronically the final state will contain six jets of which two b-jets and four non-b jets. A leptonic decay of the two W bosons will result in two leptons and two neutrinos. The third possibility is the combination of a W boson decaying hadronically and the other decaying to leptons, this final state is called semi-leptonic. Here two b-jets, a charged lepton, two non-b jets and a neutrino determine the final state.
The neutrinos in the fully leptonic and semi leptonic decays will escape detection in an experiment since they only interact very weakly. However by calculating the missing transverse energy, which will be introduced properly a bit further in this thesis, one can infer whether or not neutrinos were present in the event. The leptons from the decay of the W boson are required to be isolated. This means that in a certain area around the considered lepton, no other particles are found. The probability to obtain $t\bar{t}$ events resulting in a final state with a charged lepton, a neutrino and four jets is about $2 \times \frac{2}{3} \times \frac{1}{3} = \frac{4}{9} \approx 44$ %. If we focus however on a charged lepton of a certain flavour, this reduces to $\frac{4}{9} \times \frac{1}{3} = \frac{4}{27}$ or about ≈ 15 %.

Chapter 2

Experimental setup

The Large Electron Positron collider (LEP) reached maximum energy around 210 GeV. Although the experiments at LEP provided profound insights in physics at the GeV scale, a number of questions remained unanswered. One of the most important unanswered questions was the fact that no candidate boson for the electroweak symmetry breaking was found. To this end CERN decided to build a new particle collider in the tunnel where the LEP collider was located. As a consequence in 1994 the construction of the Large Hadron Collider (LHC) was approved [17]. This collider would accelerate two proton beams up to an unprecedented energy of 7 TeV. At its design ideas the LHC would reach an instantaneous luminosity of $10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$, providing physicists with a enormous amount of data for studies of physics phenomena at the TeV scale.
The choice for a hadron collider was related with the problem of synchrotron radiation. This phenomena appears for all charged particles accelerated in a circular path. Due to the angular acceleration the particles will loose a certain amount of their energy through synchrotron radiation. This amount of energy loss has an inverse proportionality with the mass of the particle to the fourth power. Therefore a proton collider was favoured over an electron collider.

2.1 The Large Hadron Collider

The LHC is located near Geneva below the border between France and Switzerland nearly 100 m underground and has a length of approximately 26.8 km. The LHC accelerates beams of protons and will be able to collide these bunches at a centre of mass energy of 14 TeV. Although the LHC is designed as a hadron collider the LHC is also able to provide heavy ion collisions at energies of 2.75 TeV per nucleon and a design luminosity of 10^{27} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The hadron collisions allow for studies of the standard model at the TeV scale and searches for new

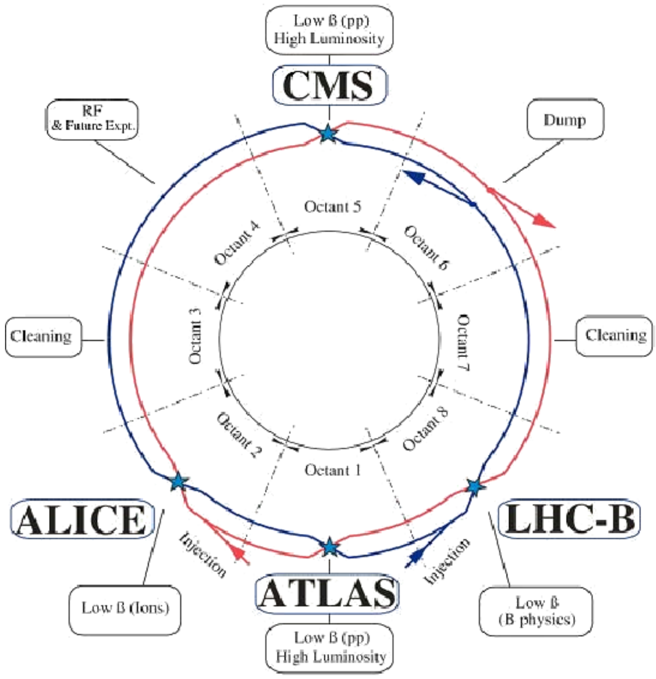

Figure 2.1: Model of the LHC structure with the four main experiments (ALICE, ATLAS, CMS and LHCb) indicated on the circumference. The two beampipes are indicated by the red and blue circles, intersecting in the detectors [17]

physics phenomena, while the heavy ion collisions aim to provide new insights in the physics of matter at extremely high densities. To record the collision data, the LHC needs detectors surrounding the interaction points (IP) of the bunch crossings. There are four main detectors installed on the circumference of the LHC, which are illustrated in Figure 2.1. Two of these detectors are multipurpose detectors which are designed to investigate for a broad range of physics phenomena. One of these two experiments is the ATLAS experiment, [18] and is by far the largest detector at the collider, with a length of approximately 44 m and height of 25 m and weight of 7000 tonnes. The second multipurpose detector is the CMS detector, [19] which will be described in more detail in Section 2.2. The ALICE (A Large Ion Collider Experiment) detector [20] is designed for detailed studies of the heavy ion collisions. The last detector at the LHC is the LHCb experiment [21]. This experiment aims to perform high precision b physics measurements. Among the many studies performed at LHCb an example is the search for hints of the matter-antimatter asymmetry in the universe.
The bunches at the LHC are accelerated by a series of magnets. At the time of

the construction it was required to push the edge of technology in superconducting research further. With a two ring superconducting collider and limited space in the tunnel, the choice was made to use a twin-bore magnet system in the accelerator. Due to this technology the two rings are magnetically and physically coupled. At the LHC the magnets are cooled to an operational temperature of 2 K.
Although the LHC is designed to provide collisions at a centre of mass energy of 14 TeV, the current reached energy was 8 TeV (4 TeV per beam) due to the consolidation of the high current splines interconnecting the magnets. A magnet quench during test runs in 2008 damaged part of the accelerator which needed repairs before being able to start running again. The cause of the accident were most likely bad interconnections between the magnets. To avoid another accident the operational energy was kept low during the first run. At this moment the accelerator is being upgraded. The second run of the LHC starting in 2015 will aim to achieve energies of 13 to 14 TeV and a luminosity that is already twice as high as the design luminosity.

2.1.1 Accelerator chain and event characteristics

To accelerate protons up to energies of 7 TeV the collider uses an injection chain to cluster and pre-accelerate the protons. The injection chain is constructed as a series of accelerators as shown in Figure 2.2. The particles start at the LINAC 2 (linear accelerator) where from a simple container of hydrogen gas the protons are obtained by stripping the electrons from the hydrogen with the use of an electric field. The protons are then accelerated to an energy of 50 MeV. Next, the protons are injected into the proton synchrotron booster (PSB) where they are accelerated to an energy of 1.4 GeV. During this process the protons are also gathered in bunches.
The following step in the acceleration is done by the proton synchrotron (PS) and raises the energy to 25 GeV. This acceleration step is followed by the super proton synchrotron (SPS) which gives the final acceleration to 450 GeV before injecting the bunches into the LHC accelerator. The final process to get each of the LHC rings filled with bunches at energies of 4 TeV per bunch takes roughly 25 minutes [22]. Once the bunches reach their maximum energy the collisions can start.
From the beam energy and luminosity, one can estimate the bunch characteristics. When the LHC operates, a beam contains at most 2808 bunches, each containing roughly 10^{11} protons. The bunches will have a minimum spacing of 24.95 ns, which is an important parameter for the design of the detectors surrounding the interaction points. At full operational capacity bunch crossings will appear every 25 ns. During the first runs of the LHC a spacing of approximately 50 ns between the bunches was used.

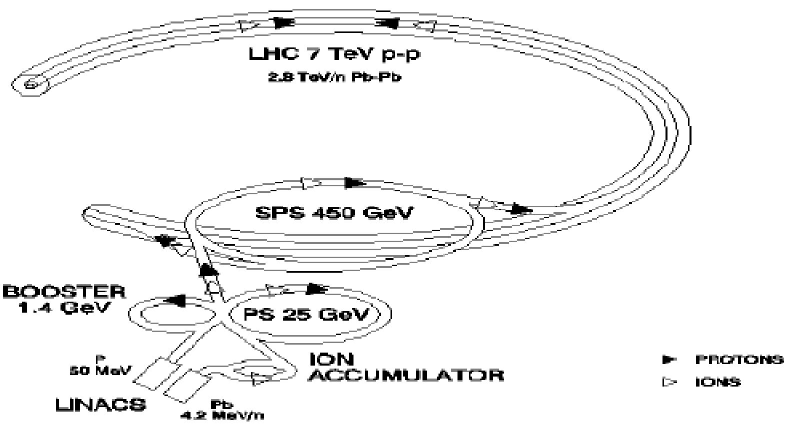

Figure 2.2: Model of the structure of LHC and the accelerator complex [17]

2.1.2 Characteristics of the proton collisions

When colliding bunches at the LHC the properties of these packages of protons are known, such as their dimensions and the approximate number of protons in each bunch. This results in the notion of instantaneous luminosity (L), which is an important parameter for the collisions and is well defined by the bunch characteristics. At the LHC the design luminosity is 10^{34} cm^{-2}s^{-1}. The integrated luminosity ($\mathcal{L}$ can be calculated by integrating L over the operation time of the LHC. With the production cross sections of $t\bar{t}$ events known one can calculate the number of events occurring at the collisions by using Equation 2.1.

$$N_{events} = \sigma \times \mathcal{L} \tag{2.1}$$

As mentioned earlier in this chapter, the choice for a hadron collider was motivated by a decrease in synchrotron radiation. However when colliding protons a lot of extra difficulties come up, because the proton is composed of a combination of quarks and gluons. Apart from the three valence quarks and gluons, the proton also contains so called sea quarks. These are quark-antiquark pairs which are continuously created and annihilated inside the proton structure due to QCD effects. When two electrons collide the interaction is fully known, since it is an interaction between two elementary particles. With protons on the other hand, one can never be certain which two particles took part in the interaction and whether there were other particles involved in the interaction. When two partons of the protons interact, the proton remnants will not be colour neutral. Hence the remnants will hadronize to form colour neutral hadrons, giving rise to the so-called underlying event. A good understanding of the underlying event is therefore needed to analyse the hard interaction.

Another factor that obscures the event of interest is due to the collision of bunches instead of just two particles. At the LHC there are on average up to 20 collisions happening simultaneously. This notion of additional collisions is called *pile up*. Pile up and underlying event pose an extra challenge on detector and analysis level for physicists to study the hard collision of the interacting partons.

2.2 The Compact Muon Solenoid

To study the phenomena occurring during the interactions at the LHC, physicists need a detector to measure the characteristics of the outgoing particles. For the study reported in this masterthesis the Compact Muon Solenoid (CMS) detector [19] is used. The CMS experiment is a multi-purpose detector and does not specifically focus on one physics topic. This has its effect on the design of the detector. Like the majority of particle detectors at accelerators, CMS measures particles by using a set of subdetectors. Each subdetector is dedicated to the measurement of specific types of particles.
A first large property that is exploited is the electric charge of the particles. For charged particles one can measure the momentum of the particle by looking at the trajectory of the particle through a well known magnetic field. The tracker records the trajectories of charged particles and is the first subdetector layer surrounding the interaction point in CMS.
Next, one needs to keep in mind that some of the particles produced during and after the collision will decay. This decay will be accompanied by a lot of electromagnetic radiation (photons and electrons). These particles can then distinctively be measured by use of an electromagnetic calorimeter (ECAL). Here the EM interacting particles are being stripped from their energy until they are completely stopped by the calorimeter. At this point one can already distinguish photons from electrons by combining the tracker and ECAL information. If a series of hits in the tracker coincides with a number of ECAL towers (as the separate sensor measurements in the ECAL are called) one can already label this particle as an EM interacting charged particle and exclude it being a photon.
The strongly interacting particles will behave differently. Due to the non-existence of free quarks, produced quarks and gluons will decay and recombine into colour stable combinations (hadrons). These hadrons are detected in the hadronic calorimeter (HCAL). One can now distinguish charged and neutral hadrons by looking at the tracker information together with the information of the HCAL. Putting also the measurements from the ECAL in the equation, one can distinguish electrons, charged and neutral hadrons and photons. The lepton neutrinos will unfortunately not be observed. These nearly massless particles will neither interact with the tracker since they don't carry electric charge, neither with the HCAL since they

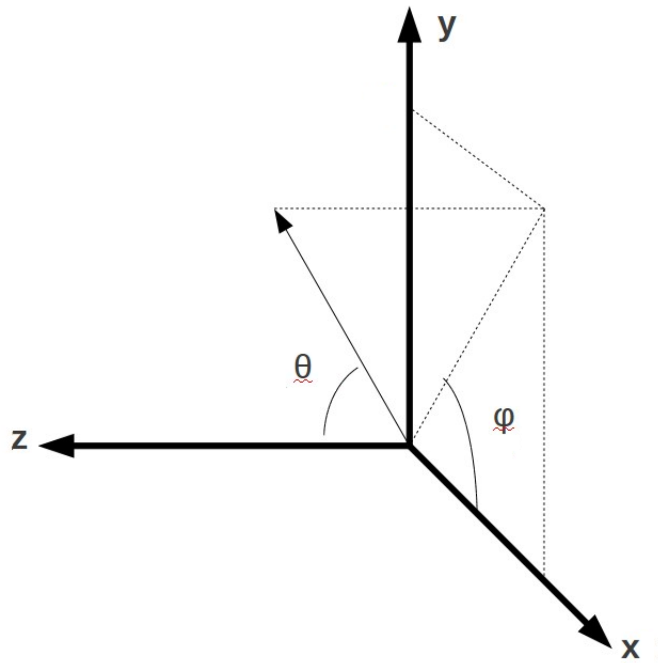

Figure 2.3: Cartesian coordinate system used in the CMS detector along with the definition of the azimuthal and polar angles, respectively ϕ and θ. The z-axis points along the beamline in the direction of the anti-clockwise beam, the y-axis points upwards to the surface and the x-axis towards the centre of the LHC ring.

have no colour charge or the ECAL since they have no EM properties.
As a last part one needs to detect muons. Muons can travel large distances before decaying and they will be the only detectable particles passing through the three previous subdetectors. Therefore muon detectors are installed surrounding the other three detectors. It is relevant to note that, since the τ lepton has a large mass, it will almost instantaneously decay into lighter particles, leading to a jet in the calorimetry system or to an electron or muon with the corresponding antineutrino.
The CMS detector is constructed according to the design outlined above. The resolutions and characteristics for the different subdetectors and sensors are motivated from the physics phenomena one wants to measure.

2.2.1 CMS coordinate system

To describe the geometry inside the CMS detector, both for detection and instrumentation and the analysis of the data, CMS uses a specific coordinate system shown in Figure 2.3. It also illustrates the definition of the azimuthal angle ϕ and polar angle θ. The orientation of the Cartesian coordinate system is as follows: the x-axis points towards the centre of the LHC ring. Together with the y-axis which is directed upwards to the surface this defines the transverse plane, labelled with subscript T. The z-axis points along the beamline in the direction of the anti-clockwise beam. The origin of the coordinate system is located in the interaction point [23]. The CMS collaboration frequently uses another set of co-

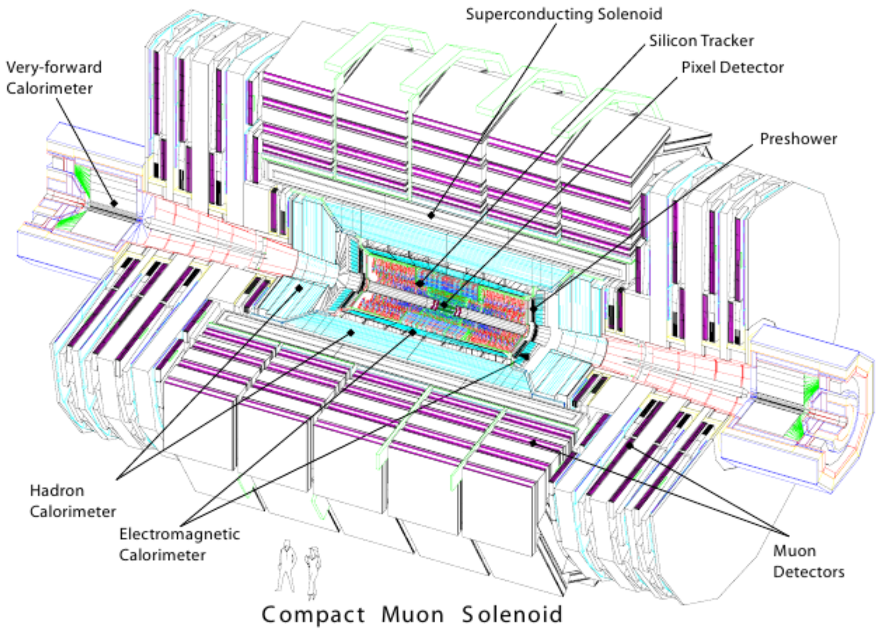

Figure 2.4: Model of the structure of the CMS detector and the different subdetectors [19]

ordinates. Two of these are the angles ϕ and θ which are defined by the Cartesian system as shown in Figure 2.3. Another coordinate which is used is the radial distance r, which is measured from the origin of the Cartesian system. A very useful coordinate commonly used is the pseudorapidity η which is defined as

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

In the transverse plane also the transverse energy E_T and transverse momentum p_T will be important. In particular for the determination of the missing transverse energy $\not{E}_T$. The transverse energy or momentum is the projection of the energy or momentum on the transverse plane, hence

$$E_T = Esin\theta$$

2.2.2 The CMS subdetectors

The CMS detector has a cylindrical shape with a diameter of 14.6 m and a length of 21 m. In total the detector has a mass of 12500 tonnes. To bend the charged particle trajectories CMS uses a superconducting solenoid, capable of generating a magnetic field of 3.8 T at an operating current of 15000 A. Except for the muon detection system all of the subdetectors are located inside this solenoid. In Figure 2.4 a schematic view of the CMS detector is shown. In the following sections, the various CMS subdetectors will be discussed in more detail.

Muon System

The CMS muon system was designed to fulfil three main functions. First of all the subdetector should provide a good muon identification. This is required because many new physics searches including the hunt for the Brout-Englert-Higgs boson rely on a good muon identification. Electrons are much more affected by radiative losses compared to muons which makes muons better suited for precise measurements. Secondly the muon system is used to trigger events. The third function aims on the momentum measurement. Since muons can travel through a large amount of material without much energy loss a muon system which can provide a good momentum measurement is a large advantage.
The structure of the CMS muon system is shown in Figure 2.5. Since CMS has a cylindrical geometry the detector can be divided in a barrel region which is itself cylindrical and two endcaps which cover the forward regions of the detector. This is also true for the muon system where these two regions are equipped with different sensors.
The barrel region contains the barrel drift tubes (DT) as sensors and covers a pseudorapidity region of $|\eta| < 1.2$. The DT are arranged in four groups of concentric circles surrounding the z-axis. Each group of DT contains three sensors for which the middle one measures the z-coordinate and the outer ones measure the coordinate in the $r - \phi$ plane with a global resolution of 100 μm. The endcaps use cathode strip chambers (CSC) for the momentum measurement of the muons. In the endcap region the trapezoidal shaped CSC detectors cover a pseudorapidity region between $1.2 < |\eta| < 2.4$. Muons traversing the endcaps will encounter 3 or 4 layers of CSCs which have either 10° or 20° of coverage in the ϕ coordinate. The muon system has an overall system of RPCs which will serve as a redundant trigger. RPCs are capable of detecting an ionising event in a much shorter time than 25 ns. Therefore this system can unambiguously identify the relevant bunch crossing of an observed muon at the LHC. Both for the barrel and endcap regions the RPCs need to meet the requirement of providing an efficient momentum measurement in the environment of particle rates up to 10^3 Hz/cm^3. Although all RPCs have the same requirements, they differ in shape and dimensions for the two distinct regions.

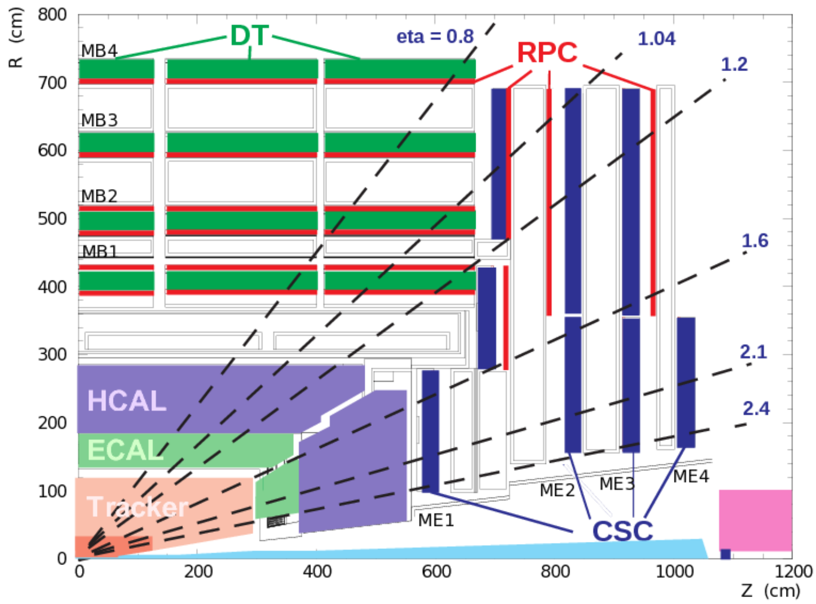

Figure 2.5: Structure of the muon system with the drift tubes, cathode strip chambers and resistive plate chambers indicated [19]

The inner tracker

In CMS the choice was made to use an all silicon tracker system. This decision was supported by a number of design challenges such as a material that is radiation hard which is needed because of the high particle flux and high occupancy. To provide precise measurements of the primary and secondary vertices the tracker consists of two parts. The first part, closest to the beamline consists of pixel sensors and is hence referred to as the pixel tracker. The pixel tracker covers an active detection area of approximately 2 m^2 covering radii from 4.4 cm to 10.2 cm.

Surrounding the pixel tracker is the silicon strip tracker (SST). Due to the lower particle flux at this distance from the interaction point strips could be used instead of pixels and therefore less service channels are required resulting in a decreased impact on the interaction of particles with non-sensitive material. The SST has 10 strip layers and covers radii between 20 cm and 1.1 m. It measures tracks of charged particles with a high granularity and precision. The tracker structure can be found in Figure 2.6 indicating the pseudorapidity coverage for $|\eta| < 2.5$. One can distinguish again the barrel and endcap modules for both the pixel and strip tracker. The pixel contains three concentric layers in the barrel region and two disks in both endcaps. For the SST the structure is a bit more complex. Ten strip layers are subdivided in two regions, the tracker inner barrel (TIB) and outer

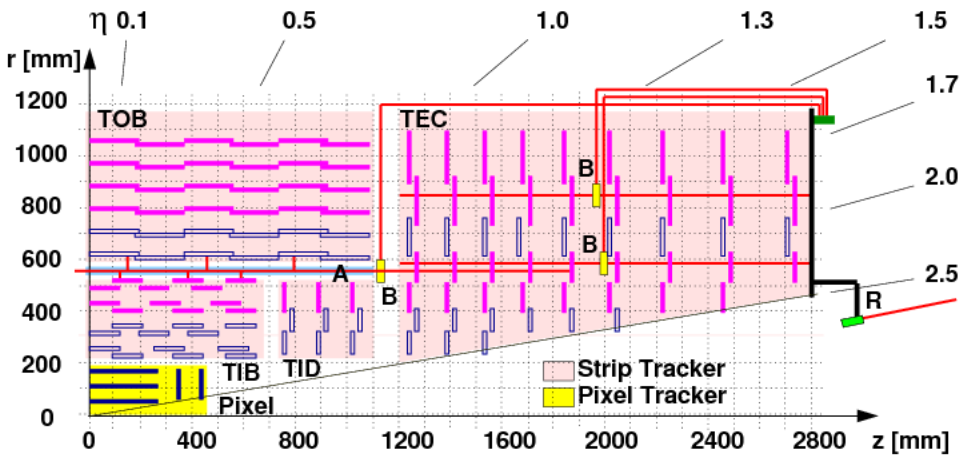

Figure 2.6: Structure of the CMS full tracker system with different module locations [19]

barrel (TOB). Modules are used to group the silicon sensors for which both shape and dimension can differ for all layers. The modules contain either double sided sensors or single sided sensors with a small relative rotation with respect to each other. By this geometry both the position in the r-ϕ plane and along the z-axis can be determined.

The SST has two disk structures. The most inner part is the tracker inner disk (TID) and has three disk layers. The outer part is the tracker endcap region (TEC) and has 9 disk layers. The two disk structures are present at both ends of the tracker detector and complete the barrel layers.

With the measurements from the tracker one can obtain the tracks of charged particles by linking the hits in the different detector layers. This linking is performed according to an algorithm which calculates the most probable trajectory for each particle and is called the Kalman filter [24].

The electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) is the first subdetector of the calorimetry system that particles encounter. The energy of particles is measured using scintillating $PbWO_4$ crystals. Particles colliding with the material of the crystals produce photons which are then gathered by avalanche photodiodes in the endcaps and phototriodes in the barrel [19]. The amount of photons gathered by the scintillators gives a measure for the energy of the particle.

The ECAL has a geometric coverage in pseudorapidity range for $|\eta| < 3.0$. As for all subdetectors the ECAL is composed of a number of different parts as illustrated

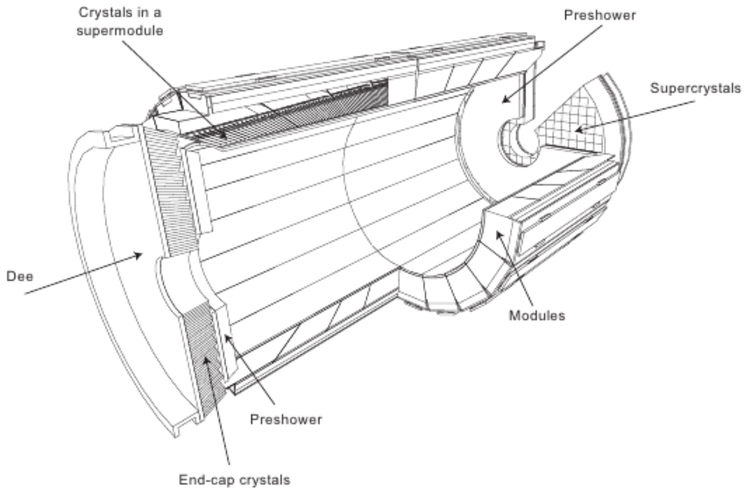

Figure 2.7: Structure of the CMS electromagnetic calorimeter [19]

on Figure 2.7. In the ECAL endcap region there is an additional preshower (PS) with the aim of detecting neutral pions in the region of $1.653 < |\eta| < 2.6$. The preshower has the working principle of a sampling calorimeter with lead radiators and silicon strip sensors. By design this calorimeter system is a homogeneous calorimeter system. Due to the chosen material it has a fast response and it is designed to be highly granular. The crystals are radiation hard and have a good energy resolution.

The hadronic calorimeter

To measure and detect hadrons CMS has installed a hadronic calorimeter (HCAL). The HCAL is located between the ECAL and the magnet and is a sampling calorimeter. In the barrel region the hadronic barrel calorimeter (HB) is located. The HB is divided into two half-barrel sections, covering a range of $|\eta| < 1.3$. The sampling is performed by brass plates and scintillation tiles. The tiles use wavelength shifting fibres to transfer the gathered light. The brass and scintillation tiles are put together in wedges. These wedges have no projective dead material. The thickness of the hadronic calorimeter can be expressed in the number of interaction lengths of particles with the material (λ_I). Since all subdetectors apart from the muon system are located inside the magnet, the width of the different layers is largely constrained by the size of the solenoid. For the HB the coverage is up to 10.6 λ_I which contains 1.1 λ_I from the ECAL.
With the HB alone the hadronic calorimeter does not cover enough radiation lengths to contain all of the hadronic showers. Therefore a tail catcher was in-

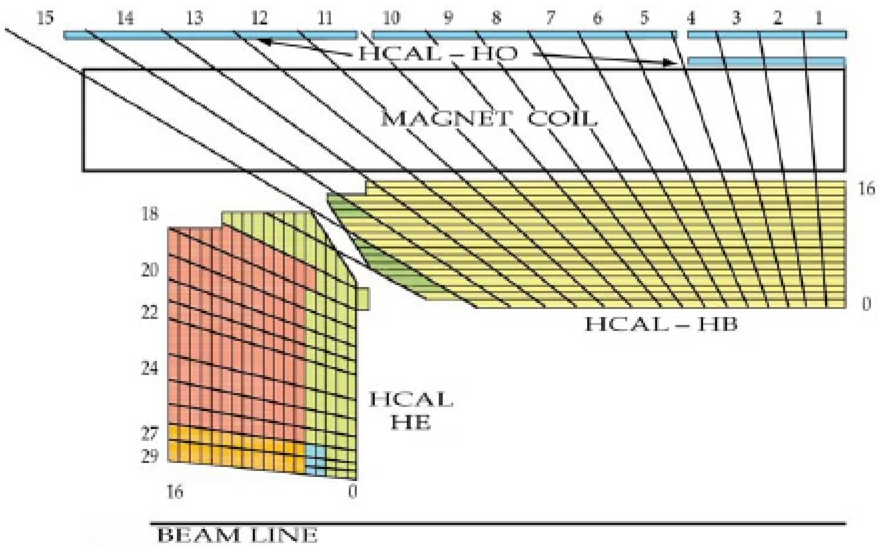

Figure 2.8: Illustration of the Hadronic calorimeter with the HB, HE and HO indicated [19]

stalled outside the HB. This outer hadronic calorimeter (HO) uses the magnetic coil as an extra absorber adding 1.4/sinθ λ_I. With addition of this HO the total calorimetry system extends up to a minimum of 11.8 λ_I. The design and location of this tail catcher is shown in Figure 2.9. It has been proven [19] that the HO recovers the effect of leakage and contributes to a better determination of the missing transverse energy. The HCAL also has two endcap regions (HE). The HE covers a range of $1.3 < |\eta| < 3$ in pseudorapidity range which contains ~ 34 % of the produced final state particles. The sensors and design of the HE meet the requirements to be radiation tolerant (10 MRad after 10 years of LHC operation [19]) and handle high counting rates. An important factor of the design of the HCAL parts and specifically for the HE is the lack of magnetic material. For example the HE lies in the region where the magnetic field is strongest ($\approx$3.8 T). Between the HB and HE an absorber is placed to minimise the cracks and gaps resulting in as little dead projective material as possible. The HE together with the ECAL covers 10 λ_I.
To close the endcap structures and extend the pseudorapidity coverage ($|\eta| = 5$) CMS has a forward calorimeter (HF) on both ends of the detector. These detectors need to be able to withstand tremendous particle fluxes. Where all layers of the CMS detector have an absorption capability of approximately 100 GeV per bunch crossing, the forward regions need to be able to tolerate approximately 760 GeV per bunch crossing. The active medium inside the HF is quartz fibre

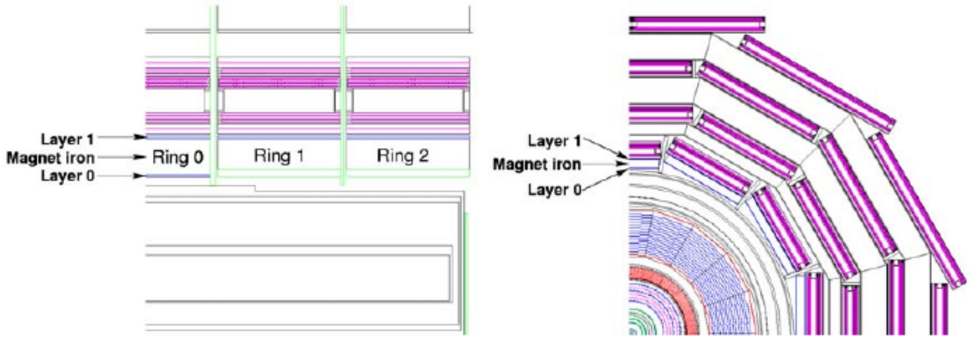

Figure 2.9: Position of the HO system in the CMS structure [19]

which is better suited for this hostile environment. The geometrical resolution of the calorimetry system can be expressed in terms of the pseudorapidity (η) and azimuthal (ϕ) coordinates, called ($\Delta\eta \times \Delta\phi$) towers. For the HB, HO and HE ($|\eta| < 1.6$) the resolutions are (0.087×0.087)), for the HE where $|\eta| \geq 1.6$ the resolution is (0.17×0.17) and for the HF (0.175×0.175).

The superconducting solenoid

As suggested by its name the CMS detector uses a solenoid magnet to produce the magnetic field inside the detector to bend the charged particle trajectories. The magnet is designed to produce a magnetic field of 4 T. To produce such a homogeneous magnetic field with a large solenoid, an at the edge technology was used. The windings inside the solenoid run in four separate layers while up to the moment of production of the magnet only one or two layers were commonly used. The windings consist of NbTi superconducting wires. This material can withstand the current of 15 000 A needed to produce the magnetic field. The magnet itself has a cold mass of 200 tonnes and has dimensions of 12.5 m of length and a free bore diameter of 6 m. With such large dimensions and capable of producing an extremely strong magnetic field, the magnet needed to be designed to endure large mechanical deformations when powering up. To close the magnetic field lines a system of iron return yokes was installed. Between these blocks of iron the muon system is interspersed. The total mass of the system of iron return yokes is 10 000 tonnes which is almost half of the full weight of the CMS detector.

The CMS trigger system

Due to the high number of bunch crossings the amount of data taken by the detector easily becomes too large. To reduce the amount of data, a trigger system is used, selecting only the most interesting events. In CMS the choice is made to use a two level trigger system with a first trigger, the Level 1 (L1) trigger, being a hardware trigger. The second trigger, the high level trigger (HLT) is a software-based trigger and together with the L1 reduces the stream of data from 40 MHz to 100 Hz.

The L1 trigger uses the input of the calorimeter and muons systems as coarsely segmented data. The remainder of the high-resolution data is stored temporarily in electronic pipelines. These pipelines will hold the data for 3.2 μs while the event is being analysed with the L1 trigger. The input is used both separately as well as combined (global). After the 3.2 μs a decision is made whether to keep or reject the data of the event. The L1 trigger reduces the event rate from 40 MHz to 100 kHz. Within the HLT the information of all subdetectors is used to reconstruct and select the event. This offline reconstruction is performed on a processor farm. The algorithms used for the HLT evolve in time. For this reason the system should be easily accessible for maintenance and updates. The HLT provides a further reduction with a factor of 10^3, such that the final amount of data is approximately 100 Hz. For more information on the trigger system one can consult reference [25].

Chapter 3

Particle reconstruction

To reconstruct the event after recording the collisions, the signals from all subdetectors are used to determine the particles produced in the collision. Due to the large particle flux, hadronization processes, the interactions with the detector material, detector inefficiencies and the amount of output signals this is an unwieldy job. To provide an accurate particle reconstruction, CMS uses the particle flow (PF) method.
For almost all of the physics studies performed at the CMS experiment, the measured data is compared to physics models. These physics models are represented by Monte Carlo (MC) simulations of the various physics processes. Some underlying models and the generation of their events at CMS are introduced in Chapter 4. The output of the event generators is similar to the particles which are produced during the hard interaction. In the CMS detector, the produced particles are physically traversing the detector. For the simulated events, large efforts have been made to simulate the interactions of the particle with the detector. In this way an equivalent and realistic detector output is obtained for the simulated events.

3.1 Particle properties and four momentum

CMS uses a number of reconstructed objects. QCD confinement makes quarks and gluons undetectable as free particles. Therefore the quarks and gluons are reconstructed as jets. The charged leptons, in particular electrons and muons, can more easily be identified. Neutral particles are distinguished by the absence of a track in the CMS tracker. Neutrinos escape detection, but their presence can be inferred from the transverse energy balance. For each of these particles one attempts to reconstruct the four momentum (p) which contains the energy of the particle (E) and the three-momentum ($\vec{p}$) as a four dimensional vector (E, $\vec{p}$). This four vector contains all information of the motion and it allows to calculate the

reconstructed particle mass.
The missing transverse energy ($\not{E}_T$) is obtained from the four momenta of all reconstructed particles. If the vectorial sum of the transverse momenta of all particles is different from zero, a particle was produced which has not been detected. This follows from the principle of momentum conservation in the transverse plane.

3.2 Particle flow

The aim of the particle flow method is to reconstruct the physics events which occur at the LHC proton collisions. In the end the reconstructed objects represent the original particles produced in the collisions.

3.2.1 Linking elements

For the reconstruction all subdetectors are used both individually and globally. Starting at the tracker, the signal comes from the individual tracker cells. Three of these cells are sufficient to reconstruct the trajectory of a charged particle. Particles with a transverse momentum as low as 150 MeV can be distinguished and measured due to the efficient tracker. For the calorimetry system the individual sensors are represented as towers. Such a tower indicates the measurement in the corresponding cell, where the height of the tower in Figure 3.1 indicates the amount of energy deposited by the corresponding particle.
The reconstruction of the final state particles uses the information from the different subdetectors, the charged particle tracks from the tracker, the clusters from the calorimetry system and muon tracks from the muon system.

Iterative tracking The momentum measurements from the tracker are far more precise than the measurements from the calorimeters. In the tracker, the direction of the particles and position of the vertices is also measured without any deviation caused by the interaction with the detector material. Since two third of the jet energy is carried by charged particles [26] it is important to reconstruct the charged particle tracks accurately. This is done by the part of the iterative tracking in the particle flow method. The tracking efficiency of the adopted iterative tracking is 90 % for charged hadrons in jets and up to 99.5 % for muons. The tracking efficiency is required to be as close to 100 % for charged particles. This requirement is due to the reduced resolution on the energy and a biased direction determination for reconstructions of charged particles which only have calorimeter information. The iterative tracking is performed in several levels. First, particle seeds are identified. Seeds are measurements that are reliable starting points for the reconstruction of a track in the tracker or a cluster in the calorimeter. First,

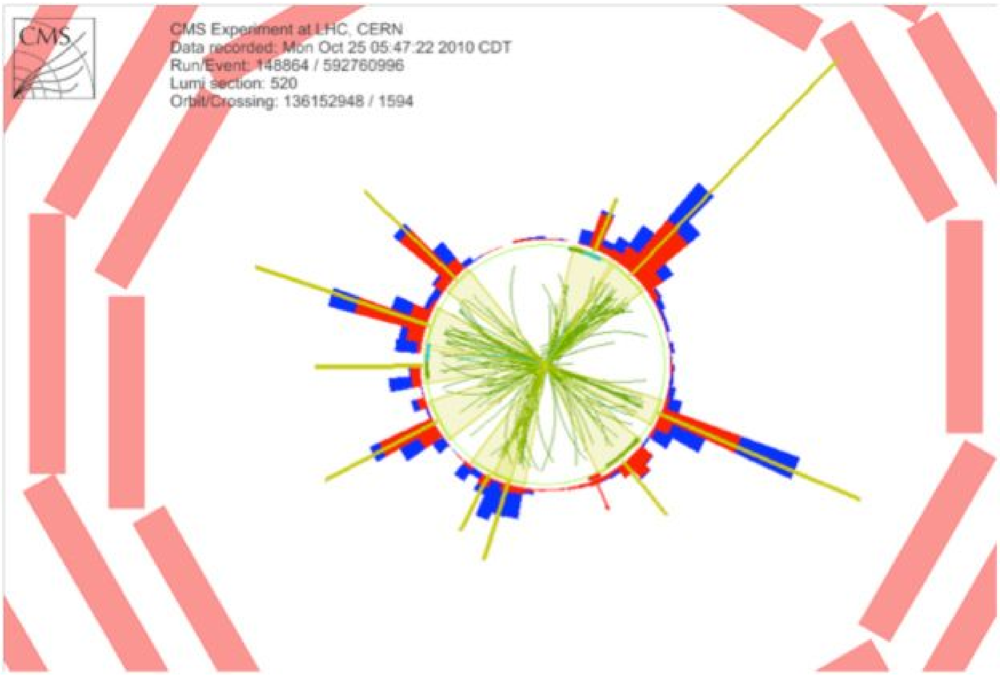

Figure 3.1: Example of a recorded data event display at the CMS detector with high jet multiplicity [27]. ECAL (red) and HCAL (blue) towers are represented as rectangles with their height representing the particles energy deposits and tracks shown as the green lines. The jets are represented as yellow cones and lines.

tight requirements are used to reconstruct tracks with a very small fake rate. Afterwards the hits assigned to these reconstructed tracks are removed. Next, the criteria on the seeds are loosened which increases the tracking efficiency. Due to the removal of the used hits, the fake rate remains low. The same procedure is used to reconstruct new tracks and the used hits are again removed where-after the criteria are loosened. Finally, the fourth and fifth iterations aim to reconstruct secondary charged particles from photon conversions and interactions between the particles and the material inside the detector. Therefore relaxed constraints on the determination of the secondary vertices are adopted.

Calorimeter clustering The second part in the particle flow method is the calorimeter clustering. Calorimeters serve a number of purposes. A first one is to measure and detect all stable neutral particles. The measurements include the direction and energy determination. Another aim is to distinct and separate neutral particles from charged ones. For the charged particles the goal is first to identify electrons and assign the Brehmsstrahlung measurements to the accompanying photon. A last part is to improve the momentum determination from the charged particles for which the tracker could not provide an accurate measurement. In Figure 3.1 an illustration of particles traversing the CMS detector is shown, along with the clustered jets (yellow cones and lines).

3.2.2 The particle flow algorithm

The most efficient reconstruction method links different subdetectors and avoids double counting of particles. The linking algorithm connects pairs of elements in the detector. The connection is defined in terms of a linking distance indicating the quality of the connection. Tracks of charged particles are linked to the signals in the preshower (PS) by looking at the signals in the most outer layers of the tracker and the first hits in the PS. Next a match for the extrapolation to the ECAL and sequentially to the HCAL is searched for. Finally a linking to a track in the muon system is performed. Each link is then stored by the algorithm as a “block”. The identification and reconstruction of the final state particles is performed by the particle flow algorithm. The list of reconstructed final state particles makes up the full event, available for analysis. The input for the particle flow method are the blocks resulting from the link algorithm, while the output are the objects reconstructed as discussed below.

Muon identification The algorithm starts by searching for particle flow muons. If the combined momentum of the particle measured in the muon system agrees within three standard deviations to the momentum of the tracker measurement, the muon is stored as a Particle Flow (PF) muon and the track is removed from the block.

Electron identification After the muons the algorithm aims to identify all electrons. Electrons will lose energy by Brehmsstrahlung. This behaviour is exploited in the PF algorithm when selecting the blocks in the tracker. These pre-identified electron tracks are then matched to the ECAL blocks and tested against a combination of tracker and ECAL variables that are able to discriminate between electrons and other particles. If a combination of blocks in the tracker and ECAL is identified as an electron, they are also removed.

Hadron identification For the remainder of the tracks a selection is made with respect to the calorimeter resolutions, for which it is required that the relative uncertainty on the measured p_T is smaller than the relative energy resolution of the calorimeter for charged hadrons. The rejected tracks from hadronic jets are however not lost. About 10 % of these tracks will be measured more precisely by using the calorimeter.
When connecting the tracks to the energy deposits in the calorimeters, the PF algorithm compares to the ECAL or HCAL clusters. These clusters can contain additional neutral particles next to the charged ones. The comparison between the momentum measurement of the track and the calorimeter cluster needs to undergo a calibration procedure [26]. This is needed due to the possible connection between one track and several ECAL and HCAL clusters. After the calibration the track momentum measurement and the calibrated calorimetric energy are compared. If there is a large difference, a relaxed search is performed where the algorithm searches for muons as additional particle flow muons. Otherwise the search is relaxed to searches for fake tracks.

Photons and neutral hadrons With the remaining blocks, a number of possibilities is still open. If the tracks correspond in momentum measurements to the calorimeter clusters, the tracks in the block give rise to particle-flow charged hadrons. If there is an energy excess with respect to the calorimeter resolution this can indicate the presence of either a particle-flow photon or particle-flow neutral hadron. At this point the remaining calorimeter clusters are identified as particle-flow photons if they are in the ECAL or particle-flow neutral hadrons if they are also in the HCAL.

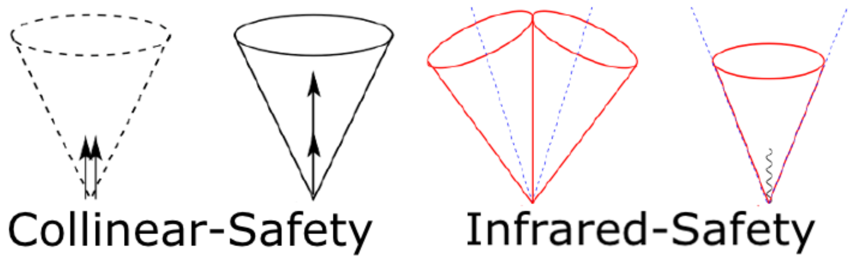

Figure 3.2: Illustration of the requirements for jet algorithms [28].

3.3 Jet clustering algorithms

3.3.1 The anti-k_t jet clustering algorithm

Jet clustering algorithms are designed to collect the observed hadrons into jets. To assign the final state particles to the appropriate jet and guarantee a proper reconstruction, a couple of requirements need to be met. Given the high rate at the LHC the algorithm is required to be fast. Another important property is the collinear safety of the algorithm. This property is required to handle the problems arising when two particles have the same direction within one jet. These two particles should not give rise to two distinct jets, which means the output of a jet algorithm should not change if the energy of a particle is distributed among two distinct collinear particles, as shown in Figure 3.2.
As a third requirement, the reconstruction has to be robust against the appearance of ghosts and soft particles in a jet. When a soft (low energy) particle is added to the jet, the output should not be altered, for any number of soft particles. This is called infrared (IR) safety and also counts for soft emissions, where a gluon is radiated from one of the jets as indicated on the right hand side of Figure 3.2.
The anti-k_t jet clustering algorithm used by default in the CMS collaboration is a fast, IR and collinear safe jet clustering algorithm. The anti-k_t jet clustering algorithm uses sequential recombination like other jet algorithms such as the k_t and Cambridge/Aachen algorithms [29]. This in contrast to the cone algorithms like the SIS cone algorithm. To do the clustering, a definition of the distance between the PF particles is introduced. Two distances are defined, on the one hand the distance between a particle and the beamline, defined in Equation 3.1, on the other hand the distance between two particles as defined in Equation 3.2.

$$d_{iB} = k_{ti}^{2p} \tag{3.1}$$

$$d_{ij} = min\left(k_{ti}^{2p}, k_{tj}^{2p}\right)\frac{\Delta_{ij}^2}{R^2} \tag{3.2}$$

$$\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 \tag{3.3}$$

Equation 3.2 explicitly relies on the distance Δ_{ij}^2 between the particles i and j, for which the definition is shown in Equation 3.3. The distance d_{ij} depends on the recombination parameter R. In the CMS the parameter R is 0.5 by default. In the above equations the parameters k_{ti}, y_i and ϕ_i represent respectively the transverse momentum of particle i, its rapidity and azimuth. The parameter p is introduced to scale the relative power of the energy with respect to the geometry Δ_{ij}.

Sequential recombination algorithms quantify the clustering of particles based on this distance. The clustering starts from a seed particle and calculates all distances between this seed and the surrounding particles. If the smallest distance is d_{ij} the particles are clustered, else if the smallest distance is d_{iB} the combined particles are clustered in a jet. Once a jet is constructed the involved particles are removed from the list. This procedure continues until no particles remain for the clustering.

For the anti-k_t algorithm, the parameter p is set to -1. As a consequence the distance between a hard particle and soft particle will be determined by inverse of the transverse momentum squared of the hard particle and the distance between them. Between two soft particles the distance d_{ij} will become very large due to the low transverse momenta, resulting in a clustering of soft particles to the hard particles instead of soft particles among each other.

The performance of different jet algorithms can be compared with respect to a number of jet characteristics. Reference [30] presents an in depth discussion of the performance of the anti-k_t algorithm with respect to some other jet clustering algorithms.

3.3.2 The FastJet package

The FastJet package provides tools for fast and efficient jet finding and analysis [31]. For an elaborate description of the usage and different features of the FastJet algorithm one should consult reference [31]. The main features used for the analysis described in this thesis are briefly discussed in the following.

The FastJet package is used to recluster the jet constituents of the original jets on the fly using the anti-k_t jet clustering algorithm with an R parameter of 0.5.

As an input for the algorithm, the four-momentum of particles in the jets selected in the analysis are used. An important remark is that the jet energy of the jets has been corrected to take into account the non-uniform and non-linear detector response and contributions of the PU to the total jet energy. When reclustering

constituents into jets with the FastJet package, these Jet Energy Corrections (JEC) are however not taken into account.
The effect of the jet energy corrections can be seen in Figure 3.3 where difference in transverse momentum (p_T) is clearly visible for the original jets and the reclustered jets.

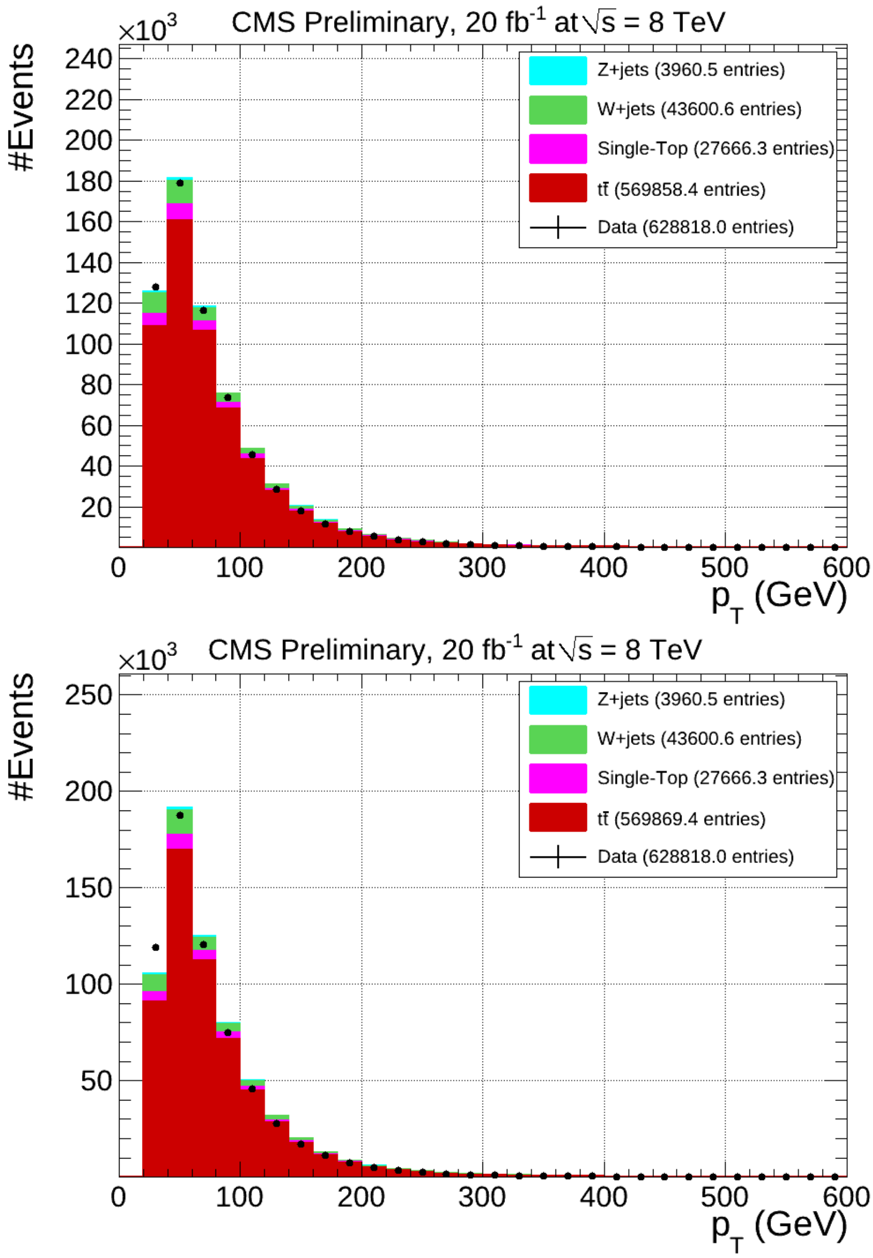

Figure 3.3: p_T distribution of the jets for all background processes, 8 TeV data and inclusive $t\bar{t}$ sample for events selected by the event selection described in Section 5.1. Original jets, selected by the event selection are shown in the upper panel. The bottom panel shows the jets after the on-the-fly reclustering with the FastJet package. The distribution shows a difference at the lower p_T values due to the absence of JEC for the reclustered jets.

Chapter 4

Event generation and simulation

When conducting an experiment one would like to compare the observed data with a physics model. These physics models are used to generate the expected output of the experiment according to a description of the physics phenomena. The event generation and the simulation happens as a sequential process. An illustration of the different steps of the event generation for proton-proton collisions is presented in Figure 4.1. First the interaction of the protons is considered using a detailed description of the hard scattering process. After the hard scattering a number of unstable particles remain, such as top quarks or W bosons which will decay, or other quarks and gluons that will first radiate other quarks and gluons (parton showering) and afterwards hadronize.
During the parton showering soft radiation gluons are produced. The evolution of the hadronization process from the free quarks and gluons cannot be described by perturbative QCD. Therefore a phenomenological point of view is used. The last part of the modelling will concern the underlying event (UE), which describes what happens to the proton remnants.
In the analysis performed for this thesis, different models of top quark events are compared to the 8 TeV data delivered by the CMS experiment. These models have different parameters for the parton showering and hadronization steps in the event generation or for the simulation of the underlying event.

Hard scattering The process of two protons interacting and producing a certain final state X can be calculated from the differential cross section $\frac{d\sigma_{pp\to X}}{dO}$, with respect to a certain observable O. This calculation depends on the partonic differential cross section $\frac{d\hat{\sigma}_{ij\to X}}{dO}$.
On the parton level the interaction of two protons is reduced to an interaction of the gluons and quarks inside the proton. In the hard scattering process top quarks are produced according to the three Feynman diagrams shown in Figure 1.2. The theoretical calculation of the probability to produce a top quark at the LHC en-

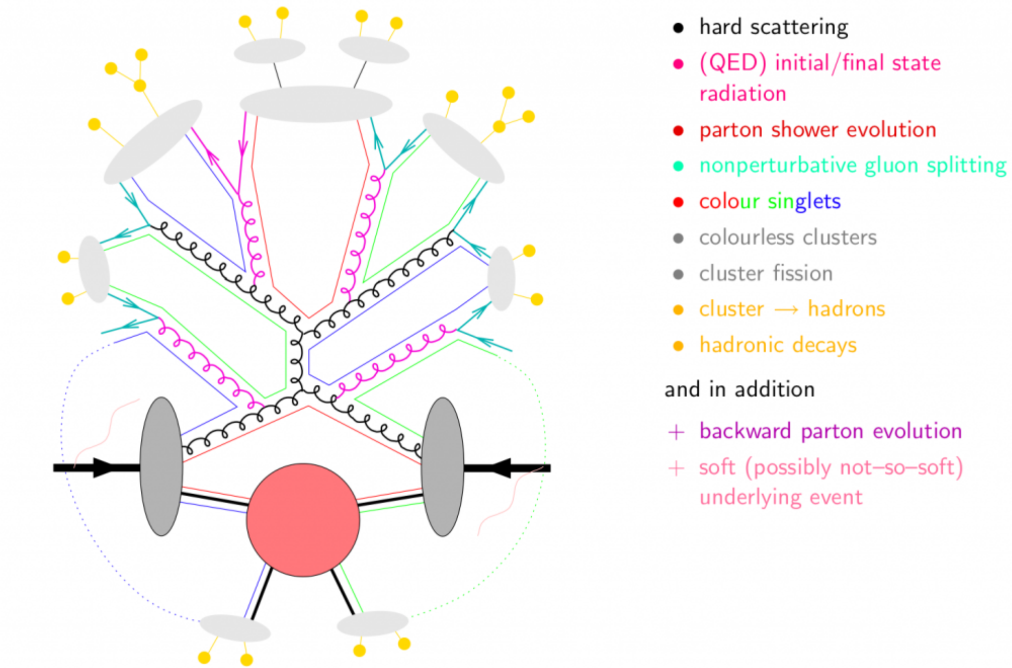

Figure 4.1: Illustration of the different steps in the simulation of a hard scattering process between two gluons [32]. The interaction starts at the bottom with the ISR represented by the red sphere. The hard scattering process starts at the gluon fusion and shows the parton showering up to point where hadronization takes over (yellow dots).

ergies results in gluon fusion as being the preferred method of the top quark pair production. The accuracy to which the calculations are performed is determined by the order of the strong coupling α_s. A number of generators are available to generate the events at leading order (LO) or next-to-leading order (NLO). For the hard scattering process the event generator `MadGraph/MadEvent` [33] is used to generate the events studied in this thesis.

Parton showering Once the hard scattering process has been generated one needs to perform the parton showering. This part of the event generation is performed by `Pythia` [34] for the events used in this thesis and describes the emission of partons by the quarks and gluons from the hard scattering. The branching of the strong interacting partons is described by the Dokshitzer-Gribov-Lipatov-Atarelli-Parisi (DGLAP) equations [8]. This equation gives the probability for a certain parton to split into two other partons at a given energy scale (Q^2). With the use of the DGLAP equations one can generate the parton shower for a given final state parton, starting from a scale Q^2_{max} calculating all of its sequential branching into other partons. This is performed up to a certain energy scale where the hadronization process will take over. Relative to a nominal model, two samples have been generated resulting in more ("scale down") and less ("scale up") Initial State Radiation (ISR) and Final State Radiation (FSR). This ISR and FSR will result in more or less reconstructed jets in the event as is shown in Figure 4.2.

Matching The matrix element describing the hard scattering process will need to be matched with the parton showering. This matching should avoid the double-counting of partons in an event. This double counting is due to the possible generation of (n+1)-jet event in two ways. Firstly through the showering of a final state from an (n+1)-parton matrix element. Secondly by a hard emission in the final state of an n-parton matrix element during the parton showering, which will result in an extra jet. The matching will ensure the (n+1)-jet final state to be produced through one of these two options and avoid it being produced by both of them at the same time.
Matching is subjected to an energy threshold at which the generation of the event by parton showering takes over from the hard-scattering matrix element. This threshold can be increased or decreased with respect to the nominal model. In the upcoming analysis two models ("matching up" and "matching down") are incorporated for with the matching threshold is varied.

Hadronization Once the simulation of the parton showers reaches the cutoff energy scale, the predictions in QCD can no longer be based on a perturbative approach. In this non-perturbative regime, the final part of the strong interacting

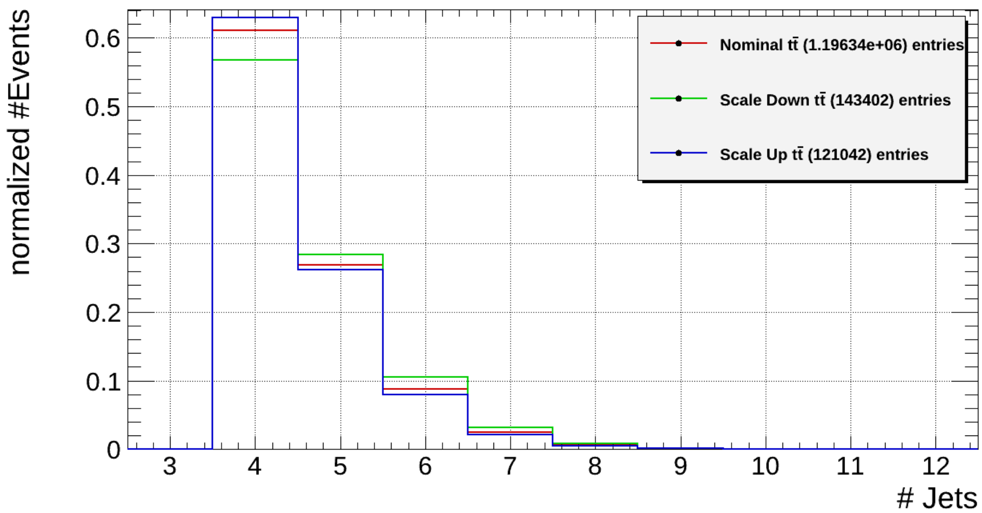

Figure 4.2: Jet multiplicity for the simulated events generated by the exclusive nominal, Scale Down and Scale up sample. The number of events is normalised for each distribution to 1.

avalanche is described by the hadronization. The particles produced by the parton showers will start to combine into stable (colour-neutral) hadrons. However not all final state particles after the hadronization will be stable, they may also decay. Typically the involved energy scales where hadronization takes over from the parton showering are of the order of $Q^2 \sim 1$ GeV. To predict the final state hadrons a phenomenological approach is needed. For the generation of the hadronization of the partons, the final states are simulated by the `Pythia` event generator with the use of the Lund string model. As for the parton showering a number of parameters can be tuned to produce different models for the hadronization process.

Underlying event While two partons, one from each proton undergo the hard scattering the remainder of the proton is no longer colour neutral. Therefore the remnants will also hadronize. These partons will also interact with each other and produce multi parton interactions (MPI). The whole of these interactions and the hadronization of the beam remnants is called the underlying event (UE).
Since the partons participating in this UE have colour charge they can have an effect on the hadronization process of the hard scattering due to the addition of soft particles. The effect of the UE is directly related to the geometry of the hard scattering. For central collisions the amount of MPI will be significantly larger then for peripheral collisions. To study the modelling of the underlying

event, samples are simulated with an increased number of multi parton interaction (*P11mpiHi* tune) and a model with less underlying event activity (*P11TeV* tune). These should be compared to the nominal P11tune.
A last effect, included in the models acts on the phenomenon of colour reconnection. During the hadronization, interactions between different colour strings of the partons of the underlying event can occur. This can change the colour structure in an event during the process. To take this into account a sample of events is produced with the colour reconnection disabled, referred to as the *P11noCR* tune.

Pile up Pile up (PU) is the presence of additional primary vertices in the detector. The presence of these additional primary vertices is due to a number of effects. A first effect is called out-of-time pile up and is caused by the presence of the particle remnants of a preceding bunch crossing when a new bunch crossing is being read out. The other main effect is caused by the large amount of protons in the colliding bunches. Therefore a number of additional protons will interact during the hard scattering. Due to this phenomenon, the events will contain more information than only the hard interaction. The distinction between the interaction of interest and pile up interactions is essential and highly depends on a good reconstruction of the primary vertices. From the reconstruction, the trigger can remove the out-of-time pile up. From the remainder of the interactions in the event the pile-up is then subtracted according to the number of primary vertices present in the event. Figure 4.3 shows the distribution of the number of primary vertices per event, which indicates the number of pile up events.

Background When a comparison is made between CMS data and a specific $t\bar{t}$ model, the background should be simulated as well. After the event selection requirement three main background processes will remain. A first background was already mentioned and is the single top quark production. A second background process which is included is the Drell-Yann process where an intermediate Z boson or virtual photon (γ^*) generates a final state similar to the $t\bar{t}$ events. Thirdly the background of leptonically decaying W bosons with additional jets is included. An overview of all simulated samples introduced in this chapter is presented in Table 4.1 for the top quark pair production and Table 4.2 for the background processes [8]. All of these models and their respective decay channel have a certain cross section at a centre of mass energy of 8 TeV and a number of simulated events have been generated corresponding to a certain integrated luminosity. The cross section and integrated luminosity are listed as well in Tables 4.1 and 4.2.

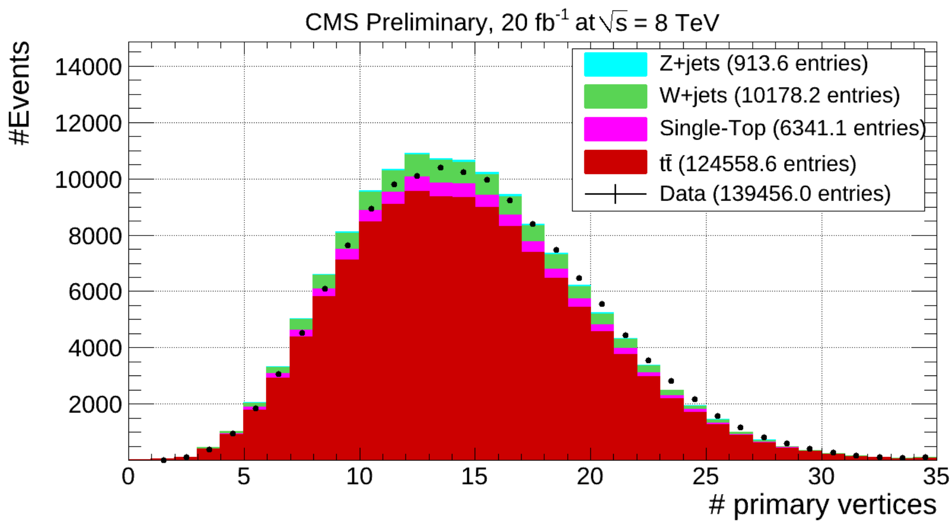

Figure 4.3: Distribution of the number of primary vertices for the simulated events and for the data. Each additional vertex represents a pile up interaction.

Table 4.1: Production cross section and integrated luminosity for all simulated samples at $\sqrt{s}$ = 8 TeV, used in the analysis [8]

Sample	Generator	σ (pb)	$\mathcal{L}$ (fb^{-1})
$t\bar{t}$ + jets	MadGraph + Pythia, tune Z2*		
Q^2 up, less ISR/FSR		245.8	20.3
Q^2 down, more ISR/FSR		245.8	21.9
matching up		245.8	21.9
matching down		245.8	22.3
$t\bar{t}$ + jets, m_t = 172.5 GeV	MadGraph + Pythia, tune P11		
l + jets		107.7	111.3
dilepton		25.8	225.4
all-hadronic		112.3	103.7
$t\bar{t}$ + jets, m_t = 172.5 GeV	MadGraph + Pythia, tune P11TeV		
l + jets		107.7	72.7
dilepton	(less MPI)	25.8	154.3
all-hadronic		112.3	70.7
$t\bar{t}$ + jets, m_t = 172.5 GeV	MadGraph + Pythia, tune P11mpiHi		
l + jets		107.7	73.9
dilepton	(more MPI)	25.8	154.2
all-hadronic		112.3	70.8
$t\bar{t}$ + jets, m_t = 172.5 GeV	MadGraph + Pythia, tune P11noCR		
l + jets		107.7	111.6
dilepton		25.8	227.6
all-hadronic		112.3	102.8

Table 4.2: Production cross section and integrated luminosity for all background samples at $\sqrt{s}$ = 8 TeV, used in the analysis [8]. Events generated by `MadGraph` + `Pythia` and `PowHeg` [35] + `Pythia`

Sample	Generator	σ (pb)	$\mathcal{L}$ (fb^{-1})
$W \to l\nu_l + jets$	`MadGraph` + `Pythia` tune Z2*		
W + 4 jets		264.0	50.7
W + 3 jets		640.4	24.2
W + 2 jets		2159.2	15.8
W + 1 jet		6662.8	3.5
$Z/\gamma^* \to l^+l^- + jets$	`MadGraph` + `Pythia` tune Z2*		
Z/γ^* + 4 jets		27.4	22.8
Z/γ^* + 3 jets		60.7	17.5
Z/γ^* + 2 jets		215.0	10.7
Z/γ^* + 1 jets		666.3	36.0
single top	`PowHeg` + `Pythia` tune Z2*		
t-channel t		56.4	66.0
t-channel $\bar{t}$		30.7	62.1
tW-channel t		11.1	44.5
tW-channel $\bar{t}$		11.1	44.5

Chapter 5

Data Analysis

To compare the data with the different simulated top quark samples presented in the previous chapter, a physics analysis is constructed. This analysis is applied both on the events simulated using the event generators and on the data recorded by the CMS detector for the 8 TeV proton collisions. The first part of the analysis discussed in this chapter presents the event selection requirements applied in Section 5.1, together with a discussion on the reconstruction of the observable for which simulated models are compared to the data (Section 5.2). Section 5.3 describes the applied method used to study the behaviour of the different models and the data. The results of the study are presented and discussed in Chapter 6.

5.1 Top quark event topology and event selection

To select top quark events from the full set of events, a number of requirements is applied. A first selection applies to the type and number of reconstructed objects present in the event. The decay channel on which this study focuses is the semi-muonic decay channel ($t\bar{t} \rightarrow b\bar{b}q\bar{q}\mu\nu_{\mu}$). It is clear that the selected events should have exactly one isolated muon and at least four jets. The events are selected with an isolated muon trigger. In addition, also offline requirements are applied.
For the muon a threshold on the pseudorapidity $|\eta| < 2.1$ is required corresponding to the muon trigger acceptance. Jets on the other hand should have a pseudorapidity of $|\eta| < 2.5$, corresponding to the tracker acceptance.
To increase the probability of selecting a $t\bar{t}$ event and eliminating background processes, requirements on the transverse momentum (p_T) of the reconstructed objects and missing transverse energy ($\not{E}_T$) are introduced. Since $t\bar{t}$ events require on average a higher energy to be produced compared to the background, the decay products and resulting final state particles will have high transverse momenta. For muons and jets a threshold of $p_T > 30$ GeV is applied. The leptonic decay of one

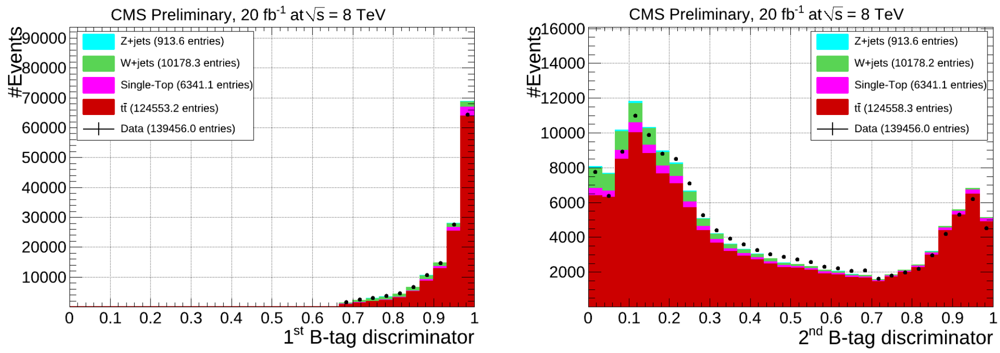

Figure 5.1: b-tag discriminator for the jet with the highest discriminator value (left) the jet with the second highest discriminator value (right). The discriminator used for b-jet identification in the analysis is the Combined Secondary vertex (CSV) algorithm with a medium criterion of 0.679 as shown in the left hand pane.

of the W bosons results in a neutrino to be produced. The presence of a neutrino can be inferred by the presence of $\not{E}_T$. The missing transverse energy is required to exceed 30 GeV.

Among the four (or more) selected jets, two of them originate from the hadronization of a b quark. Physicists are able to identify such b-jets by using b-discriminator thresholds on the jets. The b-discriminator used in this thesis is the Combined Secondary Vertex (CSV) algorithm. A medium criterion of 0.679 is applied to obtain a misstag efficiency of 1 %. Therefore only 1 % of the b-tagged jets are not originating from a b quark. The CSVM discriminator has a b-tag efficiency of 60 to 70 % [36], depending on the transverse momentum of the jet. Figure 5.1 shows the distribution of this b-tag discriminator for the two jets with highest discriminator value in the events.

With the applied threshold on the $\not{E}_T$ and CSVM requirement, all QCD multijet events that passed the previous event selection and who were not simulated are rejected. The performance of the event selection can be reviewed with a number of checks, presented below. As a first check, the transverse momentum of the muon can be reviewed, as shown in Figure 5.2. The same thresholds were introduced on all jets and the $\not{E}_T$ in the event selection. The resulting distributions are shown in Figure 5.3 for the missing transverse energy and in Figure 5.4 for the four jets with highest transverse momentum.

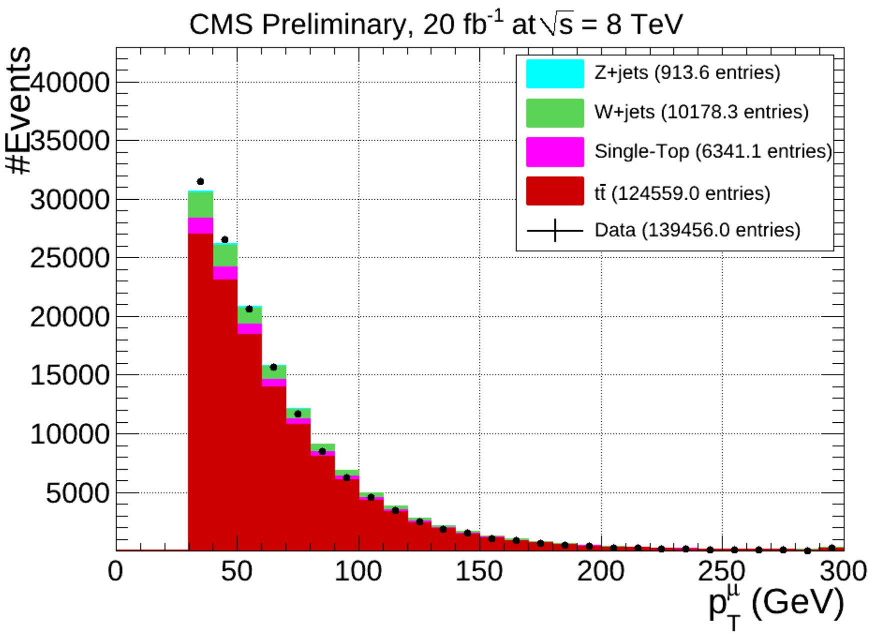

Figure 5.2: Distribution of the transverse momentum of the isolated muon in the events. For the muon, the event selection introduced a threshold of $p_T > 30$ GeV on the transverse momentum of the muon.

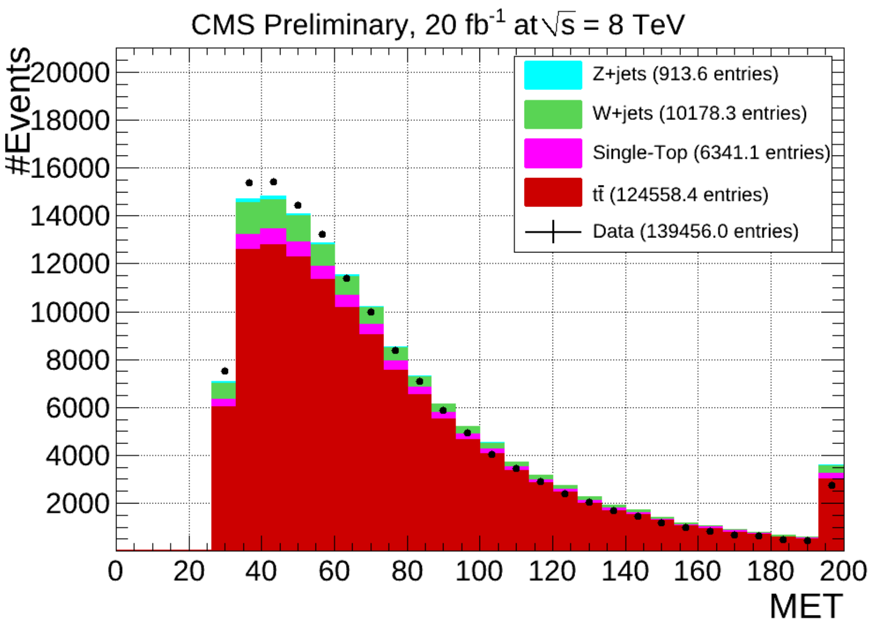

Figure 5.3: Distribution of the missing transverse energy ($\not\!\!E_T$) for the events. The threshold of $\not\!\!E_T > 30$ GeV from the event selection is clearly noticed in the distribution.

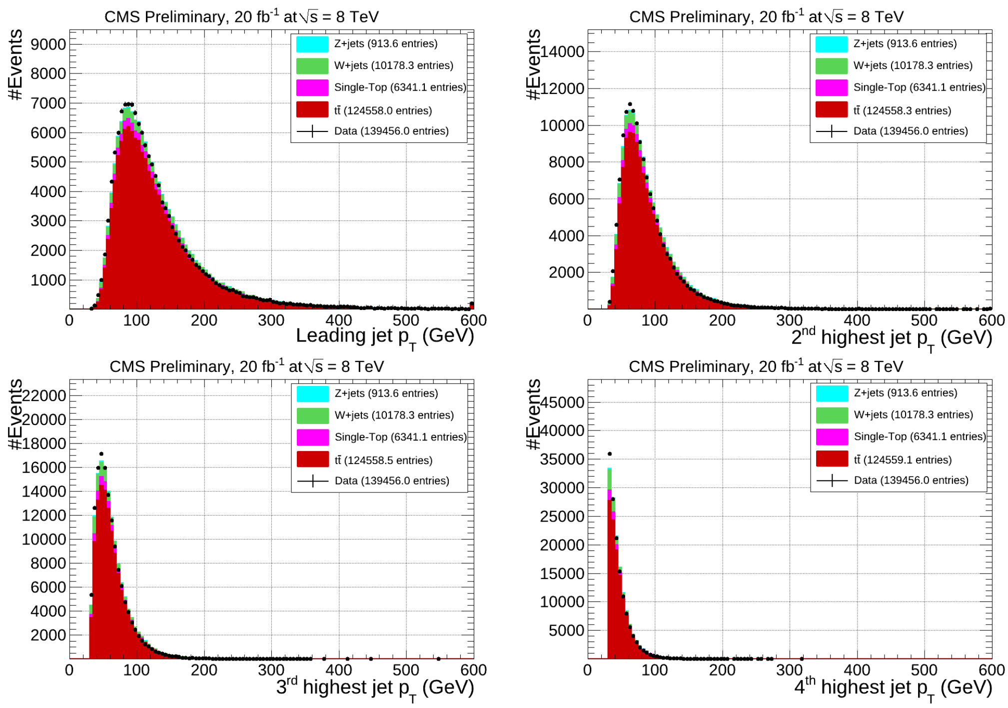

Figure 5.4: p_T distribution of the four selected jets with highest transverse momenta. From left to right and top to bottom, starting with the leading jet, the jets with second highest p_T, the jets with third highest p_T and the jets with fourth highest p_T. One can clearly see the $p_T > 30$ GeV threshold introduced in the event selection for the transverse momentum of the jets.

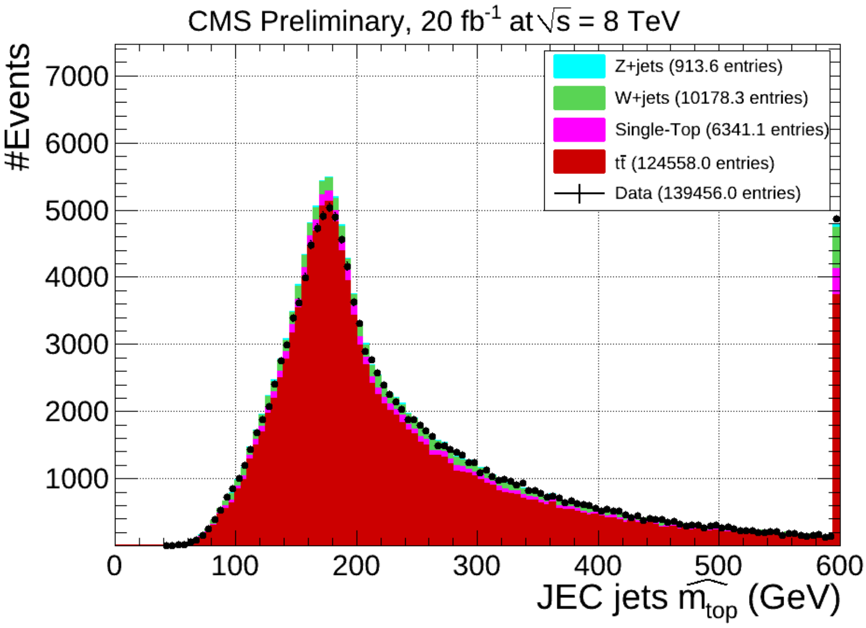

Figure 5.5: Distribution of the M3 variable as an estimate for the top quark mass, for the original jets with JEC applied and no reclustering performed.

5.2 Reconstruction of the top quark mass estimator "M3"

When studying a physics model and comparing it to the data, one needs to do this for a certain observable. In this study a top quark mass estimator is used, referred to as "M3". This mass reconstruction uses the four selected jets with the highest transverse momentum. The three-jet combination resulting in the highest p_T of the combined object is chosen to be the top quark candidate. The M3 variable is the mass of this reconstructed top quark candidate.

5.2.1 Fitting procedure

The above reconstruction is performed for each event individually. This way a distribution of the M3 variable for the top quark candidates is obtained, as is shown in Figure 5.5. This distribution is obtained for all samples and the background. To estimate the top quark mass ($\hat{m}_{top}$) and its uncertainty, the M3 distribution is fitted with a Gaussian function. The expectation value of the Gaussian represents the estimator $\hat{m}_{top}$. The M3 variable is an estimator of the top quark mass and therefore the distribution should follow a Breit-Wigner function. However due to the detector resolution, the distribution can be fitted with a Gaussian instead of a Breit-Wigner. When fitting the distribution, the function depends a lot on the range over which the fit is performed. Therefore the distribution will not follow a proper Gaussian function over the full range of the distribution. To fit only the Gaussian part of the M3 distribution, the following procedure is adopted to obtain the best fit. The quantity to determine the best fit for each distribution is the χ^2/dof value of the fit, for which the definition of the χ^2 value [37] can be found in Equation 5.1. The variables x_i indicate the values of the bins through which the fitting is performed and σ_{x_i} the uncertainty for the bin. The variable $\hat{\mu_x}$ is the expectation value of the fitted function. The sum runs over all bins included in the range of the fit.

$$\chi^2(\hat{\mu_x}) = \sum_{i=1}^{n} \frac{(x_i - \hat{\mu_x})^2}{\sigma_{x_i}^2} \tag{5.1}$$

The "dof" term represents the number of degrees of freedom, in this case the number of bins used in the fit.

The fitting procedure uses a variation on the range over which the fitting is performed. This range is defined by the Root Mean Square (RMS) [37] of the distribution. The definition of the RMS is shown in Equation 5.2. As the name suggests, the RMS is defined by taking the square root of the mean value of the distribution squared.

$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \tag{5.2}$$

The initial range is set by taking the bin with highest number of entries and adding the full RMS to both sides. The range is then decreased in steps of $\sim$ 5 GeV on each side independently. For each newly defined range the distribution is fitted. The range for which the fit returns the smallest value of the χ^2/dof is chosen as the final fit range. As an example of the procedure, four fits are shown in Figure 5.6, corresponding to the maximum fit range (full RMS), half fit range (half RMS), minimum fit range (four bins on both sides) and the range corresponding to the fit resulting in the lowest χ^2/dof. The resulting values for $\frac{\chi^2}{dof}$ for each fitted function is shown in Table 5.1. For most of the distributions, the minimum range corresponds also to the range yielding the best fit. To avoid statistical effects from the distributions on the fit, this procedure is only performed on the distribution of the reclustered jets where no p_T threshold was applied. The range specified by the resulting best fit for this distribution is then used for all distributions where p_T thresholds are applied.

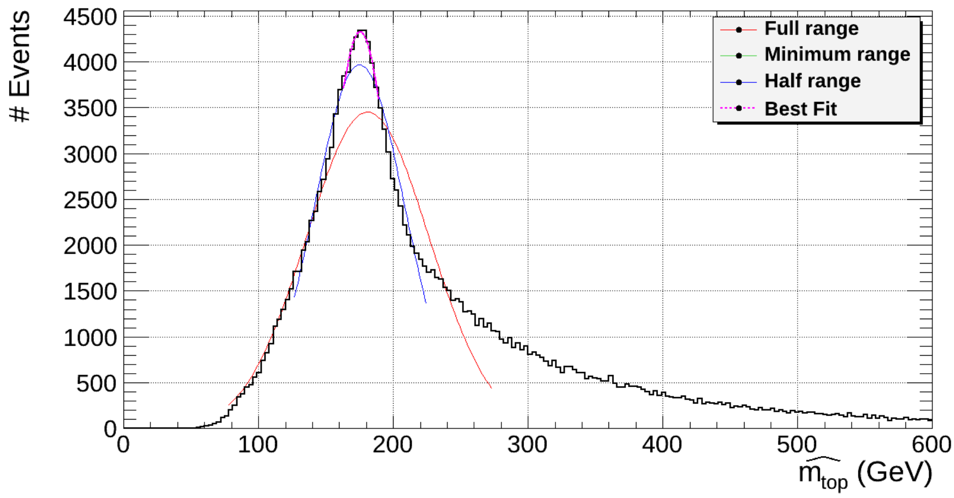

Figure 5.6: Fitting example for χ^2 optimisation. M3 distribution of the Inclusive Nominal $t\bar{t}$ events with fit range adapted. The full range corresponds to the RMS added left and right, half range to a half RMS added to both sides. The minimum range is set to four bins left and right (range of 45 GeV around maximum).

Table 5.1: Fit parameters resulting from the fitting of the M3 distribution of the Inclusive $t\bar{t}$ sample

Range	$\hat{m}_{top}$ (GeV)	σ (GeV)	χ^2/dof
full range (RMS)	180.97	0.15	4610.54
Half range ($\frac{1}{2}$ RMS)	174.91	0.14	476.51
Minimum range (4 bins left and right)	175.54	0.34	6.98
Best fit	175.54	0.34	6.98

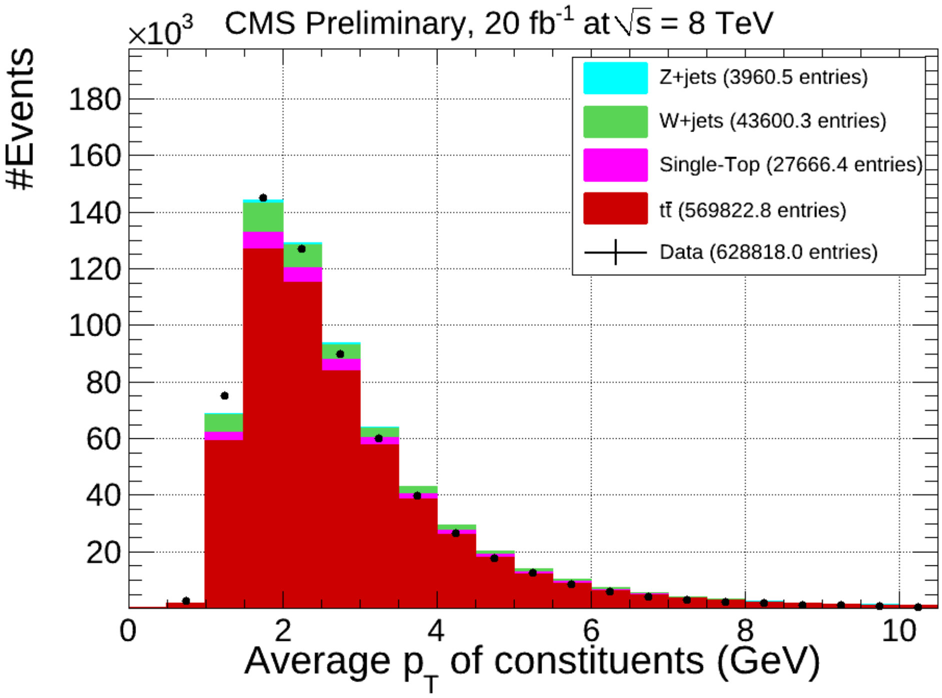

Figure 5.7: Average transverse momentum of the jet constituents without applying a threshold on constituents p_T.

5.3 Top quark mass evolution

To compare effects of the $t\bar{t}$ modelling for the different simulated samples with the data, we will apply different thresholds on the transverse momentum of the jet constituents. To achieve this we need to be able to access the constituents which are stored for the selected jets.

The constituents of the selected jets are reclustered using the FastJet algorithm. A threshold is applied on the transverse momentum of the individual constituents. As shown in Figure 5.7, the average p_T of the constituents is maximal around 1.5 GeV. The maximum threshold applied in this analysis is 2 GeV. The minimum threshold is chosen to be 0 GeV, which allows a comparison of the jets before and after the reclustering. In Figure 3.3 the transverse momentum of the jets before and after reclustering was shown. The effect of the jet energy corrections is clearly visible since jet energy corrections are not applied after reclustering. When increasing the p_T-threshold on the jet constituents, more and more will be removed from the list of particles selected for the reclustering into new jets. As shown in Figure 5.8, constituents will be removed starting from thresholds below 100 MeV. One can compare the number of accepted constituents as a function of the applied

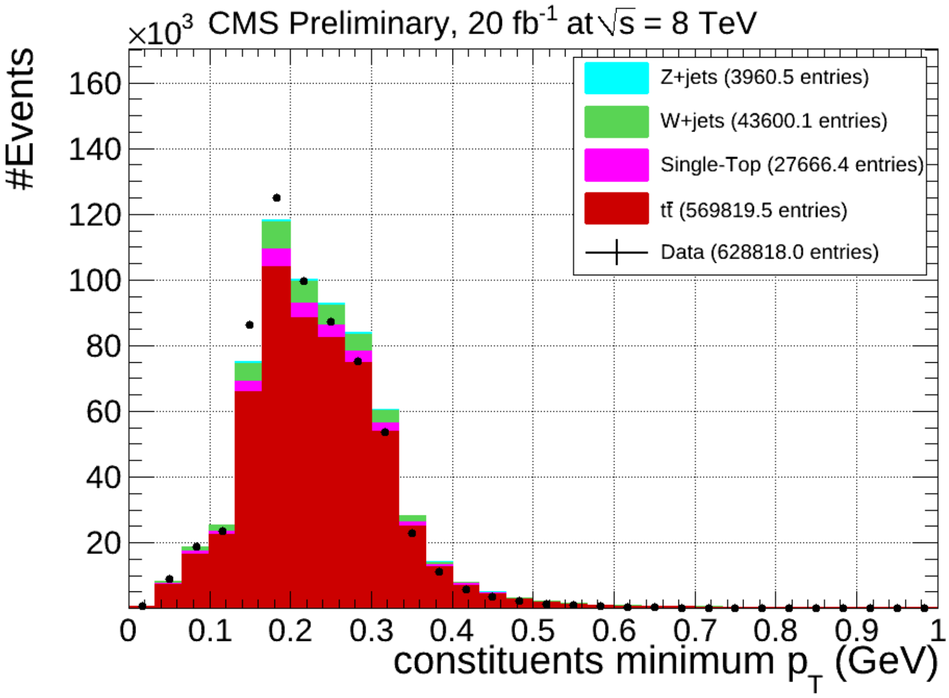

Figure 5.8: Minimum transverse momentum of the jet constituents for each event, without applying a threshold on the constituents p_T.

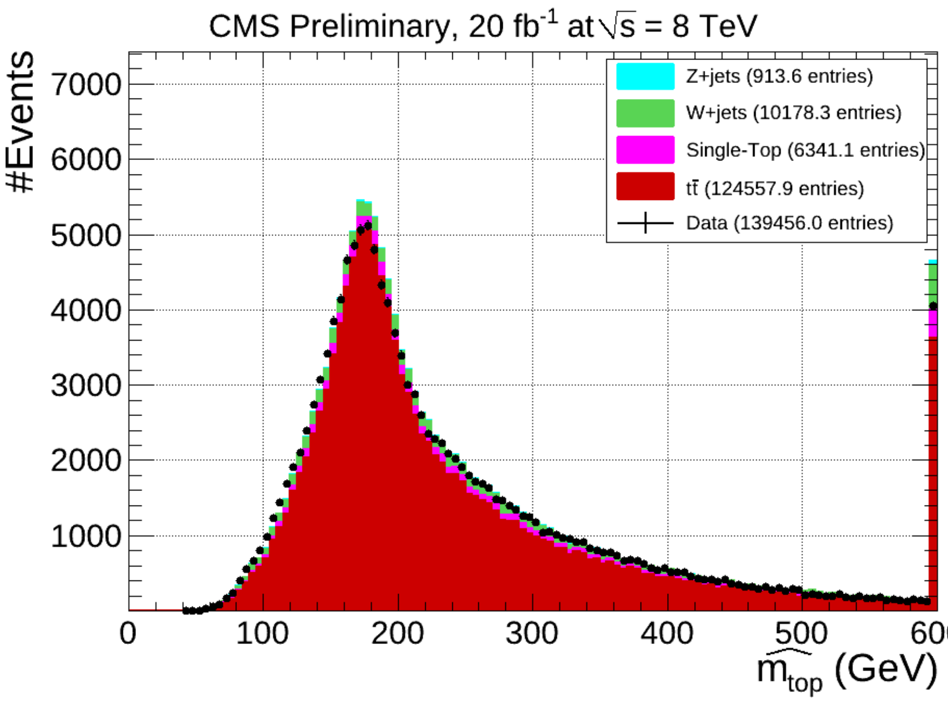

Figure 5.9: M3 distribution of the jets after reclustering the constituents without applying a threshold on the transverse momentum of the jet constituents.

threshold. After reclustering, the jets of the original event are matched with the reclustered jets. This matching minimises the distance $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ to check if the directions of the original and reclustered jets coincide. If the distance for two jets is less than $\Delta R = 0.1$ the two jets are said to be matched. An event is only accepted if the jets before and after the reclustering are all matched.
After the reclustering of the jets, the M3 variable is calculated for each simulated $t\bar{t}$ sample and each applied p_T threshold. An example of the M3 distribution with jets reclustered by the FastJet algorithm without applying a p_T threshold on the constituents is shown in Figure 5.9. One can see the distributions for the reconstructed top quark mass are not identical for the jets with JEC applied and the jets after reclustering. The distribution of both the top quark mass estimation for the jets with JEC and reclustered jets is shown in Figure 5.10.

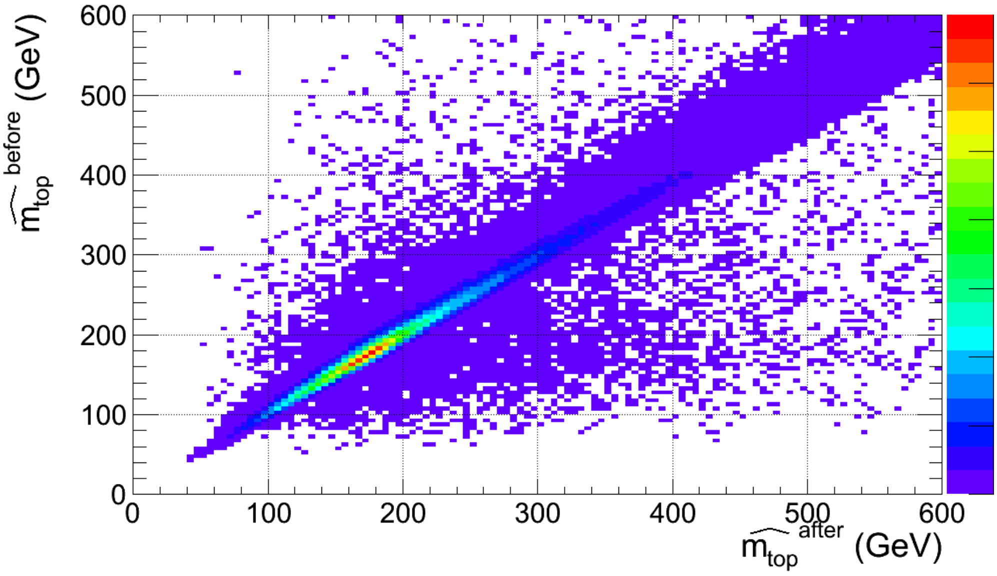

Figure 5.10: Distribution of the reconstructed top quark masses from the events for the original jets with jet energy corrections (JEC) applied against the reconstructed top quark mass for the jets after reclustering with no p_T threshold on the transverse momentum applied.

Chapter 6

Study of the modelling of top quark events

6.1 Top quark mass estimation

For each simulated top quark sample combined with the background and for the data, the top quark mass is estimated for various thresholds on the transverse momentum of the jet constituents, as shown in Figure 6.1. Figure 6.2 shows the estimated top quark mass ($\hat{m}_{top}$) for the lowest p_T threshold values. There is a clear difference between the estimated top quark mass in the data and for all simulated models.
This effect is due to the absence of the jet energy corrections after the reclustering and will result in a bias for the top quark mass estimation. We calculate this bias by determining the difference in estimated top quark mass from the original jets for which the JEC are applied and the jets after reclustering for each model. This difference is determined only for the top mass distribution of the reclustered jets where no p_T threshold on the constituents was applied.
We correct for the bias after reclustering by applying the observed difference to the estimated top quark mass, obtained for each threshold on the p_T of the constituents. The obtained bias for each model is shown in Figure 6.3. Due to the absence of JEC on the the jets after the reclustering, the transverse momentum of the jets used for the M3 calculation is different from the transverse momentum of the jets before the reclustering procedure was applied.
The "bias-corrected" top quark mass estimation is shown in Figure 6.4 for all p_T thresholds and in Figure 6.5 for the lowest threshold values.

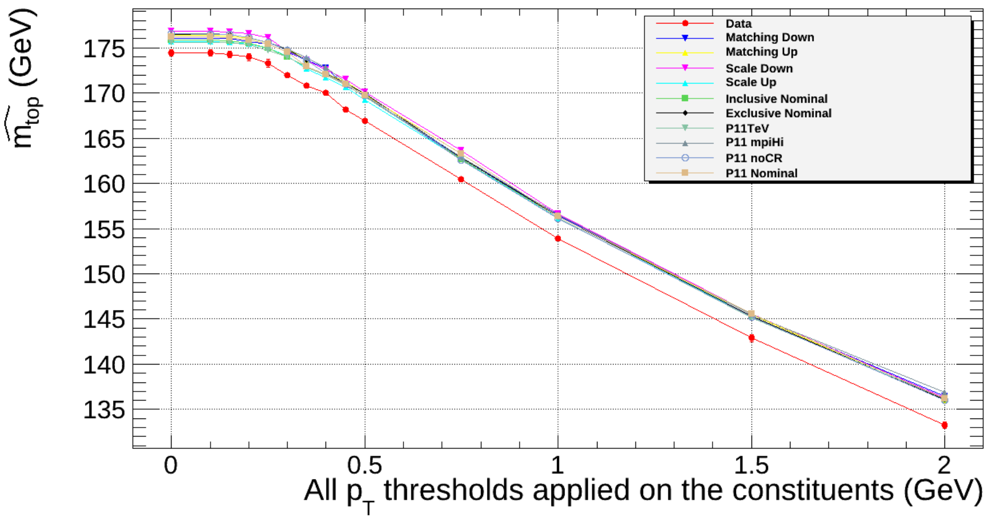

Figure 6.1: Estimation of the top quark mass as a function of the p_T threshold on the jet constituents for the data and various simulated top quark samples.

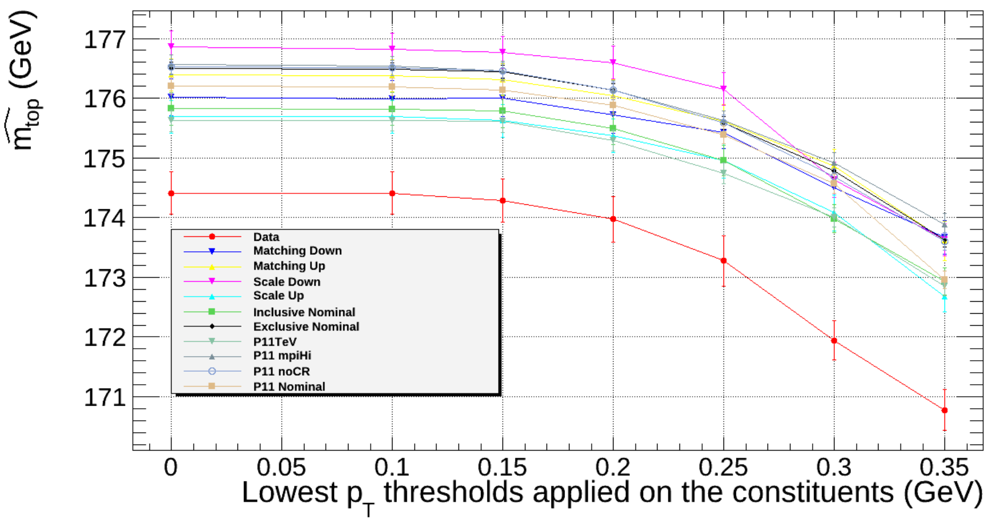

Figure 6.2: Estimated top quark mass as a function of the lowest p_T thresholds on the jet constituents for the data and various simulated top quark samples.

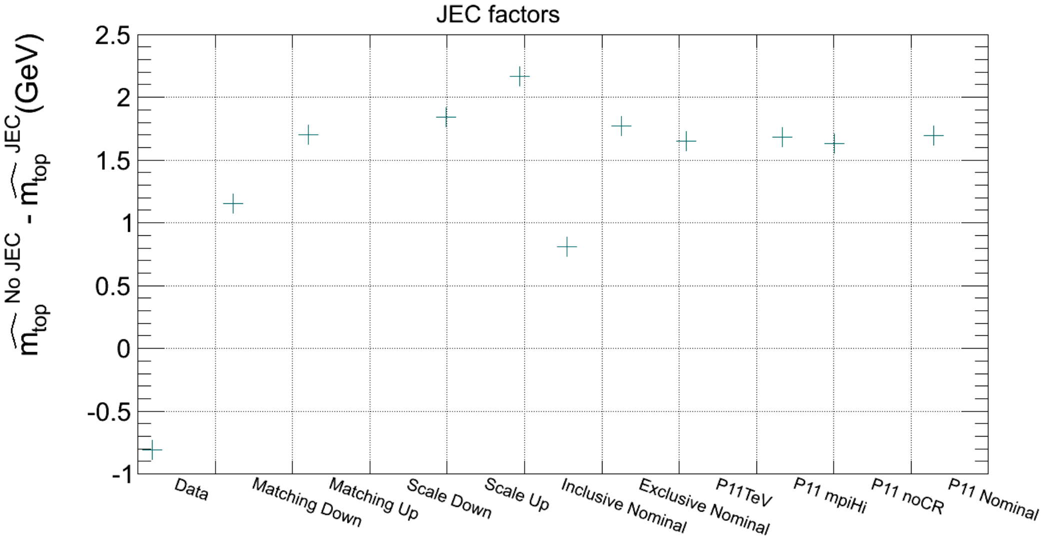

Figure 6.3: Difference between the estimated top quark mass for each model and the data, using jets before and after reclustering without applying a threshold on the p_T of the constituents. The difference between the estimated top quark mass for a certain model is due to the jet energy corrections applied on the original jets and will be used for the correction of the bias.

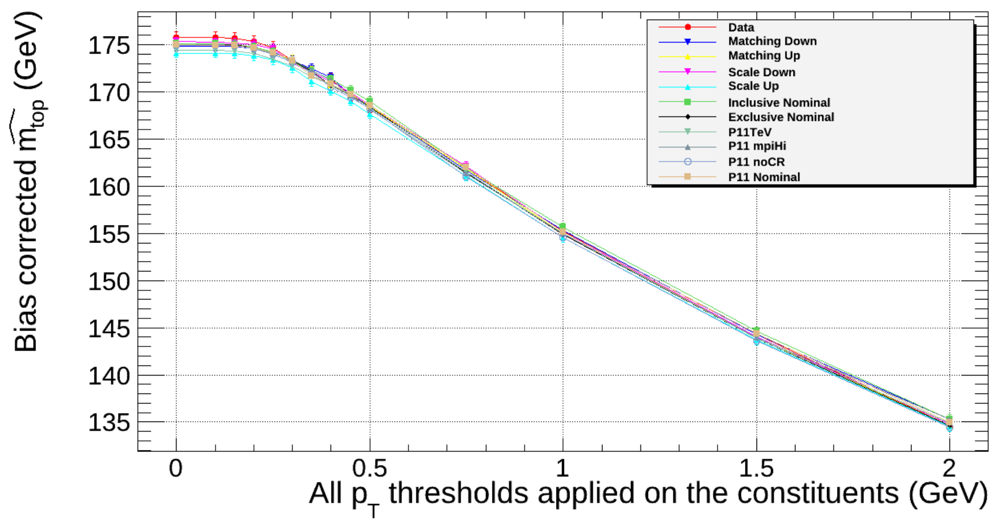

Figure 6.4: Estimation of the top quark mass as a function of the p_T threshold on the jet constituents for the data and various simulated top quark samples, with JEC correction applied to the top quark mass estimates by bias subtraction

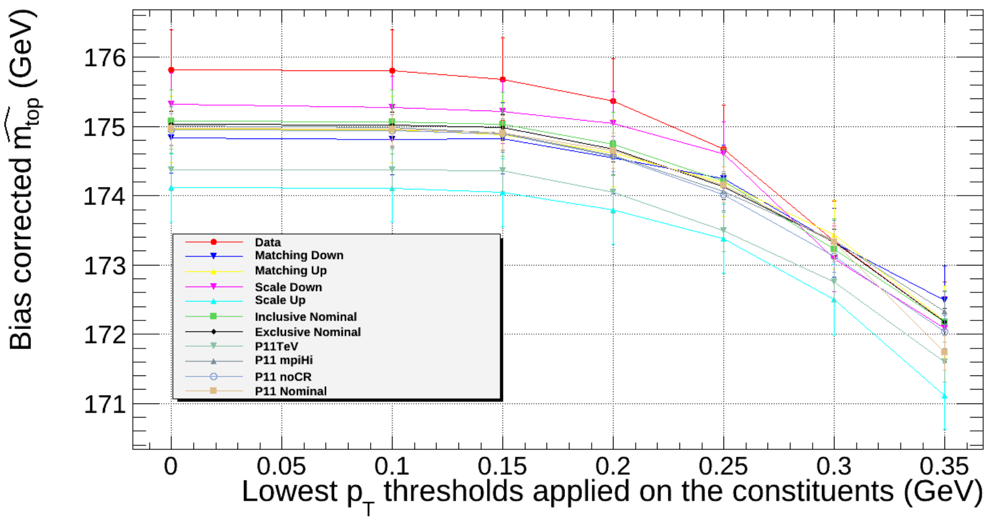

Figure 6.5: Estimation of the top quark mass as a function of the lowest p_T threshold on the jet constituents for the data and various simulated top quark samples, with the JEC correction applied to the top quark mass estimates by bias subtraction.

6.2 Evolution of the estimated top quark mass

A comparison between the top quark mass estimates for the different models is shown in Figures 6.4 and 6.5. The events used for each of the top quark mass estimates at different p_T thresholds contain the same constituents to build up the jets. Therefore the measurements are not independent. To obtain a better estimate of the uncertainty on $\hat{m}_{top}$ we will determine the correlation between the measurements at the different thresholds. The correlation factor R^{i-j} is calculated for each difference between the top quark mass estimates at thresholds "i" and "j". To calculate this factor, the semileptonic nominal $t\bar{t}$ sample is used. This sample is divided in 100 independent sets of events with an equal number of events. For each of these subsamples the top quark mass is estimated after each cut is applied on the p_T of the constituents. One can then obtain the distribution of the difference between the top mass measurements at the applied thresholds.
The correlated uncertainty ($\sigma^{corr}_{\hat{m}^i_{top}}$) between the estimates $\hat{m}^i_{top}$ and $\hat{m}^j_{top}$ is determined by the square root of the variance of a Gaussian function fitted to the distribution of their differences $\Delta\hat{m}^{i-j}_{top}$. An example of this distribution at thresholds p_T = 0.10 GeV and p_T = 0.35 GeV is shown in Figure 6.6.
In addition, a measure for the uncorrelated uncertainty ($\sigma^{uncorr}_{\hat{m}^{i-j}_{top}}$) between the thresh-

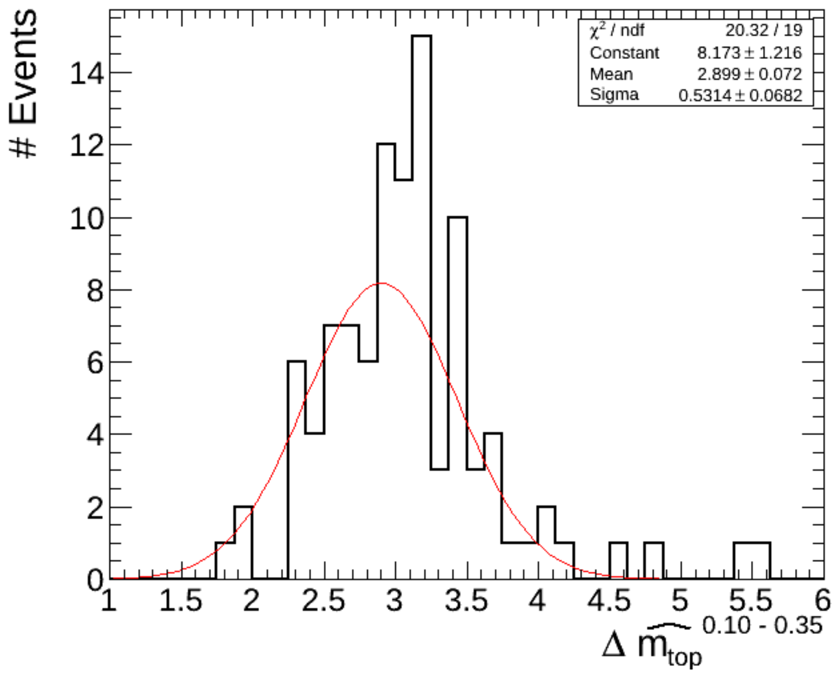

Figure 6.6: Distribution of the difference between the top quark mass estimates at applied p_T thresholds of 0.10 GeV and 0.35 GeV. The correlated uncertainty between the estimates $\hat{m}_{top}^{p_T=0.10}$ and $\hat{m}_{top}^{p_T=0.35}$ is determined by the variance of the Gaussian function fitted to this distribution.

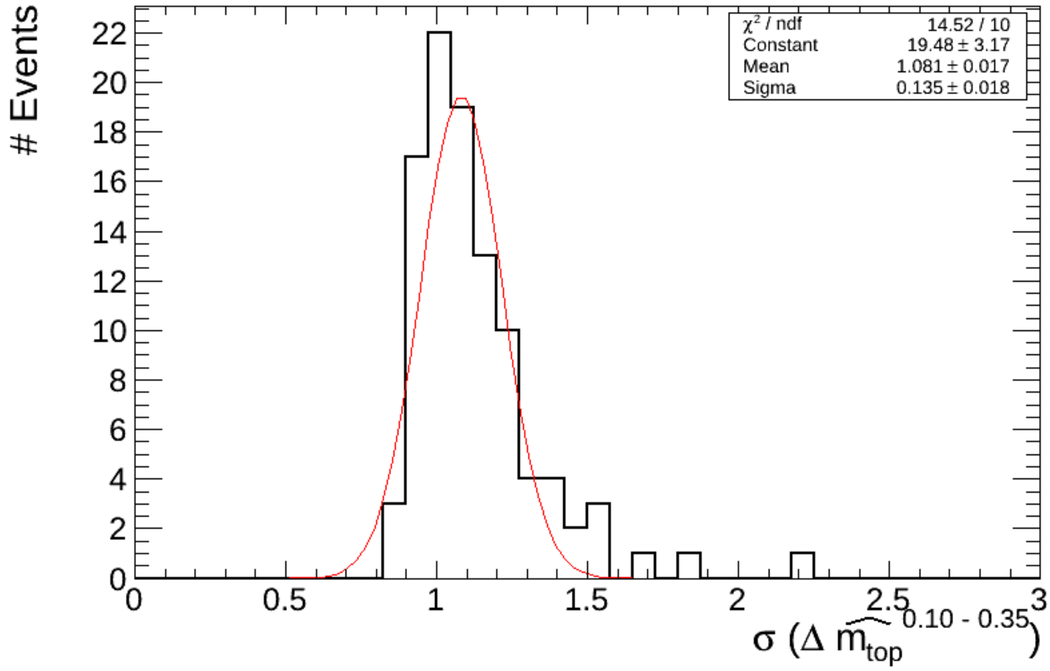

Figure 6.7: Distribution of the uncorrelated uncertainty between the top quark mass estimates at applied p_T thresholds of 0.10 GeV and 0.35 GeV. The estimated value of the uncorrelated uncertainty between $\hat{m}_{top}^{p_T=0.10}$ and $\hat{m}_{top}^{p_T=0.35}$ is determined by the expectation value of the Gaussian function fitted to this distribution.

olds "i" and "j" needs to be obtained for the top quark mass estimates. To calculate this uncertainty on the difference in top quark mass for two different p_T thresholds "i" and "j", we will use the propagation of the uncertainties:

$$\sigma_{\Delta\hat{m}_{top}^{i-j}} = \sqrt{(\sigma_{\hat{m}_{top}^{i}})^2 + (\sigma_{\hat{m}_{top}^{j}})^2}$$

The expectation value of the Gaussian fit to the distribution of $\sigma_{\Delta\hat{m}_{top}^{i-j}}$ is the uncertainty on the difference in top quark mass between these thresholds, if there would be no correlation, $\sigma_{\Delta\hat{m}_{top}^{i-j}}^{uncorr}$. An example of this fitted distribution is shown in Figure 6.7 for the uncorrelated uncertainty on the top quark mass difference between thresholds p_T = 0.10 GeV and p_T = 0.35 GeV. From parameters $\sigma_{\Delta\hat{m}_{top}^{i-j}}^{uncorr}$ and $\sigma_{\Delta\hat{m}_{top}^{i-j}}^{uncorr}$ one can obtain the correlation factor R^{i-j} for applied thresholds i and j:

$$R^{i-j} = \frac{\sigma_{\hat{m}_{top}^{i}}^{corr}}{\sigma_{\hat{m}_{top}^{i}}^{uncorr}}$$

Table 6.1: Results for the slope of the top quark mass estimations for p_T thresholds at 0.10 GeV and 0.35 GeV. Both the uncorrelated uncertainty and uncertainty with correlation factor applied are shown.

Sample	p_T thresholds (GeV)	$\Delta\hat{m}_{top}^{0.10-0.35}$ (GeV)	$\sigma_{\Delta\hat{m}_{top}}$ (GeV) (uncorrelated)	$\sigma_{\Delta\hat{m}_{top}}$ (GeV)
Data	0.10 - 0.35	3.63	$\pm$ 0.82	$\pm$ 0.05
Matching Down	0.10 - 0.35	2.32	$\pm$ 0.71	$\pm$ 0.05
Matching Up	0.10 - 0.35	2.77	$\pm$ 0.70	$\pm$ 0.05
Scale Down	0.10 - 0.35	3.18	$\pm$ 0.64	$\pm$ 0.04
Scale Up	0.10 - 0.35	3.00	$\pm$ 0.68	$\pm$ 0.04
Inclusive Nominal	0.10 - 0.35	2.88	$\pm$ 0.63	$\pm$ 0.04
Exclusive Nominal	0.10 - 0.35	2.84	$\pm$ 0.26	$\pm$ 0.02
P11 Nominal	0.10 - 0.35	3.23	$\pm$ 0.35	$\pm$ 0.02
P11TeV	0.10 - 0.35	2.77	$\pm$ 0.41	$\pm$ 0.03
P11 mpiHi	0.10 - 0.35	2.65	$\pm$ 0.40	$\pm$ 0.03
P11 noCR	0.10 - 0.35	2.90	$\pm$ 0.35	$\pm$ 0.02

The correlation factor can be used to correct the statistical uncertainty on the difference of the top quark mass estimates. Table 6.1 shows the difference of the top quark mass estimates at p_T thresholds of 0.10 GeV and 0.35 GeV for all samples. The final uncertainty on the top quark mass estimate is shown in the last column, where the correction factor of Equation 6.1 was applied on the originally determined uncertainties $\sigma_{\hat{m}_{top}}$ to take into account the correlation between the top quark mass estimates at p_T thresholds of p_T = 0.10 GeV and p_T = 0.35 GeV.
When reviewing the effect of the correlation factor on the uncertainty of the top quark mass estimates, one can see there is a large correlation. In Appendix A we show the correlation between the estimates at successive values becomes even larger, resulting in a smaller uncertainty. The correlation factors for the between the different top quark mass estimates are shown in Appendix B. This shows an important effect can be included by taking the correlation of the jet properties into account when reconstructing top quark events. From the slopes shown in table 6.1 one can see a large difference for the data with respect to the various simulated models.
The large difference for the slope of the data and the simulated samples can be interpreted when returning back to Figure 5.7 and Figure 5.8. For the data, more constituents with low p_T values are clustered in the jets. Therefore when applying thresholds on the jet constituents, the jets from the data samples will have more rejected constituents, resulting in a steeper slope when looking at the difference of the top quark mass estimate for two p_T thresholds.

Chapter 7

Conclusions and outlook

To study the modelling of top quark events at the CMS experiment we have compared a number of $t\bar{t}$ models with the data recorded by the CMS detector at a centre of mass energy of $\sqrt{s}$ = 8 TeV. For each of these models we estimated the top quark mass as the expectation value of a Gaussian function fitted to the M3 distribution. The top quark mass is estimated for a number of threshold values on the transverse momentum applied to the constituents of the original jets. After introducing these thresholds the selected constituents are reclustered into new jets, which are matched to the original jets using the direction of these jets.
One can study the evolution of the estimated top quark mass ($\hat{m}_{top}$) as a function of the applied threshold on the p_T of the jet constituents for each simulated $t\bar{t}$ model and the data.
A first observation is that the estimator is biased with respect to the estimated value for the original jets. This bias is due to the absence of the Jet Energy Corrections (JEC) on the reclustered jets and we correct for it. To study the evolution of the top quark mass estimation we calculate the slope for each of the models between two of the applied p_T thresholds. These slopes have however a large uncertainty and should be corrected for the correlation between the measurements.
Using the corrected small uncertainty we were able to study the change in the estimated top quark mass as a function of the applied p_T thresholds on the jet constituents. From this we can see that the simulated models do not agree with the data for which the slope is steeper. This is due to a difference that is observed between the data and the simulated samples for the p_T distribution of the constituents. A clear indication of this difference is shown in Figure 5.8. Here one can see the data has a larger number of constituents with low p_T, which is not accurately described by any of the used models. Therefore this study should be continued by investigating why data and the models disagree after which the models can be tuned to obtain more jet constituents with a lower p_T.

Appendices

Appendix A

Top quark mass difference for succesive p_T thresholds

Table A.1: Results for the slope of the top quark mass estimations for successive p_T thresholds at 0.15 GeV and 0.2 GeV. Both the uncorrelated uncertainty and uncertainty with correlation factor applied are shown. One can see there is a large correction to the preliminary uncertainties due to the correlation of the measurements.

Sample	p_T thresholds (GeV)	$\Delta\hat{m}_{top}^{0.10-0.35}$ (GeV)	$\sigma_{\Delta\hat{m}_{top}}$ (GeV) (uncorrelated)	$\sigma_{\Delta\hat{m}_{top}}$ (GeV)
Data	0.15 - 0.20	0.311	$\pm$ 0.845	$\pm$ 0.02
Matching Down	0.15 - 0.20	0.288	$\pm$ 0.724	$\pm$ 0.01
Matching Up	0.15 - 0.20	0.271	$\pm$ 0.685	$\pm$ 0.01
Scale Down	0.15 - 0.20	0.167	$\pm$ 0.648	$\pm$ 0.0.1
Scale Up	0.15 - 0.20	0.250	$\pm$ 0.695	$\pm$ 0.01
Inclusive Nominal	0.15 - 0.20	0.290	$\pm$ 0.6508	$\pm$ 0.01
Exclusive Nominal	0.15 - 0.20	0.308	$\pm$ 0.2590	$\pm$ 0.004
P11 Nominal	0.15 - 0.20	0.255	$\pm$ 0.361	$\pm$ 0.006
P11TeV	0.15 - 0.20	0.309	$\pm$ 0.423	$\pm$ 0.007
P11 mpiHi	0.15 - 0.20	0.322	$\pm$ 0.397	$\pm$ 0.007
P11 noCR	0.15 - 0.20	0.311	$\pm$ 0.321	$\pm$ 0.005

Appendix B

Correlation factors of the top quarks mass estimates for the different p_T thresholds

Table B.1: Correlation correction factors between the top quark mass estimations for all thresholds applied on the p_T of the jet constituents

p_T thresholds (GeV)	R^{i-j}		p_T thresholds (GeV)	R^{i-j}
0.10 - 0.15	0.01		0.20 - 0.25	0.03
0.10 - 0.20	0.02		0.20 - 0.30	0.04
0.10 - 0.25	0.03		0.20 - 0.35	0.05
0.10 - 0.30	0.05		0.20 - 0.40	0.06
0.10 - 0.35	0.07		0.25 - 0.30	0.04
0.10 - 0.40	0.06		0.25 - 0.35	0.06
0.15 - 0.20	0.02		0.25 - 0.40	0.07
0.15 - 0.25	0.05		0.30 - 0.35	0.03
0.15 - 0.30	0.04		0.30 - 0.40	0.07
0.15 - 0.35	0.06		0.35 - 0.40	0.03
0.15 - 0.40	0.08			

Bibliography

[1] Guido Altarelli, *Collider physics within the standard model: a primer*, arXiv:1303.2842v2, [hep-ph] 4 Nov 2013

[2] F. Mandl, G. Shaw, *Quantum field theory (second edition)*, (April 2010) 492 p.

[3] S.F. Novaes, *Standard model: an introduction*, arXiv:0001283v1 [hep-ph], 27 Jan 2000.

[4] Science tumbled, *while we're on the topic of quantum physics*, http://science.tumblr.com.

[5] J. Beringer et al. (Particle Data Group), PR **D86**, 010001 (2012) and 2013 (partial update 2014) (http://pdg.lbl.gov).

[6] The ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, arXiv:1207.7214 [hep-ex], 31 August 2012.

[7] The CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, arXiv:1207.7235 [hep-ex], 28 January 2013.

[8] S. Blyweert, *Measurement of the top-quark mass and the mass difference between top and antitop quarks at the LHC*,Phd thesis ,August 2013.

[9] P. Langacker, *Grand unified theories and proton decay*, Physics reports volume 72, issue 4, June 1981.

[10] CDF Collaboration, *Observation of Top Quark Production in Pbar-P Collisions*, Phy. Rev. Lett. **74** (1995) 2626-2631, doi:10.1103/PhysRevLett.74.2626, arXiv:hep-ph/9503002.

[11] *DØ* Collaboration, *Observation of the top quark*, Phys. Rev. Lett. **74** (1995) 2632-2637, doi:10.1103/PhysRevLett.74.2632, arXiv:hep-ex/9503003.

[12] *The CDF Experiment*. http://www-cdf.fnal.gov/.

[13] *The DØ Experiment*. http://www-d0.fnal.gov/.

[14] *The Tevatron collider*. http://www-bdnew.fnal.gov/tevatron/.

[15] T.M. Liss, A. Quadt, *The top quark*, Particle Data Group, particle reviews 2011 (updated 2013).

[16] L. Fiorini, *Top-Quark Physics Results From LHC*, arXiv:1201.5844 [hep-ph], January 2012.

[17] Lyndon Evans and Philip Bryant, *LHC Machine*, 2008 JINST 3 S08001.

[18] ATLAS Collaboration, *The ATLAS experiment at the CERN Large Hadron Collider*, 2008 JINST S08003 .

[19] CMS Collaboration, *The CMS experiment at the CERN LHC*, 2008 JINST 3 S08004.

[20] ALICE Collaboration, *The Alice experiment at the CERN LHC*, 2008 JINST 2 S08002.

[21] lhc-b collaboration, *The LHCb detector at the CERN LHC*, 2008 JINST 5 S08005.

[22] "CERN", *The accelerator complex*, http://home.web.cern.ch.

[23] F. Pandolfi, *Search for the Standard Model Higgs Boson in the H→ZZ→l+l-qq Decay Channel at CMS*, Springer International publishing, 2013.

[24] E. Widl and R. Frühwirth, *A large-scale application of the Kalman alignment algorithm to the CMS tracker*, 2008 J.phys.: Conf.Ser.119 032038.

[25] The CMS Trigger and Data Acquisition group, *The CMS high level trigger*, arXiv:hep-ex/0512077-v1.

[26] The CMS Collaboration, *Particle-flow event reconstruction in CMS and performance for jets, Taus and* E_T^{miss}, CMS PAS PFT-09-001, 28 April 2009 (Update 3 April 2013).

[27] Achintya Rao, *Search for microscopic black hole signatures at the Large Hadron Collider*, cms.web.cern.ch, 15 December 2010.

[28] P. Schieferdecker (KIT), *Jet algorithms*, Vivian's meeting April 17 2009.

[29] Y.L. Dokshitzer, G. D. Leder, S. Moretti and B. R. Webber, JHEP 9708, 001 (1997) [hep-ph/9707323]; M. Wobisch and T. Wengler, hep-ph/9907280.

[30] M Cacciari, G. P. Salam, G. Soyez, *The anti-k_t jet clustering algorithm*, arXiv:0802.1189v2 [hep-ph], 21 April 2008..

[31] M Cacciari, G. P. Salam, G. Soyez, *FastJet 3.0.0 user manual*.

[32] Quantum Diaries, *When Feynman Diagrams fail*, www.quantumdiaries.org, 11 December 2010.

[33] J. Alwall et al., *MadGraph 5: Going Beyond*, JHEP **06** (2011) 128, doi:10.1007/JHEP06(2011)128, arXiv:0709.2092.

[34] T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 Physics and Manual*; JHEP **05** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.

[35] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, JHEP"**11** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

[36] The CMS Collaboration, *Performance of b tagging at $\sqrt{s}$ = 8 TeV in multijet, $t\bar{t}$ and boosted topology events*, CMS PAS BTV-13-001, 15 August 2013.

[37] J. D'Hondt, *Statistiek voor fysici*, Interuniversitair Instituut voor Hoge Energieën, Vrije universiteit Brussel, 2006.

Summary

At elementary particle level, the standard model (SM) provides a description for all of the fundamental interactions, except for gravity. Up to now the SM has been successfully tested but unfortunately it does not describe all of the observed phenomena and is only valid for interactions up to the TeV scale. To provide new insights in physics beyond the TeV scale and search for new physics phenomena, the LHC was built, colliding protons at centre of mass energies of 7 and 8 TeV. To study the interactions at the LHC physicists use large detectors like the Compact Muon Solenoid (CMS) experiment. Due to the high energy of the collisions, top quarks will be produced at a very high rate.
In this thesis, a study of the modelling of these top quarks is performed using the reconstructed top quark mass. Events are selected corresponding to the semimuonic decay channel of $t\bar{t}$ events, at an interaction energy of $\sqrt{s}$ = 8 TeV. The top quark mass is estimated as a function of cuts on the transverse momentum, p_T, of the jet constituents where-after the jets are reclustered. As a result, the estimated top quark mass changes for each applied threshold. The slope of the top quark mass evolution is determined as the difference of the top quark mass estimation for two different p_T thresholds applied on the constituents. This slope is obtained for each simulated model and the data. We found that none of $t\bar{t}$ simulated samples models correctly the slope of the data because the transverse momentum distribution of the jet constituents is not accurately modelled.

Samenvatting

Het standaard model (SM) van de deeltjes fysica beschrijft alle fundamentele interacties op elementair niveau, behalve zwaartekracht. Tot op dit moment is het standaard model uitvoerig en met succes getest. Dit model beschrijft echter niet alle fenomenen die worden waargenomen en is het model enkel geldig tot op de TeV schaal. Met het oog op het verwerven van nieuwe inzichten in fysica die niet worden beschreven door het standaard model is de large Hadron Collider (LHC) gebouwd. Hier worden protonen versneld en op elkaar gebotst bij een massamiddelpuntsenergie van 7 en 8 TeV. Om de interacties van deze botsingen te bestuderen gebruiken fysici detectoren zoals het Compact Muon Solenoid (CMS) experiment. Door de hoge energie die gecreëerd wordt in de botsingen worden top quarks in grote aantallen geproduceerd.
In deze thesis wordt een studie van de modellering van top quarks uitgevoerd aan de hand van hun gereconstrueerde massa. Gebeurtenissen worden geselecteerd overeenkomstig met het semimuonische verval kanaal van de $t\bar{t}$ gebeurtenissen, aan een energie van $\sqrt{s}$ = 8 TeV. De massa van de top quark wordt geschat als functie van sneden toegepast op het transversaal moment, p_T van de deeltjes waaruit de jets bestaan. Hierna worden de jets opnieuw geclusterd. Hierdoor verandert de top quark massa voor elke toegepaste snede. Het verschil in de geschatte massa van de top quark tussen twee p_T sneden op de deeltjes kan bepaald worden om de gesimuleerde modellen te vergelijken met de data. Het verschil komt voor geen van de modellen overeen met het verschil van de data. Dit komt omdat de verdeling van de transverse impuls van de deeltjes waaruit de jets bestaan niet juist gesimuleerd worden in de modellen.

Acknowledgements

As a final part of my thesis I would like to show my gratitude to a number of people who helped me in one way or another to establish this thesis. First of all I would like to thank my promotor professor Jorgen D'Hondt for providing me with the interesting topic for my thesis research and for his guidance during this project and the whole of my studies. I also want to thank doctor Petra Van Mulders, my co-promotor for her guidance, help and understanding during this year and for the advice on the writing of my thesis report. In addition I want to htank Petra for the many discussions on several physics topics and the advice for the analysis in my thesis.
I also want to thank all of my fellow students for their help and the good times we had during the past years. Also all people of the IIHE, especifically the people working in the top quark group for their help and advise during the past year.
Finally I would like to thank my family, all of my friends and everyone who supported me during my studies. Especially my parents for their support and motivation during my study and this final project.