Vrije Universiteit Brussel

Faculteit Wetenschappen en Bio-ingenieurswetenschappen
Vakgroep Fysica

Measurement of the top-quark mass and the mass difference between top and antitop quarks at the LHC

Stijn Blyweert

Promotor: Prof. Dr. Jorgen D'Hondt

Proefschrift ingediend met het oog op het behalen van
de academische graad van Doctor in de Wetenschappen

Augustus 2013

Doctoral examination commission:

Prof. Dr. J. D'Hondt (Vrije Universiteit Brussel)
Prof. Dr. C. De Clercq (Vrije Universiteit Brussel)
Prof. Dr. F. Blekman (Vrije Universiteit Brussel)
Prof. Dr. S. Lowette (Vrije Universiteit Brussel)
Prof. Dr. B. Craps (Vrije Universiteit Brussel)
Prof. Dr. P. Claeys (Vrije Universiteit Brussel)
Dr. Martijn Mulders (CERN)
Prof. Dr. A. Onofre (Universidade de Minho)

Contents

Introduction

The most fundamental aspects of the building blocks of our universe are studied within the field of elementary particle physics. Our current understanding of these particles and the interactions amongst them is summarised in a theory called the Standard Model. This theory has been experimentally verified with a great precision and all results are still in agreement with its predictions. One of the most important achievements of the theory was the successful prediction of the existence of the so-called Brout-Englert-Higgs boson, which was discovered about a year ago. Although the theory has been proven to be extremely successful, it also has a number of shortcomings. Gravity, for example, is not included in this theory. Therefore further experimental tests of the Standard Model's predictions need to be carried out.

This was one of the main motivations for the construction of the Large Hadron Collider at CERN near Geneva, which is a 27 km long circular particle accelerator designed to deliver proton-proton collisions at record-breaking centre-of-mass energies up to 14 TeV. Within this thesis top quark pair events, which were produced during the 2010-2012 run of the LHC at centre-of-mass energies of 7 and 8 TeV and recorded by the CMS detector, are used in a detailed study of the top-quark mass.

The exact value of the top-quark mass is a very important parameter of the Standard Model, since it is used for internal consistency checks of the theory. Therefore a very precise measurement of the top-quark mass is performed in this thesis based on the Ideogram method. To study the sensitivity of the measured top-quark mass to the theoretical modelling of $t\bar{t}$ events a measurement of this quantity is also performed in bins of several kinematic variables and compared to different theoretical predictions. Finally one of the fundamental symmetries within particle physics, the CPT symmetry, is tested by measuring the difference in mass between the top and the antitop quark.

In Chapter 1 the theoretical framework of the Standard Model is introduced, together with an overview of the current experimental status within the field of top-quark physics. The experimental setup used to perform the measurements within this thesis, namely the Large Hadron Collider and the CMS detector which are both located at CERN, are discussed in Chapter 2. An overview of the generation and simulation of proton-proton collisions is given in Chapter 3, together with the different reconstruction algorithms used in this thesis which translate electronic hits in the CMS detector into the measurements of physical objects like electrons and muons. The selection of top-quark events and the reconstruction of the top-quark mass in every collision event is explained in Chapter 4. In Chapter 5 the technique to measure the top-quark mass is introduced and the results which are obtained when this technique is applied to the data are discussed. The optimisation of this measurement technique and the measurement in bins of kinematic variables are also given here.

The measurement of the mass difference between the top and the antitop quark based on a similar technique is discussed in Chapter 6. Finally in Chapter 7 an overview of the results is given together with a comparison to previous measurements of these quantities and some projections how these measurements can be further improved.

Some of the results in this thesis have already been published. This includes the measurement of the mass difference between the top and the antitop quark with the 7 TeV dataset [1]. A preliminary result was also made public based on the 8 TeV dataset [2]. The publication of this measurement is in progress.

Chapter 1

Top-quark physics within the Standard Model

One of the most extensively studied areas of physics during the last decades was the quest for a so-called 'theory of everything': one theory which could describe the entire particle content of our universe together with the interactions amongst them at all possible energies. There is a theory which does describe all interactions taking place at energies below 1 TeV: the Standard Model of elementary particle physics. In Section 1.1 a description of the particles in this theory and the forces among them is given, together with its main properties relevant for this thesis as well as its shortcomings.

The heaviest particle within the Standard Model, which is also the most recently discovered quark, is the top quark. The importance of a detailed study of this quark can be found in Section 1.2. In this section an overview of the current theoretical and experimental status within the field of top-quark physics is given.

1.1 The Standard Model of elementary particle physics

Within the Standard Model of elementary particle physics [3–8] all currently known elementary particles and their mutual interactions are described in the framework of relativistic quantum field theories. An overview of these particles and their interactions are described in Section 1.1.1. Section 1.1.2 introduces the theoretical framework of the Standard Model; while Section 1.1.3 introduces the Brout-Englert-Higgs mechanism which is believed to be responsible for the mass of all elementary particles. One of the fundamental symmetries of the Standard Model, the CPT symmetry, is discussed in Section 1.1.4. The shortcomings of this model are described in Section 1.1.5, together with some possible extensions of the Standard Model.

1.1.1 The Standard Model particles and their mutual interactions

According to the Standard Model, all known matter is composed of 12 fermions[1] with spin 1/2. An overview of these 12 particles can be found in Table 1.1. Each of these fermions

Generation	Quarks		Leptons	
1st	up u	down d	electron neutrino ν_e	electron e^-
2nd	charm c	strange s	muon neutrino ν_μ	muon μ^-
3rd	top t	bottom b	tau neutrino ν_τ	tau τ^-
Electrical charge	+2/3	-1/3	0	-1

Table 1.1: An overview of the fermions in the Standard Model and their corresponding electrical charge.

also has a corresponding anti-particle, with the same quantum numbers but an opposite electrical charge. These anti-particles are denoted with a bar over their symbol (e.g. $\bar{f}$ is the anti-particle of a fermion f), except for the charged leptons. The anti-particles of the electron, muon and tau are the positron e^+, the anti-muon μ^+ and the anti-tau τ^+ respectively.

The Standard Model contains 3 generations of particles, where each subsequent generation is a heavier copy of the previous one. Within each generation, there are two types of particles: quarks and leptons. Quarks are particles which interact via the electromagnetic, weak and strong force, while leptons are particles which interact only via the electromagnetic and/or the weak force: the neutral leptons (neutrino's) do not interact via the electromagnetic force. All visible stable matter surrounding us is made out of particles from the first generation. A proton is formed by two up quarks and one down quark, while a neutron is formed by one up quark and two down quarks. Together with the electron, these protons and neutrons bind together into all different kinds of atoms.

All the interactions between these fermions are mediated via so-called force-carrying particles. These bosons[2] all have spin 0 or 1. The bosons within the Standard Model are listed in Table 1.2.

First there is the massless gluon, which is the carrier of the strong force. Secondly there is the massless photon, carrying the electromagnetic force, and thirdly there are the massive $W^\pm$ and Z^0 bosons, carrying the weak force. The Standard Model also predicts the existence of an additional boson: the Brout-Englert-Higgs (BEH) boson, which is responsible for giving mass to all other particles via the mechanism of electroweak symmetry breaking. This particle was discovered only very recently [11, 12] and will be discussed in Section 1.1.3.

[1] Fermions are particles with a half-integer spin

[2] Bosons are particles with an integer spin

Boson	Mass (GeV)	Spin
gluon g	0	1
photon γ	0	1
W^+ and W^-	80.385 ± 0.015	1
Z^0	91.1876 ± 0.0021	1
H^0	125.7 ± 0.4	0

Table 1.2: An overview of the bosons in the Standard Model, their masses and their spins [9, 10]. The spin of the recently discovered H^0 boson is still hypothetical, although current results are compatible with the spin-zero hypothesis.

1.1.2 Theoretical framework of the Standard Model

The theoretical framework which forms the mathematical description of the Standard Model is that of relativistic quantum field theories [3, 4]. As suggested by the name of this theoretical framework, it combines two of the greatest achievements within physics in the 20$^{\text{th}}$ century: quantum mechanics and special relativity. In this section the Lagrangian of the Standard Model is built.

Fermions and their interactions

Fermions within the Standard Model are represented by a Dirac-spinor field ψ. The Dirac Lagrangian, representing a free fermion, is defined as

$$\mathcal{L}_{Dirac} = i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi. \tag{1.1}$$

In this equation γ_{μ} are the Dirac matrices, anti-fermions are represented by $\bar{\psi}$ and m is the mass of the fermions.

Following the principle of gauge invariance the Dirac Lagrangian is required to be invariant under a local phase transformation

$$\psi' = U\psi = e^{i\vec{\epsilon}(x)\cdot\frac{\vec{\tau}}{2}}\psi \tag{1.2}$$

with rotation parameters $\vec{\epsilon}(x)$ and generators of a given Lie-group $\vec{\tau}$. This invariance can be enforced by the introduction of a covariant derivative

$$\mathcal{D}_{\mu} = \partial_{\mu} - ig\frac{\vec{\tau}}{2}\cdot\vec{A}_{\mu}, \tag{1.3}$$

where g is proportional to the interaction strength and is called the coupling constant, and where $\vec{A}_{\mu}$ are the interacting vector gauge fields. This covariant derivative then transforms the Dirac Lagrangian into

$$\begin{aligned}\mathcal{L} &= i\bar{\psi}\gamma^{\mu}\mathcal{D}_{\mu}\psi - m\bar{\psi}\psi \\ &= i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi + g\bar{\psi}\gamma^{\mu}\frac{\vec{\tau}}{2}\cdot\vec{A}_{\mu}\psi.\end{aligned} \tag{1.4}$$

The last term in this equation describes the coupling between the fermion field and these new gauge fields.

This clearly shows that the requirement of a theory to be invariant under a local phase transformation introduces additional vector fields which are responsible for the interactions within this theory. The gauge transformations can be based on both Abelian or non-Abelian Lie groups[3]. In the Abelian case only interactions between the fermions and the gauge fields are allowed, while in the non-Abelian case also couplings among the gauge fields themselves are present.

The fundamental interactions of the Standard Model

As explained in Section 1.1.1, the Standard Model contains three fundamental interactions: the strong force, the weak force and the electromagnetic force. As explained in the previous section, these interactions are introduced by requiring the theory to be invariant under local phase transformations. In the case of the Standard Model these transformations are generated by the gauge group

$$G_{SM} = SU(3)_C \otimes SU(2)_L \otimes U(1)_Y, \tag{1.5}$$

where the group $SU(3)_C$ introduces the strong interaction, and the group $SU(2)_L \otimes U(1)_Y$ introduces the unified electroweak interaction.

Quantum chromodynamics

Within the theory of quantum chromodynamics (QCD), the strong interaction is introduced by requiring the theory to be invariant under the local phase transformations generated by the non-Abelian group $SU(3)_C$. This introduces eight gauge fields G^a_μ, known as gluons, with $a = 1 \dots 8$, and a covariant derivative

$$\mathcal{D}_\mu = \partial_\mu - ig_s \frac{\lambda^a}{2} G^a_\mu. \tag{1.6}$$

In this equation g_s is the strong coupling constant and λ^a are the eight Gell-Mann matrices. The gluons only interact with particles carrying a so-called colour charge C, which are only the quarks and the gluons themselves. The leptons do not carry colour charge so they do not interact with gluons and are considered as singlets under $SU(3)_C$.

The electroweak interaction

The remaining part of the Standard Model gauge group, $SU(2)_L \otimes U(1)_Y$, generates the electroweak force. The requirement of invariance under $U(1)_Y$ transformations introduces a single field B_μ; the same requirement under $SU(2)_L$ transformations generates three fields W^k_μ, with $k = 1 \dots 3$; and they require the introduction of a covariant derivative

$$\mathcal{D}_\mu = \partial_\mu - ig \frac{\tau^k}{2} W^k_\mu - ig' \frac{Y}{2} B_\mu \tag{1.7}$$

for the theory to be invariant under $SU(2)_L \otimes U(1)_Y$ local phase transformations. Here g and g' are the coupling constants, τ^k are the Pauli matrices and Y is the

[3] An Abelian group is defined as a group with commutating generators $\vec{\tau}$: $[\tau_i, \tau_j] = \tau_i\tau_j - \tau_j\tau_i = 0$

hypercharge. To incorporate the observed parity violation in the weak interaction, the three fields W_μ^k can only couple to left-handed fermions[4].
Linear combinations of the fields B_μ and W_μ^k correspond to the experimentally observed bosons. First there is the photon γ

$$A_\mu = W_\mu^3 \sin\theta_W + B_\mu \cos\theta_W, \tag{1.8}$$

secondly there is the neutral Z^0 boson

$$Z_\mu^0 = W_\mu^3 \cos\theta_W - B_\mu \sin\theta_W, \tag{1.9}$$

and thirdly there are the $W^\pm$ bosons

$$W_\mu^\pm = \sqrt{\frac{1}{2}} \left(W_\mu^1 \mp i W_\mu^2 \right). \tag{1.10}$$

In these equations the Weinberg angle θ_W is defined as

$$\tan\theta_W = \frac{g'}{g}. \tag{1.11}$$

Since the Z^0 and $W^\pm$ bosons have a mass different from zero, additional mass terms for these need to be added to the Lagrangian, but this would break the gauge invariance of the Lagrangian. To solve this problem a procedure to spontaneously break the symmetry is introduced, as explained in Section 1.1.3.

As it is experimentally observed that for quarks the mass (or strong force) eigenstates differ slightly from the eigenstates under the weak interaction, a matrix transforming the mass eigenstates to the weak eigenstates is needed:

$$\begin{pmatrix} d^{weak} \\ s^{weak} \\ b^{weak} \end{pmatrix}_L = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L. \tag{1.12}$$

This matrix is called the Cabibbo-Kobayashi-Maskawa or CKM matrix. The elements of this matrix $|V_{qq'}|$ are related to the probability of a quark q to decay into another quark q' via the weak interaction.

1.1.3 The Brout-Englert-Higgs mechanism and the corresponding boson

The local gauge invariance forbids the introduction of mass terms for bosons and also for the fermions, as explained in the previous Section. To accommodate the observed massive Z^0 and $W^\pm$ bosons within the Standard Model the Brout-Englert-Higgs (BEH)

[4]Right-handed fermions are defined by $\psi_R = \frac{1}{2}(1+\gamma_5)\psi$ and they have their spin lined up to the direction of motion, while left-handed fermions are defined by $\psi_L = \frac{1}{2}(1-\gamma_5)\psi$ and these have their spin opposite to the direction of motion. The γ_5 matrix in these equations is defined by $\gamma_5 = i\gamma_0\gamma_1\gamma_2\gamma_3$

mechanism [13–15] is used, where a scalar field is introduced which leaves the Lagrangian invariant but which breaks the symmetry of the vacuum state.

The electroweak symmetry within the Standard Model is broken by the introduction of a scalar field Φ that is an electroweak doublet

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \tag{1.13}$$

and where ϕ^+ and ϕ^0 are both complex scalar fields. With these field the following gauge invariant term can then be added to the Standard Model Lagrangian

$$\begin{aligned} \mathcal{L}_{BEH} &= (\mathcal{D}^\mu \Phi)^\dagger \mathcal{D}_\mu \Phi - V(\Phi) \\ &= (\mathcal{D}^\mu \Phi)^\dagger \mathcal{D}_\mu \Phi - \mu^2 \left(\Phi^\dagger \Phi\right) - \lambda \left(\Phi^\dagger \Phi\right)^2, \end{aligned} \tag{1.14}$$

with μ^2 representing a mass parameter and $\lambda > 0$ the strength of the field's self interaction. By taking $\mu^2 < 0$ the potential $V(\Phi)$ reaches a non-unique minimum for

$$\Phi^\dagger \Phi = \frac{|\mu^2|}{2\lambda} \equiv \frac{v^2}{2}, \tag{1.15}$$

where v is the vacuum expectation value. A particular vacuum state is chosen and an expansion is made around this minimum

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}. \tag{1.16}$$

Here $H(x)$ is the only remaining field of Φ, called the BEH field. It has a spin 0 particle associated to it (the BEH boson H^0) with a mass $m_H = \sqrt{2\lambda} v$. The other three fields are absorbed by the Z^0 and $W^\pm$ bosons when acquiring a mass. When the $SU(2)_L \otimes U(1)_Y$ covariant derivative of Equation 1.7 is entered into Equation 1.14 they obtain masses

$$m_W = \frac{1}{2} v g \qquad\qquad m_Z = \frac{1}{2} v \sqrt{g^2 + g'^2}. \tag{1.17}$$

The masses of the fermions, on the other hand, are not generated in a similar way; Yukawa coupling terms need to be added to the Lagrangian by hand. These terms describe the interaction between the fermion fields and the BEH field. They have the form

$$\mathcal{L}_{Yukawa} = g_{Yukawa} \Phi \bar{\psi} \psi, \tag{1.18}$$

where g_{Yukawa} is the Yukawa coupling constant for a fermion with mass $m = g_{Yukawa} v/\sqrt{2}$.

Until very recently, the existence of the BEH boson was the biggest unknown of the Standard Model, but this has changed completely in the summer of 2012 with the discovery of a new boson [11, 12] that is compatible with the Standard Model BEH boson.

1.1.4 CPT symmetry within the Standard Model

All local relativistic quantum field theories like the Standard Model are believed to possess another fundamental symmetry: the invariance of the theory under so-called CPT transformations [16]. This CPT transformation is a combination of three individual transformations:

the charge conjugation transformation C which converts a particle into its antiparticle, the parity transformation P which changes a right-handed coordinate system into a left-handed one or vice versa, and the time reversal transformation T which changes the direction of flow of time.

The invariance under CPT transformations of these local relativistic quantum field theories is called the *CPT theorem*. It states that local quantum field theories which are invariant under Lorentz transformations (the so-called relativistic quantum field theories) must also possess the CPT symmetry. From this theorem it can also be shown that a particle and its antiparticle must have certain identical properties, like its mass, lifetime and size of its electrical charge. This means that if CPT violation would be observed (e.g. by the measurement of a difference in mass between a particle and its corresponding antiparticle), this would be a clear sign of physics not described by the Standard Model.

Multiple experimental tests of CPT symmetry have been conducted, but no significant violation of CPT symmetry has been observed so far. One of the most stringent limits on CPT violation are set by using neutral kaons, which are mesonic bound states of a d and an $\bar{s}$ quark. From the measured neutral kaon decay rates the particle-antiparticle mass difference in the kaon system is estimated to be $|m_{K^0} - m_{\bar{K}^0}|/m_{average} < 6 \times 10^{-19}$ at 90% CL [9, 17, 18]. Although this extremely precise measurement is an important confirmation of CPT symmetry, this measurement is performed using bound states of fundamental particles and relies on theoretical calculations to extract the mass difference from the decay rates. Therefore direct measurement of the mass difference between a fundamental particle and its antiparticle are also performed since these would suffer less from these issues. An example is the measurement of the mass difference between the top and the antitop quark as discussed in Section 1.2.2.

1.1.5 Shortcomings of the Standard Model

Although the Standard Model of elementary particle physics is a very successful theory – all experimental results show compatibility with its predictions [19] – there are a number of important shortcomings within the Standard Model. A non-exhaustive list of shortcomings is given here.

Gravity Since the Standard Model does not include a description of gravity, it can never be a complete 'theory of everything'. The effects of gravity on the current experimental results are completely negligible so its inclusion is not mandatory at this point, but eventually the Standard Model will need to be modified with the addition of a quantum mechanical description of gravity.

Grand Unification Because of the unification of the electric and the magnetic force within electromagnetism and of the unification of the electromagnetic and the weak force within the electroweak interaction of the Standard Model, it is generally believed that the strong and the electroweak interaction should also be unified at some very large energy scale. Such a unification is not described by the Standard Model.

The hierarchy problem The typical scale of electroweak physics of ~ 100 GeV is multiple orders of magnitude smaller than the fundamental scale of gravity, the Planck scale of order $\sim 10^{19}$ GeV. The large discrepancy between both scales is called the hierarchy

problem. Because the Standard Model predicts no new physics up to the Planck scale (where the effects of gravity become important), the explanation of the BEH boson mass would require an extreme fine-tuning of all the parameters of the Standard Model.

Cosmological problems Multiple astrophysical observations have shown that our universe consists of 4.9% of ordinary matter, 26.8% of dark matter and 68.3% of dark energy [20]. The Standard Model does not include a proper dark matter candidate, and it can not explain what this peculiar dark energy actually is.

The aforementioned shortcomings of the Standard Model have led to the belief that there must exist extensions of the Standard Model. A very popular and promising extension is the inclusion of supersymmetry, which provides a link between bosons and fermions. Supersymmetric theories add to every particle a superpartner with the same properties but with a difference in spin of half a unit. These theories can provide a dark matter candidate, solve the hierarchy problem and make the electroweak and the strong forces unify at high energies. Many other possible extensions of the Standard Model exist, for example theories introducing extra dimensions or additional interactions. To verify experimentally if the Standard Model is still valid and to search for any of the predictions of these theories beyond the Standard Model, particle colliders at increasingly high energies are necessary.

1.2 Top-quark physics

Since its discovery in 1995 [21, 22] by the CDF [23] and DØ [24] experiments at the Tevatron collider [25], different aspects of the top quark have been studied in great detail. Also the ATLAS [26] and CMS [27] experiments located at the Large Hadron Collider [28] (LHC) have also performed very precise measurements within the field of top-quark physics. The current understanding of the physics related to the top quark will be discussed in this section. The production and decay of top quarks will be discussed in Section 1.2.1, while its main properties will be discussed in Section 1.2.2. Finally the importance of a precise measurement of the top-quark sector of the Standard Model will be discussed in Section 1.2.3.

1.2.1 Top-quark production and decay

Because of their heavy mass top quarks can only be produced via extremely energetic processes. The only two laboratories where they have been produced are the Tevatron and the LHC. The Tevatron delivered proton-antiproton collisions at a maximum centre-of-mass energy of 1.96 TeV, while the LHC produced proton-proton collisions at a maximum centre-of-mass energy of 8 TeV. In these collisions top quarks can either be produced singly or in a quark-antiquark pair (a $t\bar{t}$ pair). Since the pair production occurs more frequent, these processes are generally used in detailed studies of the top-quark properties.

Very precise theoretical predictions of the $t\bar{t}$ production cross-section are available [29, 30] for both the Tevatron and the LHC. The experiments located at both colliders have also conducted very precise measurements of these cross-sections [31–33]. These theoretical predictions are summarised in Table 1.3 together with the most precise experimental

measurements, where a good agreement can be seen between the theoretical predictions and the experimental measurements.

	$\sigma_{theoretical}$ (pb)	$\sigma_{experimental}$ (pb)
Tevatron: $p\bar{p}$ @ $\sqrt{s} = 1.96$ TeV	$7.16^{+0.39}_{-0.48}$	7.65 ± 0.42
LHC: pp @ $\sqrt{s} = 7$ TeV	$172.0^{+12.1}_{-13.4}$	$161.9^{+6.7}_{-6.6}$
LHC: pp @ $\sqrt{s} = 8$ TeV	$245.8^{+16.6}_{-18.7}$	227 ± 15

Table 1.3: Overview of the most precise theoretical predictions [29, 30] and experimental measurements [31–33] of the cross-section for $t\bar{t}$ production.

Due to its very large mass, the lifetime of the top quark is predicted to be extremely short: $\sim 5 \cdot 10^{-25} s$ [34]. This has been experimentally verified to be $\tau_t = (3.29^{+0.90}_{-0.63}) \times 10^{-25}$ s [35]. Its decay happens predominantly to a b quark and a W boson ($t \to W^+ b$ or $\bar{t} \to W^- \bar{b}$). The decay into lighter quarks ($t \to Ws$ or $t \to Wd$) is highly suppressed since the $|V_{tb}|$ element of the CKM matrix is measured to be very close to one: $|V_{tb}| > 0.972$ at 95% CL [36].

The W boson itself is known to be unstable and decays immediately into either a charged lepton and a neutrino ($W \to \ell^+ \nu_\ell$ with $\ell = e, \mu$ or τ) or into a quark and an antiquark (mainly $W \to u\bar{d}$ or $c\bar{s}$). Since each quark carries one of the three possible colour charges there are 6 possible hadronic decay channels, while there are only 3 leptonic decay channels. Therefore the W-boson has a leptonic branching ratio of $\sim 1/3$ and a hadronic branching ratio $\sim 2/3$,

This means that the decay of a $t\bar{t}$ pair can happen in three channels: the hadronic channel where both W bosons decay hadronically, the semi-leptonic channel where one W boson decays hadronically and the other decays leptonically, and the dilepton channel where both W bosons decay leptonically. The decay channel considered in this thesis is the semi-leptonic channel[5]

$$t\bar{t} \to WbWb \to q\bar{q}' b\ell\nu_\ell b, \tag{1.19}$$

where the lepton ℓ is a muon or an electron. This decay has a branching ratio of $\sim 14.8\%$ per lepton flavour. The light and b quarks formed in the final state of $t\bar{t}$ decays will not be observed directly. They will undergo the showering and hadronisation as explained in Chapter 3, which will result in collimated groups of final state particles called jets.

There have also been searches for top-quark decays into other channels. The branching ratio of flavor changing neutral current top-quark decays ($t \to Zq$) is measured to be smaller than 0.07% at 95% CL [37], while the branching ratios of the baryon number violating decay channels $t \to bc\mu$ and $t \to bue$ have been measured to be smaller than 0.16% and 0.17% at 95% CL, respectively [38].

All the experiments have also performed differential cross-section measurements [39–42] where the $t\bar{t}$ production cross-section is measured as a function of some variable, like for

[5]When a particle like a b quark is mentioned in this thesis its antiparticle is implicitly assumed as well.

example the number of additional jets, the transverse momentum and pseudorapidity[6] of the decay products, the transverse momentum and pseudorapidity of the $t\bar{t}$ system, etc. An example of these distributions can be seen in Figure 1.1 for the rapidity[7] of the $t\bar{t}$ system and the transverse momentum of the b jets produced during top-quark decay. In general a

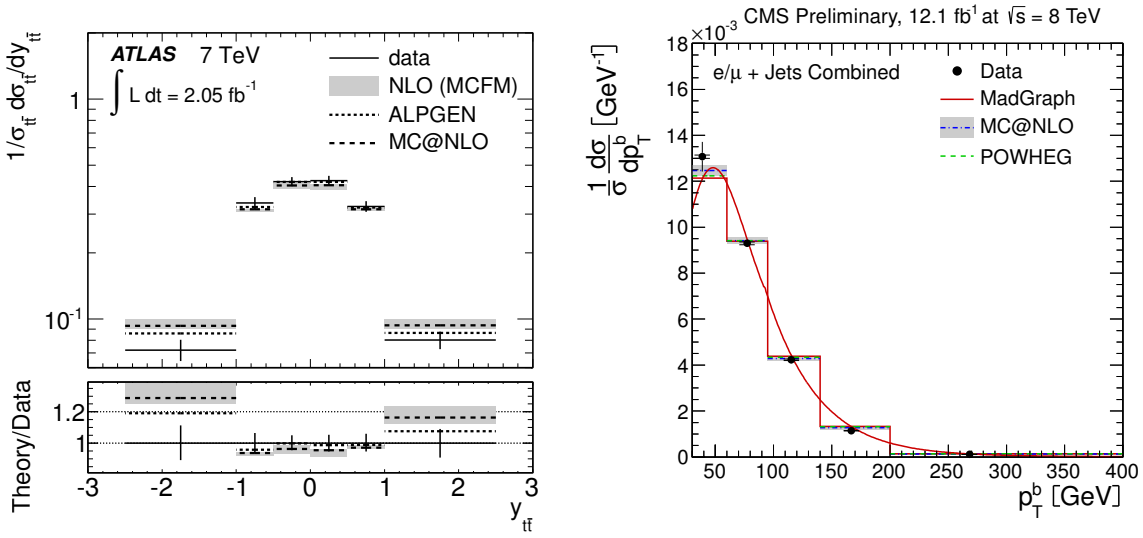

Figure 1.1: Differential $t\bar{t}$ production cross-section as a function of the rapidity of the $t\bar{t}$ system (left) and as a function of the transverse momentum of the b jets produced during top-quark decay (right).

decent agreement between the predicted differential cross-sections and the experimentally measured ones is found, which shows the capability of the Standard Model in providing a description of these processes.

The production of top quarks does not always happen in pairs. They can also be produced singly via the s-channel or t-channel exchange of a W boson or via the associated production of a top quark and a W boson. These processes all have a significantly lower production cross-section compared to pair production, making their study more difficult. In recent years significant improvements have been made in this sub-field with the discovery of single top-quark production and the precise measurement of the single top-quark production cross-section in both the s- and the t-channel by the Tevatron experiments [43, 44], resulting in a combined s+t-channel single top-quark production cross section of $\sigma_{s+t} = 4.11^{+0.60}_{-0.55}$ pb in $p\bar{p}$ collisions at a centre-of-mass energy of 1.96 TeV.

At the LHC the measurement of the t-channel cross-section [45–48] and the discovery of the associated tW production [49–51] has been performed. At a centre-of-mass energy of 7 TeV a cross-section of $\sigma_t = 67.2 \pm 6.1$ pb is measured, while at 8 TeV the measured cross-section is equal to $\sigma_t = 80.1 \pm 13.0$ pb. The associated tW production cross-section is measured to be equal to $\sigma_{tW} = 16^{+5}_{-4}$ pb and $\sigma_{tW} = 23.4^{+5.5}_{-5.4}$ pb at 7 and 8 TeV,

[6] The transverse momentum ($p_T = p \sin\theta$) is the momentum in the direction transverse to the beamline. The pseudorapidity is defined in terms of the polar angle θ relative to the counterclockwise-rotating beam as $\eta = -\ln(\tan\theta/2)$.

[7] The rapidity of a particle is defined as $y = \frac{1}{2}\ln\frac{E+p_z}{E-p_z}$, where p_z is the size of the momentum parallel to the counterclockwise-rotating beam.

respectively. The ratio $R_{t/\bar{t}}$ of the t-channel top to antitop quark production cross-sections is also measured at the LHC [52, 53], resulting in values of $R_{t/\bar{t}} = 1.81^{+0.23}_{-0.22}$ at 7 TeV and $R_{t/\bar{t}} = 1.76 \pm 0.27$ at 8 TeV. All these results are in good agreement with the Standard Model predictions.

1.2.2 Precise measurement of the top-quark properties

Together with the precise measurement of $t\bar{t}$ production and decay, several other top-quark properties have been measured. One of its main properties is the mass of the top quark which has been measured very precisely at the Tevatron and at the LHC. An overview of the measurements performed at the LHC can be found in Figure 1.2, while Figure 1.3

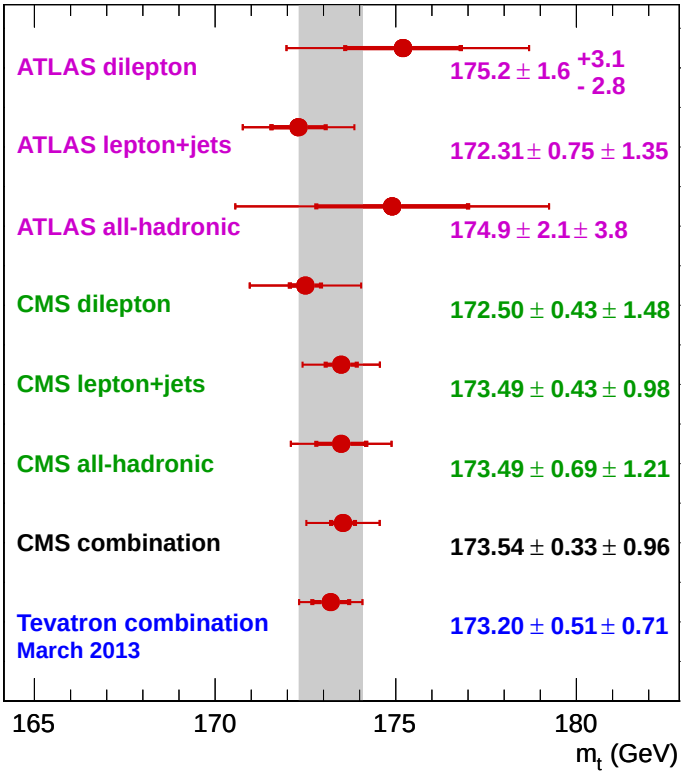

Figure 1.2: The latest top-quark mass measurements from the LHC experiments [54–60].

shows an overview of the Tevatron measurements of the top-quark mass. The most precise measurement is the combination of all the individual Tevatron measurements, which results in a value of

$$m_t = 173.20 \pm 0.51(\text{stat}) \pm 0.71(\text{syst}) \text{ GeV} = 173.20 \pm 0.87 \text{ GeV}, \quad (1.20)$$

corresponding to a relative precision of 0.50%. A good agreement between the measurements performed at the Tevatron and the LHC is also observed, although the LHC measurements have a slightly larger uncertainty.

Similar techniques as used in the top-quark mass measurements are also used to measure the difference in mass between the top and the antitop quark ($\Delta m_t = m_t - m_{\bar{t}}$)

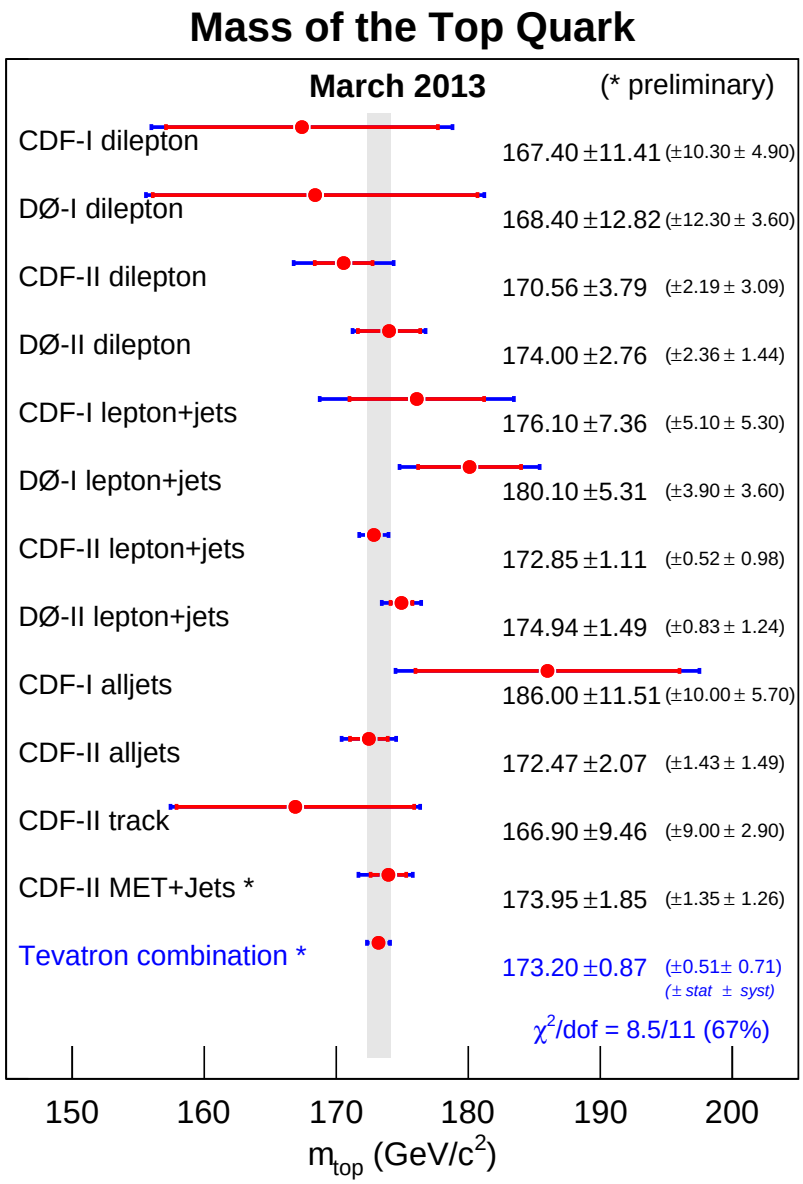

Figure 1.3: The latest top-quark mass measurements from the Tevatron experiments and their combination [61].

which provides a check of the CPT symmetry of quantum field theories, as explained in Section 1.1.4. These measurements are summarised in Table 1.4, where a good agreement with the expectation from CPT symmetry is found: no measurement shows a significant deviation from zero.

Due to its very short lifetime the spin information of the top quark is conserved and transferred to its decay products. To study this, precise measurements of the spin correlation between both top quarks in $t\bar{t}$ pair production and of the polarisation of these top quarks have been made. The observation of spin correlations in the $t\bar{t}$ system was only made quite recently [64, 65] and showed good compatibility with the expectations from the Standard Model. Also the top-quark polarisation has been measured [66, 67], and a good agreement

	Δm_t (GeV)
DØ, 3.6 fb^{-1}	0.8 ± 1.8 (stat) ± 0.5 (syst)
CDF, 8.7 fb^{-1}	-1.95 ± 1.11 (stat) ± 0.59 (syst)
CMS, 4.7 fb^{-1}	-0.44 ± 0.46 (stat) ± 0.27 (syst)

Table 1.4: Overview of the most precise measurements of the mass difference between the top and the antitop quark [1, 62, 63].

with the Standard Model was found.

Another important property of the top quark is its electrical charge. Since the top quark is known to decay into a b quark and a W boson, only two options for the charge of the top quark are possible: $2/3$ (as expected from the Standard Model) and $-4/3$. The exotic top-quark charge of $-4/3$ has been completely excluded by the experimental observations [68, 69].

A detailed study of the Wtb-coupling is also performed by studying the W-boson helicity fractions. The fractions of events which contain W bosons with longitudinal, left-handed and right-handed polarisation are measured to be 0.626 ± 0.059, 0.359 ± 0.035 and 0.015 ± 0.034, respectively [70], which is in good agreement with the expectations from the Standard Model.

1.2.3 Importance of the top-quark sector

Since the discovery of the BEH boson and the corresponding measurement of its mass m_H, all the unknown parameters of the Standard Model have been measured. This means that now a global consistency check of the Standard Model can be performed, providing a test of the relations between the parameters of the Standard Model [19]. Such a global electroweak fit results in a p-value of 0.07 for the compatibility of the Standard Model with the current experimental observations.

The compatibility of the measured masses of the top quark and the W boson with all other experimental observations can also be checked, as is shown in Figure 1.4. Here the electroweak fit is repeated two times: once without the inclusion of the measured top-quark and W-boson masses (resulting in the blue band) and once without the inclusion of the measured top-quark, W-boson and BEH boson masses (resulting in the grey band). In both cases the coloured bands show the predicted values of the W-boson and the top-quark masses. From the difference in size between the grey and the blue band it is clear that the measurement of the mass of the BEH boson provides a significant improvement in the precision of these predictions. A comparison of the blue band with the black point (the measured top-quark and W-boson masses) shows a decent agreement amongst them. It also clearly shows that an improved measurement of these masses is very important in checking the consistency of the Standard Model.

Apart from the mass measurement, top quarks can also be used for other important cross-checks of the Standard Model. First of all they can be used to test the CPT symmetry

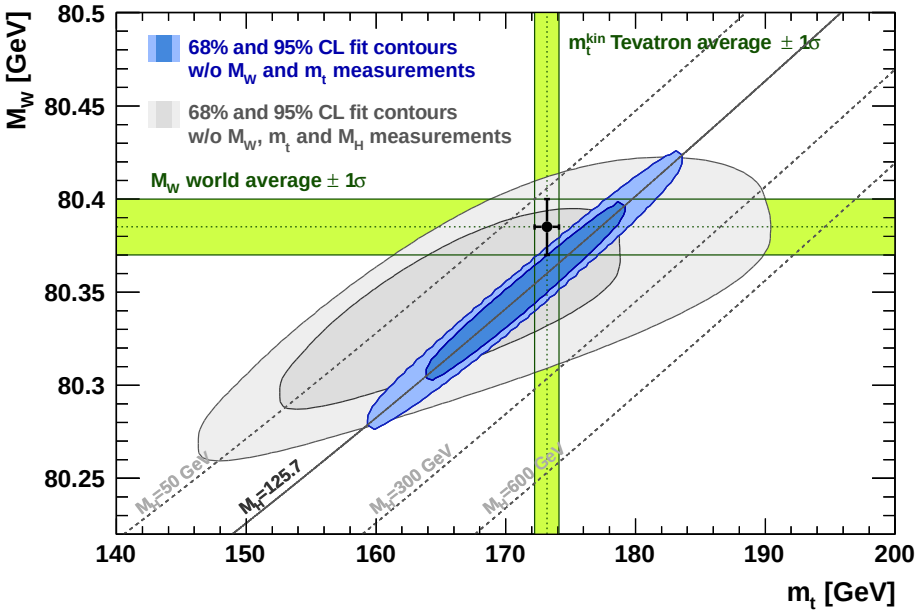

Figure 1.4: Result of the electroweak fit showing the compatibility of the measured top-quark and W-boson mass with all other experimental measurements used as input for the fit.

in the Standard Model by measuring the difference in mass between top and antitop quarks. Secondly they provide a direct probe for anomalous couplings in the Wtb vertex, which can be studied with the angular distributions of the decay products. Thirdly they can be used for all kinds of searches for physics beyond the Standard Model, like for example unobserved top-quark decay channels, or heavy particles decaying into one or more top quarks. Finally, top quarks are also a very important background in several searches for physics beyond the Standard Model as well as in the precise measurement of the properties of the BEH boson. Therefore a profound understanding of the top-quark sector will result in improved searches for new physics.

Another important aspect of top quarks is their usage as a calibration tool. Because of their decay into a b quark and a W boson they provide a sample of b jets which can be used for data-driven measurements of the b-jet identification efficiency [71, 72]. An additional interesting possibility is the measurement of the jet energy scale of light-quark jets and b-jets using the $t\bar{t}$ decay products [73].

Chapter 2

The CMS detector at CERN's Large Hadron Collider

As discussed in the previous section, a good agreement is found between experimental measurements and the predictions from the Standard Model. To test the consistency of the Standard Model more precisely and to conclude on the existence of the BEH boson as well as to search for possible phenomena beyond the Standard Model a new particle collider was built at CERN (European Organisation for Nuclear Research) [74], the Large Hadron Collider (LHC) [28], which delivers proton-proton collisions at the highest energies. In Section 2.1 more details about the LHC will be discussed.

Around the collision points immense particle detectors were constructed to allow a precise measurement of the LHC's proton-proton collisions. In this thesis the data recorded by the Compact Muon Solenoid (CMS) detector [27] is used. A detailed description of this experiment can be found in Section 2.2.

2.1 The Large Hadron Collider

The LHC was constructed during the previous decade in the tunnel which also housed the Large Electron-Positron collider [75]. This is a quasi-circular tunnel with a circumference of 26.7 km that was excavated between 1984 and 1989. It is located on the border between France and Switzerland between the lake of Geneva and the Jura mountains, where it lies between 45 m and 170 m below the surface of the Earth. The design of the LHC will be discussed in Section 2.1.1 together with its operation. All the different experiments using the LHC's proton-proton collisions to perform certain measurements are described in Section 2.1.2.

2.1.1 The design and operation of the LHC

Since the LHC's main goal is to allow a detailed study of the interactions amongst elementary particles and the search for new physics at the highest possible energy scale, it was designed to accelerate protons up to an energy of 7 TeV and to deliver proton-proton collisions at a centre-of-mass energy of 14 TeV. This energy is a factor 7 larger than at the Tevatron collider [25], which was the previous highest-energy collider.

The LHC's design

Because it was decided to build the LHC in the existing tunnel of the LEP collider, stringent requirements were put on the magnets which need to bend the proton beams onto their orbit through the already existing tunnel. To keep the 7 TeV proton beams onto their orbit, these magnets need to produce a magnetic field of 8.33 T. To produce such a high magnetic field superconducting magnets are used that operate at a temperature of 1.9 K.

Since the LHC collides protons with protons – and not protons with anti-protons as was done at the Tevatron – two separate beam pipes with opposite magnetic fields were needed. Another important limiting factor was the limited inner diameter of the tunnel, 3.7 m, which led to the choice of a twin-bore magnet design as can be seen in Figure 2.1. Here the two beam pipes are clearly visible, each surrounded with superconducting coils to

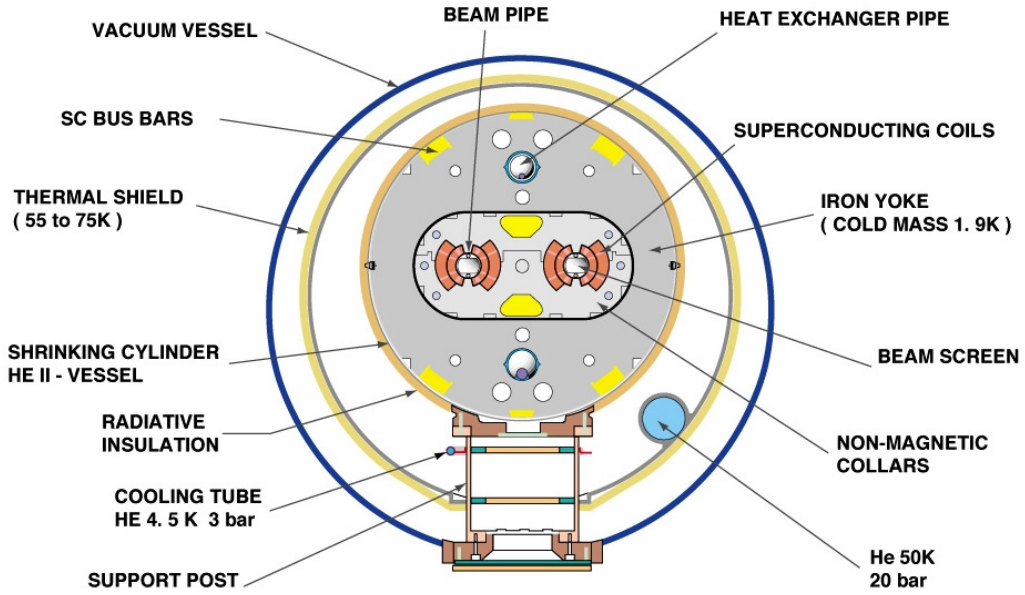

Figure 2.1: A schematic cross-section of a LHC dipole magnet.

produce the magnetic fields. In total 1232 of these $\sim$ 15 m long dipole magnets are used, each weighing about 27.5 tons, together with many other magnets for the focusing and steering of the proton beams.

In total the LHC has four interaction regions. These are regions where both beams cross each other and where protons are brought into collisions. Large particle detectors are constructed around these collision points. Their aim is to precisely measure all the particles produced when collisions take place. More information about these detectors is given in Section 2.1.2.

Apart from the centre-of-mass energy, another important aspect to conduct detailed studies is the number of collisions produced per second. In a particle collider the number of events of a certain process (e.g. $t\bar{t}$ events) produced per second is given by

$$\frac{dN_{process}}{dt} = \mathcal{L}\ \sigma_{process}, \tag{2.1}$$

where $\mathcal{L}$ is the instantaneous luminosity produced by the collider and $\sigma_{process}$ is the cross-section of this particular process. To increase this event rate the LHC was designed to deliver an instantaneous luminosity of 10^{34} cm^{-1}s^{-2}, which is almost two orders of magnitude larger than the maximum instantaneous luminosity achieved by the Tevatron.

In order to achieve such a high luminosity the LHC was designed to collide bunches of 10^{11} protons every 25 ns, which is equivalent to 40 million bunch-crossings per second. In total each LHC beam consists out of 2808 bunches of protons. Apart from this high bunch-crossing rate the luminosity is also increased by a strong focusing of the colliding bunches of protons, resulting in multiple proton-proton collisions in the same bunch-crossing. These additional proton-proton interactions taking place during the same bunch-crossing as the process under study (e.g. $pp \rightarrow t\bar{t}$) are called in-time pileup interactions. Because of the very short time between two bunch crossings the signals in some detectors are also influenced by the bunch crossings immediately preceding or following the primary crossing. This effect is called out-of-time pileup.

The accelerator complex located at CERN

Before protons are ready to be injected into the LHC, they first have to go through a number of smaller particle accelerators which gradually increase their energy and which produce the proton bunches which are separated by 25 ns. An overview of the entire accelerator complex at CERN is shown in Figure 2.2. Here the LHC with its four interaction regions (the yellow dots) is clearly visible.

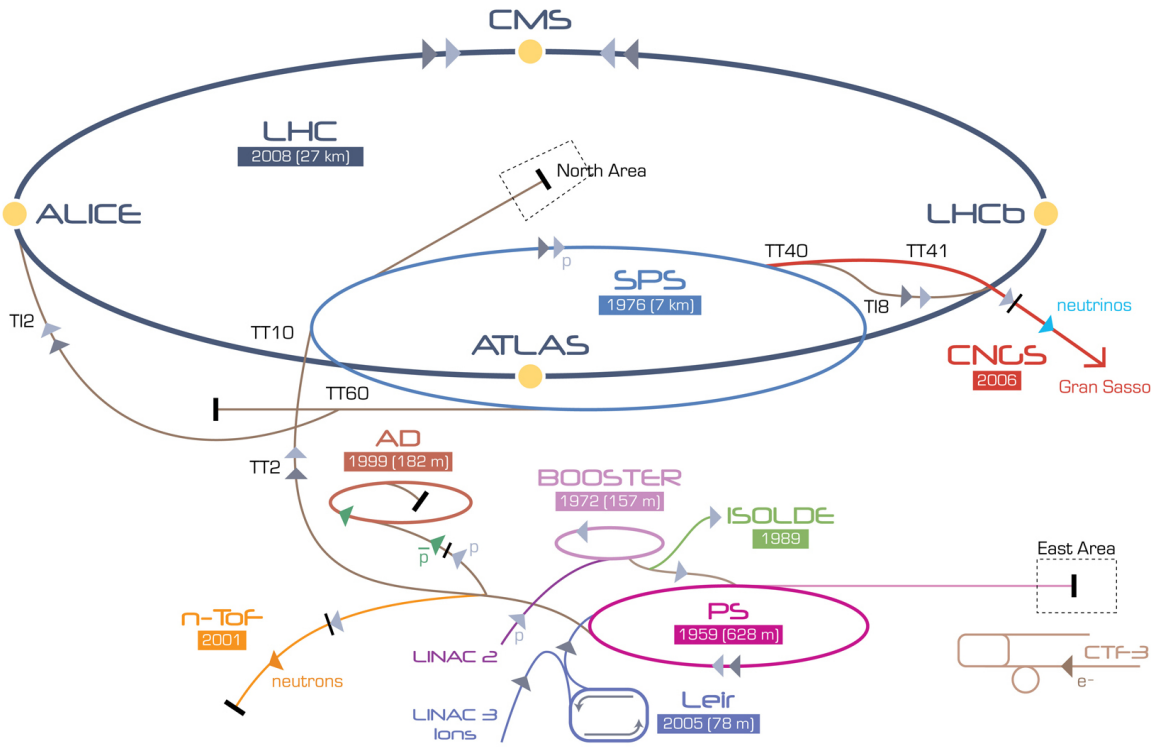

Figure 2.2: Overview of CERN's accelerator complex.

First there is the LINAC2, a linear particle accelerator, delivering 50 MeV protons to

the Proton Synchrotron Booster. The Booster accelerates these protons further until they reach an energy of 1.4 GeV and can be injected into the Proton Synchrotron (PS). Here the bunch trains with the 25 ns spacing are produced and these proton bunches are then accelerated until they reach an energy of 26 GeV. Finally these 26 GeV protons are injected into the Super Proton Synchrotron (SPS), where their energy is increased to 450 GeV and from which they are directly injected into the LHC. The final acceleration is done by the LHC, which also brings the protons into collision with each other.

Apart from these high-energy proton-proton collisions, this accelerator complex is also used for many other experiments. The LHC has also been used successfully to deliver the most energetic collisions between lead ions and between protons and lead ions. With the proton beams other experiments in the field of for example nuclear physics and neutrino physics are performed.

The 2010-2012 run of the LHC

The first protons were circulating in the LHC during September 2008, but due to some technical incidents, the first high-energy proton-proton collisions were only delivered on 30 March 2010 at a centre-of-mass energy of 7 TeV (3.5 TeV per proton beam), which was the highest energy ever achieved. After these successful proton-proton collisions the LHC ran in 2010 and 2011 at a centre-of-mass energy of 7 TeV; and in 2012 at a centre-of-mass energy of 8 TeV. In between the proton-proton runs the LHC also delivered collisions between lead ions and proton-lead collisions.

During its first year, 2010, the LHC delivered $\sim$ 45 pb^{-1} to its main experiments achieving a maximal instantaneous luminosity of $2.04 \cdot 10^{-32}$ $s^{-1}cm^{-2}$, which is a factor 50 lower than the nominal instantaneous luminosity. The year after the delivered luminosity was significantly larger, up to 6.14 fb^{-1}. The maximal instantaneous luminosity was also significantly increased to a value of $4.02 \cdot 10^{-33}$ $s^{-1}cm^{-2}$, which was the highest luminosity ever reached at a hadron collider. During the year 2012 the LHC delivered again more luminosity to its main experiments – up to 23.27 fb^{-1} – with a record-breaking maximal instantaneous luminosity of $7.67 \cdot 10^{-33}$ $s^{-1}cm^{-2}$.

The very high instantaneous luminosity which was reached in 2011 and 2012 resulted in a huge amount of data to be analysed by the experiments, but it also has a big disadvantage: a large amount of pileup. The distribution of the mean number of interactions as observed by the CMS experiment is shown in Figure 2.3. This high number of pileup interactions introduces a lot of additional particles and corresponding energy deposits in an experiment, which makes it much harder to correctly reconstruct and identify the interesting particles produced in these collisions.

2.1.2 The experiments located at the LHC

Around each of the four interaction regions large experiments were constructed to perform a detailed analysis of the particle collisions delivered by the LHC. The two largest experiments are the ATLAS [26] and CMS [27, 76] general-purpose experiments. These are both designed to study a wide variety of physics phenomena. This includes high-precision measurements of Standard Model quantities, the search for and measurement of the BEH boson, and the search for all kinds of signals from physics beyond the Standard Model. To

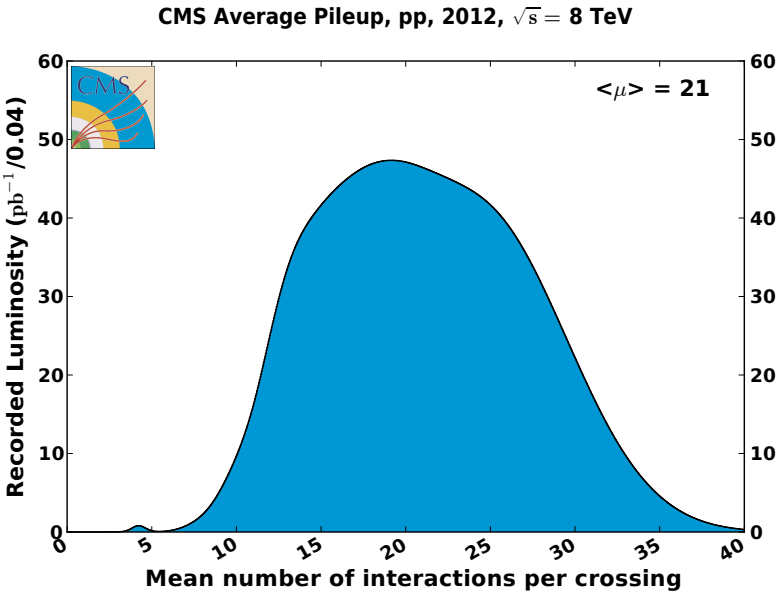

Figure 2.3: The mean number of proton-proton interactions in the 2012 data recorded by the CMS detector.

achieve this both experiments use complementary detection techniques. More information about the CMS detector can be found in Section 2.2.

Apart from these two large experiments there are also two somewhat smaller particle detectors. Firstly there is the ALICE experiment [77], which is designed to study the heavy-ion collisions produced by the LHC. Secondly there is the LHCb experiment [78], which is optimised to deliver very precise measurements in the field of *b*-quark physics.

In addition to these four main experiments, there are also two small experiments: the LHCf [79] experiment and the TOTEM experiment [80]. Both experiments are optimised to detect particles emitted in the very forward region during proton-proton collisions.

2.2 The Compact Muon Solenoid

As one of the two general-purpose detectors, the CMS detector is designed to perform a large variety of measurements. To successfully accomplish this goal, it needs to be able to precisely measure all particles emerging from the interaction point where the proton-proton collisions occur. The detector has several layers built around the collision point. Its geometry consists out of a central *barrel* part which is closed at its two ends by an *endcap* part, as shown in Figure 2.4.

The inner layer of the CMS detector is a silicon-based tracking detector, as explained in Section 2.2.1, which reconstructs the trajectories of all charged particles traversing the detector. The tracking detector is surrounded by the electromagnetic and hadronic calorimeters, which measure the energy of most particles produced during collisions. More information about the calorimeter system is given in Section 2.2.2. The barrel part of the calorimeter system is located within the superconducting magnet, which produces a solenoidal field of 3.8 T inside its volume. The final detector layer is the muon system, as explained in more

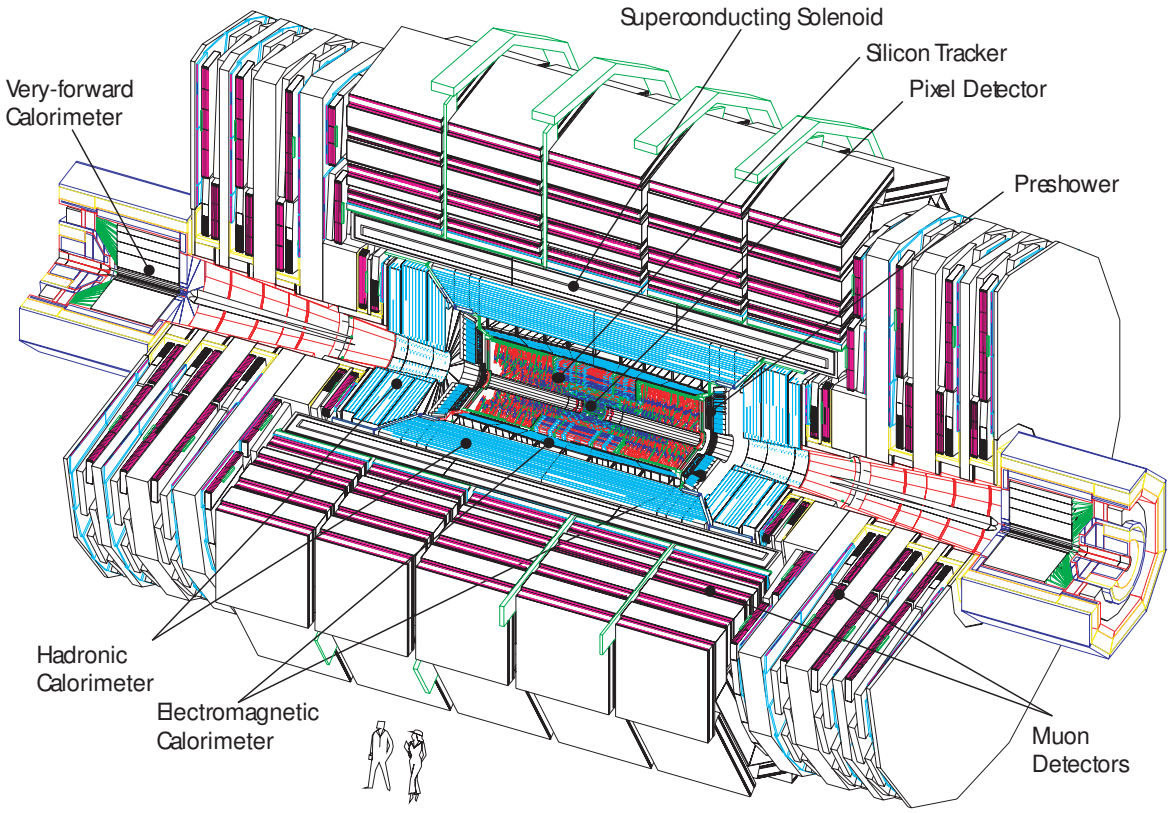

Figure 2.4: Graphical overview of the entire CMS detector.

detail in Section 2.2.3 and measures the energy and momenta of the muons traversing all detector layers.

The overall dimensions of the CMS detector are a length of 21.6 m, a diameter of 14.6 m and a total weight of 12500 tons. The origin of its coordinate system is centred at the nominal collision point, with the y-axis pointing vertically upward, the x-axis pointing inwards to the centre of the LHC, and the z-axis pointing along the direction of the counterclockwise-rotating proton beam. The azimuthal angle ϕ is measured from the x-axis in the $x - y$ plane and the polar angle θ is measured from the z-axis. The pseudorapidity is defined as

$$\eta \equiv -\ln \tan\left(\frac{\theta}{2}\right) \tag{2.2}$$

and has the advantage that differences in η are invariant under Lorentz boosts along the beamline. Another useful quantity is the rapidity, which has numerical values that are close to the pseudorapidity values. It is also denoted with the symbol y and is defined as

$$y \equiv \frac{1}{2} \ln \frac{E + p_z}{E - p_z}. \tag{2.3}$$

At the nominal bunch crossing rate of 40 MHz the CMS experiment will produce 40 TB of data per second, since each collision contains on average 1 MB of data. Because the current state-of-the-art computer systems are not able to handle such a large amount of

data, the CMS detector is equipped with an online event selection system as explained in Section 2.2.4. The computing system designed to handle the vast amount of data produced by CMS is explained in Section 2.2.5. The performance of the CMS detector during the 2010-2012 run of the LHC is discussed in Section 2.2.6.

2.2.1 The inner tracking system

The central part of the CMS detector is a large tracking detector which is designed to precisely measure the direction and momentum of the charged particles produced during the LHC's collisions. This subdetector is centred around the collision point within a cylindrical volume with a length of 5.8 m and a diameter of 2.5 m. A schematic overview of the geometry of tracking system is given in Figure 2.5.

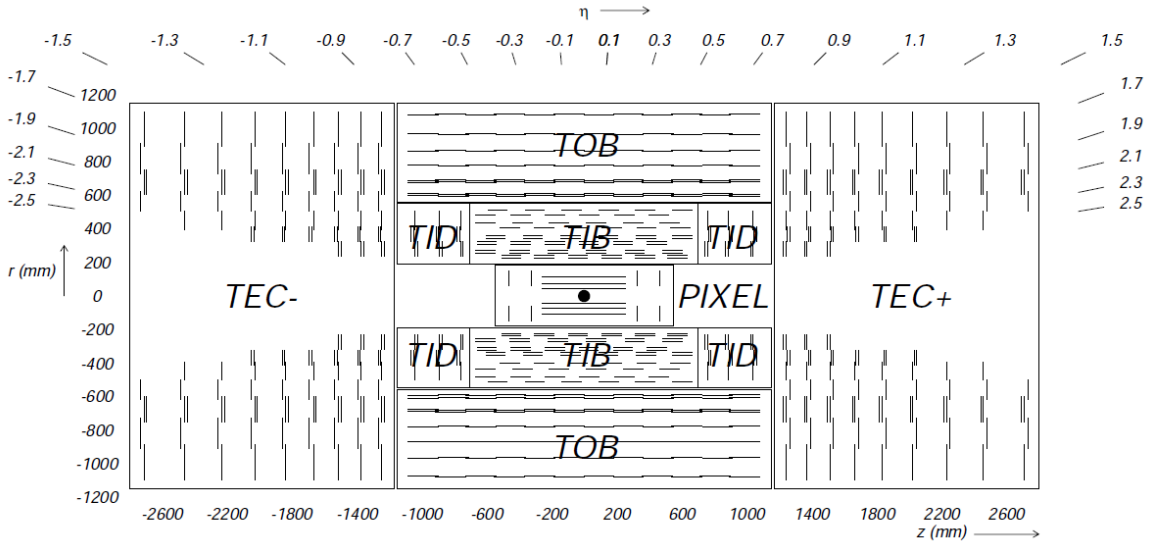

Figure 2.5: Geometry of the CMS tracking detector.

The inner part of the tracking detector is the pixel detector, which consists out of three barrel layers located at radii of 4.4, 7.3 and 10.2 cm and two endcap disks at each side located at $z = \pm 34.5$ and ± 46.5 cm. This arrangement results in at least 3 tracking points over almost the full pseudorapidity range. The entire detector consists out of 66 million pixels which all have a size of 100×150 μm^2. This high granularity is needed for an efficient reconstruction of charged particle tracks in the high pileup environment of the LHC.

In the radial region between 20 cm and 116 cm the silicon strip tracking detector is located, which is composed of four subsystems. The most central part is the Tracker Inner Barrel (TIB) which extends up to a radius r of 55 cm within $|z| < 65$ cm. The TIB subdetector is complemented with 3 endcap disks at each side within the region $|z| <$ 100 cm called the Tracker Inner Disks (TID). The combined TIB/TID system has strip pitches between 80 and 120 μm. It is completely surrounded by the Tracker Outer Barrel (TOB), which has an outer radius of 116 cm, covers the volume $|z| < 118$ cm and has strip pitches between 122 and 183 μm. The final subsystem is the Tracker EndCap (TEC), covering the volume within 124 cm $< |z| <$ 282 cm and 22.5 cm $< r <$ 113.5 cm, and

having pitches between 97 and 184 μm. This layout ensures at least 9 hits in the silicon strip tracker within the range $|\eta| < 2.4$.

The main objective of the tracking detector is the precise reconstruction of the trajectories of charged particles crossing the detector. These can then be used to identify and reconstruct primary vertices originating from proton-proton collisions or other vertices originating from the decay of long-lived particles.

Reconstruction of charged particle tracks

The starting point of the CMS track reconstruction algorithm is the collection of tracker hits made by charged particles traversing the tracker modules. The reconstruction of these hits into tracks is done by the iterative-tracking algorithm [76, 81]. Within each iteration the combinatorial track finder is applied, which consists of four logical parts: seed generation, trajectory building, ambiguity resolving, and the final track fit.

Seed generation
: In this first step the tracker hits are used to build seeds. These seeds are initial trajectory candidates to be used as a starting point in the next reconstruction step. They are generated by combining three tracker hits or two tracker hits and a beam constraint which are compatible with a single charged particle traversing the detector. The most precise seeds are obtained using only hits in the pixel detector, but to increase the track reconstruction efficiency the hits in the silicon strip detector are also used.

Trajectory building or pattern recognition
: The trajectories of charged particles are built using a combinatorial Kalman filter method. This filter starts from a seed and extrapolates the track to the next layer of the tracking detector, where a compatible tracker hit is looked for. Finally the track parameters are updated taking this possible additional hit into account and this process is repeated until the outermost layer of the tracker is reached.

Ambiguity resolving
: Ambiguities can occur for example when multiple tracks are built starting from the same seed. They are resolved based on the fraction of hits shared between two tracks and based on the number of hits and the χ^2 of the fit of these tracks.

Final track fit
: For their final estimates the track parameters are refitted with the same Kalman filter method, this time taking into account all the hits attributed to the track during the pattern recognition step. This is performed two times: once starting from the beamline and running inside-out towards the calorimeters and once starting from the most outward hit and running outside-in towards the beamline. This procedure yields optimal estimates of the track parameters.

During the first iteration tracks are reconstructed with very tight criteria, resulting in a moderate efficiency and a negligible fake rate. During the next iterations tracker hits unambiguously assigned to the tracks found in the previous iteration are removed and the track reconstruction is repeated using the remaining hits, with the track reconstruction

criteria being loosened progressively in each iteration. These looser criteria increase the tracking efficiency, while the fake rate is kept low due to the hit removal.

For muons within the region $|\eta| < 2.4$ and with $p_T = 100$ GeV this algorithm has a reconstruction efficiency larger than 99%. The transverse momentum resolution is of the order of $1 - 2\%$ for 100 GeV muons within $|\eta| < 1.6$ and rises up to 7% at $|\eta| = 2.4$ [81].

Primary vertex reconstruction

The identification of primary vertices [76] uses the reconstructed tracks which are originating from the beamline. These are combined into clusters of tracks based on the z-coordinate of their point of closest approach with respect to the beamline. These clusters are then promoted to primary vertices. The main primary vertex is the one with highest $\sum p_T^2$ of its associated tracks.

To improve the position resolution each cluster of tracks is fitted separately to determine the coordinates of the primary vertex. In these fits tracks are downweighted according to their distance from the common vertex to reduce the effect of long-lived particles decaying into charged particles. After this fit a position resolution of ~ 20 μm is obtained in the x and y directions and ~ 30 μm in the z direction [81].

2.2.2 The calorimeters

The energy of most particles produced during the LHC's collisions is measured in the calorimeter system of CMS which is built around the tracker. This consists of two main parts: an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL). Each of these calorimeters is optimised for a precise measurement of different kinds of particles.

The electromagnetic calorimeter

The inner part of the calorimeter system is the ECAL detector which is designed to precisely measure the energy of electrons and photons. It is a hermetic homogeneous calorimeter built out of scintillating lead tungstate crystals ($PbWO_4$). A geometrical overview of the ECAL is shown in Figure 2.6.

The barrel part of the ECAL (EB) is made of 61200 crystals and covers the range $|\eta| < 1.479$. At their front face these crystals have a size of 22×22 mm^2, which corresponds approximately to 0.0174×0.0174 in $\eta - \phi$ space, and they each have a total length of 230 mm. Each of the two ECAL endcaps consists of 7324 crystals, which cover the range $1.479 < |\eta| < 3.0$, have a front face size of 28.62×28.62 mm^2 and have a length of 220 mm.

In front of the ECAL endcaps (EE) covering the range $1.653 < |\eta| < 2.6$ a preshower detector is located. This is a tracking device where two layers of lead are interleaved with silicon strip sensors to measure both the position and deposited energy of particles traversing the detector. Its main goal is to discriminate photons from π^0 particles.

The electromagnetic calorimeter allows a very precise measurement of the energy of electrons and photons. Their energy resolution has been measured with electrons originating from $Z \to e^+e^-$ decay, which results in resolutions of $1 - 3\%$ in the barrel and $2 - 5\%$ in the endcaps [82].

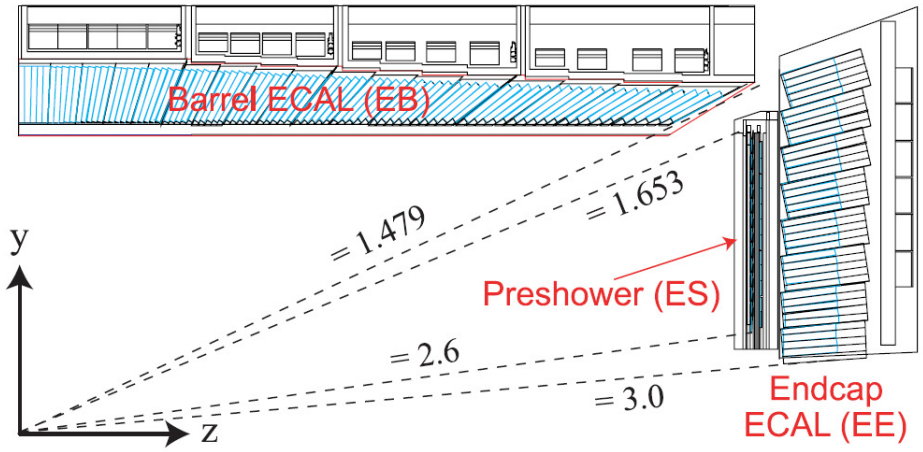

Figure 2.6: Overview of one quarter of the ECAL subdetector.

The hadronic calorimeter

The electromagnetic calorimeter of the CMS detector is surrounded by the hadron calorimeter (HCAL), which is designed to measure the energy of hadrons produced during the LHC's collisions. This is necessary for a precise reconstruction of jets and missing transverse energy. The main part of the HCAL is a sampling calorimeter where brass absorber plates are interleaved with plastic scintillators to measure the deposited energy. A schematic overview of the HCAL subdetectors is given in Figure 2.7.

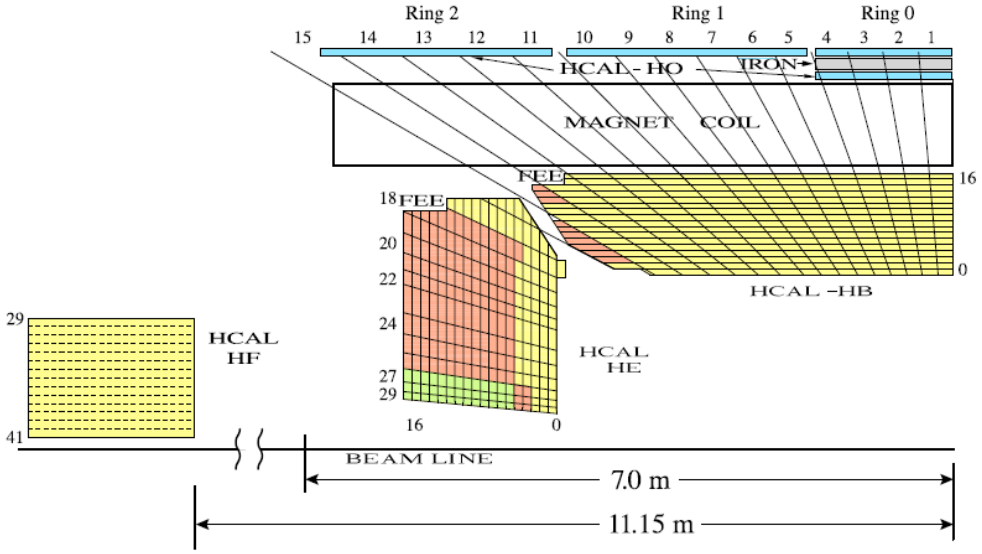

Figure 2.7: Layout of the different HCAL subdetectors.

The central part of the HCAL calorimeter, extending out to $|\eta| = 1.4$, is the HCAL barrel subdetector (HB). Since the HB subdetector is located within the limited space inside the magnet coil, hadronic shower leakage can sometimes happen. Therefore the HB subdetector is complemented with additional scintillator tiles located outside the magnet

called the hadronic outer subdetector (HO). The combined geometry of the HB and HO subdetectors consists of projective towers with a size of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$.

The HCAL endcap subdetector (HE) overlaps slightly with the HB subdetector and covers the pseudorapidity range $1.3 < |\eta| < 3.0$. The $\Delta\eta \times \Delta\phi$ tower size matches that of the barrel for $1.3 < |\eta| < 1.74$, while for larger $|\eta|$ it increases stepwise to a maximal size of 0.350×0.174. In the very forward region ($2.9 < |\eta| < 5.2$) the HE subdetector is complemented by the HCAL forward subdetector (HF). For these calorimeters steel absorbers with embedded quartz fibres are used to measure the energy of incoming particles. The $\Delta\eta \times \Delta\phi$ tower size of the HF subdetector ranges between 0.111×0.174 and 0.302×0.348.

When the ECAL and HCAL calorimeters are combined an energy resolution of 24% is obtained for 20 GeV pions [76]. For 100 GeV pions the energy resolution is further improved to 13%.

2.2.3 The muon system

For an efficient detection and a precise reconstruction of muons the CMS detector is surrounded by an extensive muon system. It covers the region $|\eta| < 2.4$ and its geometry is shown in Figure 2.8. The different muon chambers are hosted in the iron magnet return

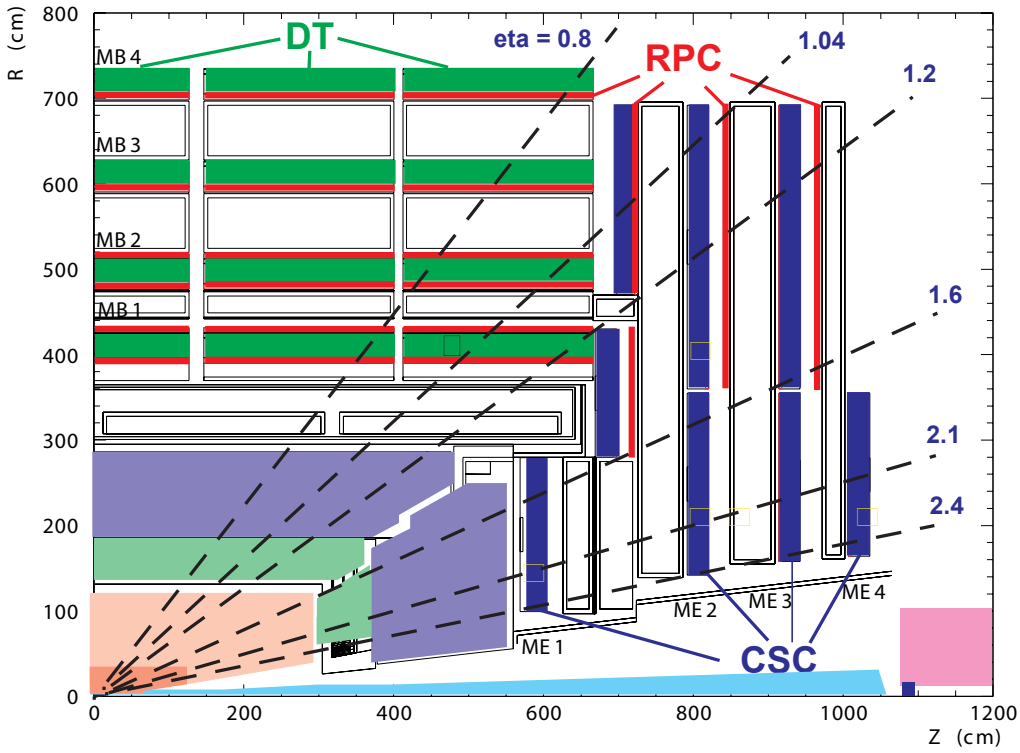

Figure 2.8: Longitudinal view of the muon system geometry.

yoke to benefit from the strong magnetic field and of its hadron absorber capabilities.

In the barrel part of the CMS detector ($|\eta| < 1.2$) 4 layers of drift tubes (DT) are present, while in the endcaps ($|\eta|$ values between 0.9 and 2.4) up to 4 layers of cathode

strip chambers (CSC) are used. These two systems provide a precise measurement of the muon transverse momenta and trajectories. To improve the time resolution of the muon system the region $|\eta| < 1.6$ is also equipped with resistive plate chambers (RPC).

2.2.4 The online event selection

When operating at nominal conditions the LHC will produce bunch crossings at a rate of 40 MHz. Because of storage and computing limitations events can only be recorded at a maximal rate of 400 Hz, which means that an online event selection needs to be performed to reduce the number of events by a factor of 10^5. The online event selection or trigger system of the CMS experiment consists of two levels: the Level-1 (L1) trigger and the High Level Trigger (HLT).

The Level-1 trigger is made of custom-designed, largely programmable electronics. It uses coarsely segmented data from the calorimeters and the muon system to decide to keep or reject an event within 3.2 μs. The maximal output rate of the L1 trigger is 100 kHz, which is a reduction of at least a factor of 400 with respect to the bunch crossing rate of 40 MHz.

Events which are selected by the L1 trigger are processed by the HLT, which is a software system implemented on a large farm of commercial processors. The HLT system has access to the complete reconstructed event data and performs complex algorithms on these events, which are very similar to those used in the offline software.

2.2.5 The computing aspect of CMS

After being selected by the online trigger system the collision events are passed on to the CMS computing system. This is designed to support the storage, transfer and manipulation of the recorded data and the necessary simulated collision events. It was designed in a combined way for all the LHC experiments so it is often referred to as the Worldwide LHC Computing Grid (WLCG) [83].

Because of the very large scale of this project a design consisting of multiple levels (or Tiers) of computing centres was chosen:

Tier-0 (T0)
: A single Tier-0 computing centre is hosted at CERN. It is the entry point of the data recorded by the CMS detector, where this data is copied to permanent mass storage and a first reconstruction is performed. The T0 computing centre is also responsible for the transfer of all the data to at least two independent Tier-1 computing centres.

Tier-1 (T1)
: A total of 7 Tier-1 centres are located around the world. They provide large batch CPU facilities to carry out data reprocessing, reliable mass storage systems to ensure long-term safe storage of the data from CMS (including simulated data), and very high speed international network links to allow a fast and efficient distribution of these data to the different Tier-2 centres for analysis.

Tier-2 (T2)
: Multiple ($\sim$ 55) Tier-2 centres are also distributed around the world. These are

smaller computing centres of varying sizes with two main tasks. Firstly they support analysis activities via a CPU farm and local storage of the necessary data. Secondly they support the production of simulated data and their transfer to the T1 centres.

2.2.6 The CMS detector during the 2010-2012 run of the LHC

The CMS detector was operated very efficiently during the 2010-2012 run of the LHC. An overview of the integrated luminosity delivered by the LHC, recorded by the CMS detector, and validated for analysis can be seen in Figure 2.9.

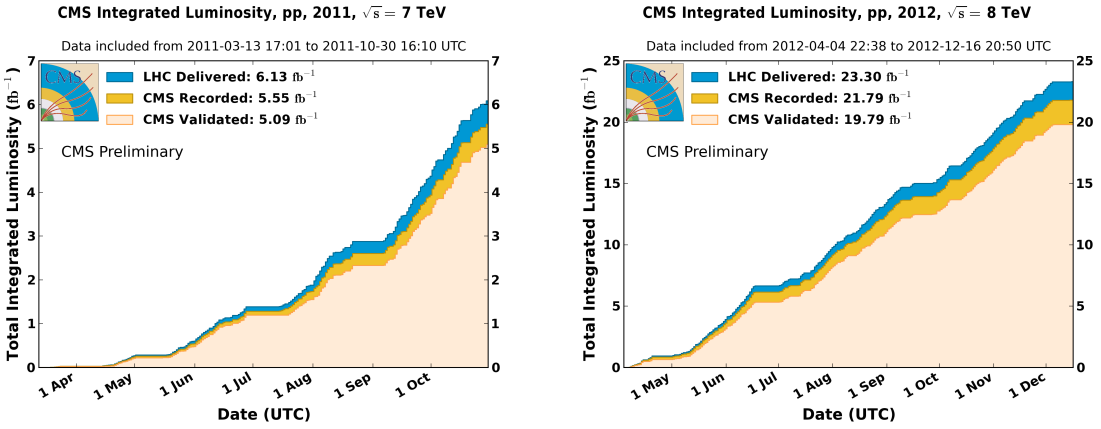

Figure 2.9: The integrated luminosity delivered to, recorded by and certified by the CMS experiment in 2011 (left) and 2012 (right).

From the numbers in these figures it can clearly be seen that the CMS detector recorded the proton-proton collisions delivered by the LHC with a high efficiency. The data recording efficiency was 90.5% in 2011 and 93.5% in 2012. There are several possible reasons for data loss, like for example issues with the online trigger system which decides whether events should be recorded or not, or problems with the power supplies of the CMS detector.

Since not all data recorded by the CMS detector is of the necessary high quality, a separate offline validation of this data is performed. During this validation it is checked that all the subcomponents of the CMS detector (the magnet, the different subdetectors, ...) are working as expected and that the reconstructed objects like electrons, muons and jets are also behaving according to the expectations. By comparing the numbers of the validated and recorded integrated luminosity a validation efficiency of 91.7% and 90.8% was observed in 2011 and 2012, respectively.

Chapter 3

Simulation and reconstruction of proton-proton collisions

To allow a detailed comparison between the real data taken by the CMS experiment and the theoretical expectations from the standard model two additional aspects need to be considered. The first aspect is a detailed simulation of the proton-proton collisions and the detector response to the particles produced during these collisions. The generation of proton-proton collisions is explained in Section 3.1, while more information about the simulation of the detector response can be found in Section 3.2. The second aspect is the reconstruction of physical objects like charged particles and energy deposits out of the electronic signals accumulated by the CMS detector. The same algorithms, which are discussed in Section 3.3, are applied to real data and to simulated events.

3.1 Generating proton-proton collision events

The complex process of generating individual $pp \to X$ collision events can be factorised into different steps [84–86]. These start from the incoming protons and continue until the hadrons that form the experimentally observed final state. These steps are illustrated in Figure 3.1 and are summarised below.

Incoming protons
: Each of the two incoming protons can be considered as a group of partons – quarks and gluons. The momenta of these partons inside the proton are distributed according to the so-called parton distribution functions, which are discussed in Section 3.1.1.

Hard scattering
: Two partons, one parton from each incoming proton, collide with each other during the hard scattering. When very short-lived particles like top quarks or heavy gauge bosons are produced their decay is also included in this process. This part of the event generation is also discussed in Section 3.1.1.

Parton shower
: The incoming and outgoing partons can emit additional gluons and quarks. Radiation produced by the incoming partons is called Initial State Radiation (ISR), while

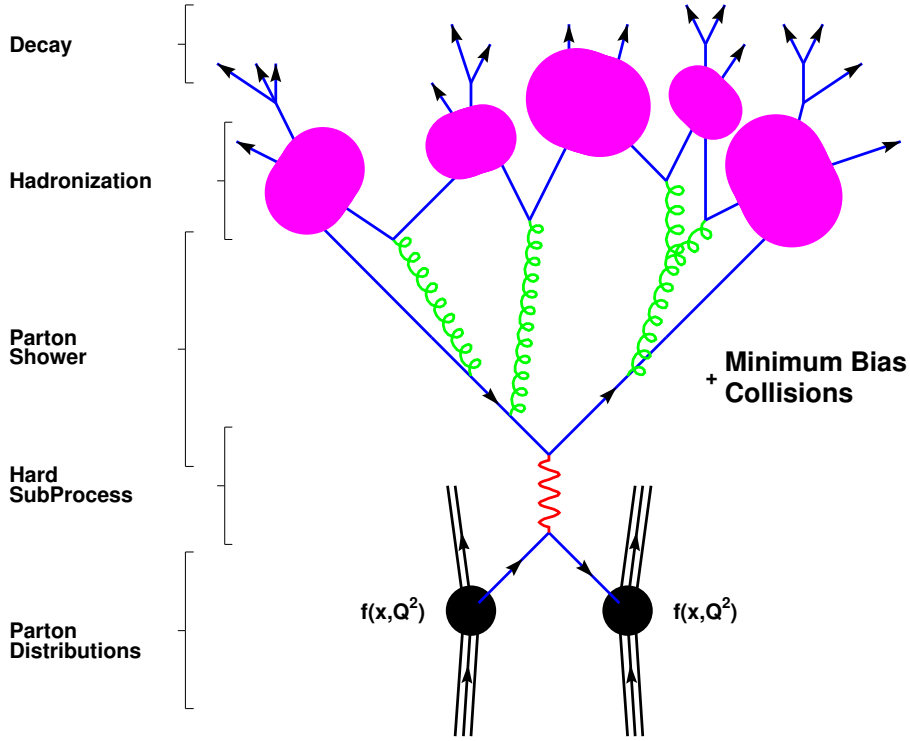

Figure 3.1: Graphical overview of the different event generation steps.

radiation emitted by the outgoing partons is called Final State Radiation (FSR). This is modelled using parton showering and discussed in Section 3.1.2.

Hadronisation

When the partons produced in the previous steps recede from each other their evolution can not be described anymore with perturbative QCD. Phenomenological hadronisation models are then used to describe how the partons give rise to colour neutral hadrons. This is described in Section 3.1.3.

Underlying event

Since the colliding partons take only a fraction of the incoming proton energy, much of the initial energy remains in the beam remnants. These remnants carry also colour charge so via radiation and hadronisation they will produce the so-called underlying event, as explained in Section 3.1.4.

3.1.1 The hard scattering

The calculation of the production cross section of a final state X as a function of an observable $\mathcal{O}$ can also be factorised into different parts. This differential cross section for $pp \to X$ production can be calculated as

$$\frac{\mathrm{d}\sigma_{pp\to X}}{\mathrm{d}\mathcal{O}} = \sum_{i,j=q,\bar{q},g} \iint \mathrm{d}x_i \mathrm{d}x_j \; f_i(x_i, Q^2) \; f_j(x_j, Q^2) \; \frac{\mathrm{d}\widehat{\sigma}_{ij\to X}}{\mathrm{d}\mathcal{O}}. \tag{3.1}$$

In this equation the sum runs over all partonic constituents i and j of the two incoming protons, carrying momentum fractions x_i and x_j. The partonic differential cross section $\frac{\mathrm{d}\hat{\sigma}_{ij\to X}}{\mathrm{d}\mathcal{O}}$ can be calculated within QCD as an expansion in the strong coupling constant α_s, where the inclusion of higher orders improves the theoretical precision of these calculations. Calculations where only the term in the lowest α_s order is included are called leading order (LO) calculations. These LO calculations are limited in precision and have some important shortcomings. For example, they do not include the possibility of additional final-state partons that can be produced during the hard scattering. Therefore the inclusion of higher order terms is necessary. A significantly better precision is achieved via next-to-leading order (NLO) or next-to-next-to-leading order (NNLO) calculations, which are becoming available for an increasing amount of processes.

The parton distribution functions (PDF) $f_i(x_i, Q^2)$ and $f_j(x_j, Q^2)$ in Equation 3.1 represent the probabilities of finding partons i and j in the incoming protons carrying momentum fractions x_i and x_j at a scale Q^2. These functions can not be calculated from first principles so they are determined via global fits to experimental data. The PDFs used here are the ones obtained by the CTEQ collaboration [87] and are plotted in Figure 3.2 for $Q^2 = (350\ \mathrm{GeV})^2$. These are described by 22 independent parameters and their corresponding uncertainties.

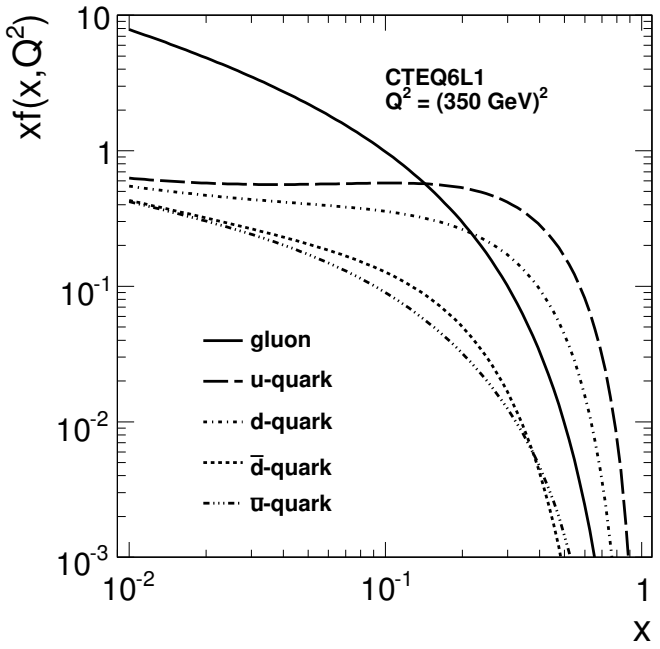

Figure 3.2: CTEQ PDFs for different partons at a scale $Q^2 = (350\ \mathrm{GeV})^2$.

For the production of a pair of top quarks a minimal partonic centre-of-mass energy $\sqrt{\hat{s}}$ given by $\hat{s} \approx (2m_t)^2 \approx (350\ \mathrm{GeV})^2$ is needed, assuming a top-quark mass of 175 GeV. The partonic centre-of-mass energy for the collision of two partons i and j is given by $\hat{s} = x_i x_j s$, so when setting $x_i \approx x_j \approx x$ this results in momentum fractions $x \approx 2m_t/\sqrt{s}$. This gives values of x equal to 0.050 and 0.044 for 7 and 8 TeV centre-of-mass energies, respectively. At these low x values the gluon PDF is completely dominant as shown in the Figure 3.2, which means that $t\bar{t}$ production at the LHC happens mainly via the gluon-gluon fusion process $gg \to t\bar{t}$.

To generate an event sample which can be compared with real collision events Equation 3.1 can be used. In this case the energy and momenta of the final state particles can be used as observables $\mathcal{O}$ and events are generated according to the theoretical $\frac{d\sigma_{pp\to X}}{d\mathcal{O}}$ distribution. Several LO event generators are available, for example PYTHIA [88], but due to their more precise description of the data other event generators are used more frequently, like the multi-leg LO MADGRAPH [89] generator, and the NLO generators POWHEG [90] and MC@NLO [91].

MadGraph

MADGRAPH is a multi-purpose matrix element generator which automatically generates tree-level matrix elements for decays and $2 \to n$ scatterings. This allows the inclusion of real higher-order corrections (additional final-state partons), but no virtual higher-order corrections are considered. In this way the matrix elements for $t\bar{t}$ production with up to 3 additional partons and W-boson production with up to 5 additional partons can be calculated. With these matrix elements events can be generated with the MADEVENT event generator.

POWHEG and MC@NLO

These two generators are able to generate individual events for a limited number of processes with NLO accuracy, like for example processes involving one or more top quarks or one or more gauge bosons. They are based on theoretical matrix-element calculations which have been calculated with NLO precision. The difference between POWHEG and MC@NLO lies in the way these theoretical NLO calculations are matched to the parton showers, which will be explained in Section 3.1.2.

3.1.2 Parton showering

The description of the additional quarks and gluons emitted by the incoming and outgoing partons of the hard interaction is described by the parton showering formalism. Here the probability $\mathcal{P}_{a\to bc}$ of a parton a at a scale Q^2 to split into two partons b and c is given by the DGLAP (Dokshitzer-Gribov-Lipatov-Altarelli-Parisi) equations [92–94]

$$d\mathcal{P}_{a\to bc} = \frac{\alpha_s}{2\pi}\frac{dQ^2}{Q^2}P_{a\to bc}(z)dz, \tag{3.2}$$

where z defines the energy sharing between the two final-state partons b and c, defined as $E_b = zE_a$ and $E_c = (1-z)E_a$. The different splitting functions $P_{a\to bc}$ are given by

$$P_{q\to qg}(z) = \frac{4}{3}\frac{1+z^2}{1-z}, \tag{3.3}$$

$$P_{g\to gg}(z) = 3\frac{(1-z(1-z))^2}{z(1-z)}, \tag{3.4}$$

$$P_{g\to q\bar{q}}(z) = \frac{n_f}{2}(z^2+(1-z)^2), \tag{3.5}$$

where n_f if the number of quark flavours.

These equations can then be applied to every final-state parton to describe their shower. A parton a is evolved downwards from some initial scale Q^2_{max} until it branches. Here the

mother parton a is replaced by two daughter partons b and c, which in their turn are evolved downwards in Q^2 and may branch. This process gives rise to a cascade of partons and continues until some lower Q^2 cutoff scale around $(1\ \text{GeV})^2$ is reached to avoid the $Q^2 \to 0$ divergence. Below this scale hadronisation effects take over, which are modelled in a different way as explained in Section 3.1.3.

The $\mathcal{P}_{a\to bc}$ probabilities obtained from these equations are clearly larger than one in the soft limit ($z \to 1$). This can be solved by the introduction of a Sudakov form factor, which changes the DGLAP equations into

$$\mathrm{d}\mathcal{P}_{a\to bc} = \frac{\alpha_s}{2\pi}\frac{\mathrm{d}Q^2}{Q^2}P_{a\to bc}(z)\mathrm{d}z\ \exp\left(-\sum_{b,c}\int_{Q^2}^{Q^2_{max}}\frac{\mathrm{d}Q'^2}{Q'^2}\int\frac{\alpha_s}{2\pi}P_{a\to bc}(z')\mathrm{d}z'\right). \quad (3.6)$$

This Sudakov factor can be interpreted as the probability that no branching of parton a occurred when it evolved from a scale Q^2_{max} to a scale Q^2.

The parton shower approach is also applied to generate ISR showers. In this case the starting point are the partons entering the hard interaction and they are also evolved to lower and lower Q^2, which means that the evolution goes backwards in time. Both the ISR and FSR parton showers are implemented into different programs, like PYTHIA [88] and HERWIG [95]. Most of the simulated data samples used in this thesis are showered with the transverse momentum $p^2_\perp$ ordered shower approach of PYTHIA, while some samples are produced with the $E^2\theta^2$ ordered showers from HERWIG to estimate the effect of different shower approaches.

Matching hard-scattering matrix elements with parton showers

When a parton shower formalism is used in combination with a hard-scattering matrix-element calculation to generate events, an unambiguous separation between the hard scattering and the parton showering is necessary [96]. Its primary goal is to avoid the double-counting of events, since a $(n+1)$-jet event can be obtained in two ways: from the showering of a $(n+1)$-parton matrix-element final state, or from an n-parton matrix-element final state with a hard emission during the parton shower leading to an extra jet. A proper matching scheme between the matrix element and the parton shower will also avoid dead regions by ensuring that each configuration is generated by at least one of the allowed paths.

Most simulated event samples used in this thesis are generated with MADGRAPH matched to PYTHIA for the parton showering. This matching is performed with the $k_\perp$-based variant of the MLM matching scheme [97], which uses the same definition to measure distances as the $k_\perp$ jet algorithm as explained in Section 3.3.4. The minimal distance d_{min} among the distances d_{ij} between any two partons i and j, and the distance d_{iB} between a parton i and the beam is used. Both distances are defined as:

$$d_{ij} = \min\left(k^2_{\perp i}, k^2_{\perp j}\right)\frac{\Delta R^2_{ij}}{R^2} \quad \text{and} \quad d_{iB} = k^2_{\perp i}, \quad (3.7)$$

where $k_{\perp i}$ is the transverse momentum of parton i, R is a radius parameter, and $\Delta R_{ij} = \sqrt{\Delta y^2_{ij} + \Delta\phi^2_{ij}}$ is the separation between two partons i and j in the (y, ϕ)-plane where y is the rapidity as defined in Equation 2.3.

Initially a set of partonic events with n partons from the matrix element is generated, where the minimal distance d_{min} needs to exceed a certain threshold d_{cut}. Afterwards these events are showered with PYTHIA and the resulting final-state partons are clustered into jets with the $k_\perp$ jet algorithm with a cutoff $d_{cut}^{jets} > d_{cut}$. The resulting jets are then matched to the n original matrix-element partons by requiring the distance d_{ij} between a matrix-element parton i and a jet j to be smaller than d_{cut}^{jets}. An event is then rejected unless each jet is matched to a matrix-element parton, except for the highest multiplicity sample. Here extra jets are allowed below the d_{min} value of the softest matrix-element parton in the event, since these additional jets can only be generated via the parton showering.

Events generated with POWHEG and showered with PYTHIA are also used in this thesis. In this case events are generated using NLO calculations, so including events where an additional parton is emitted in the hard process. Afterwards the parton showering is performed where the subsequent softer radiation at lower Q^2 compared to this additional hard parton is generated. This results in events with up to one additional emission with full NLO accuracy, while any additional radiation is generated by the parton shower.

A sample of $t\bar{t}$ events generated with MC@NLO and showered with HERWIG is also used. Here NLO corrections to an n-parton final state are calculated, including $(n+1)$-parton real corrections and n-parton virtual corrections. It is also calculated how a first branching in a shower starting from an n-parton topology would populate the $(n+1)$-parton final state. This is then subtracted from the $(n+1)$ matrix element calculation to define the true $(n+1)$ events, while all other events are then belonging to the n-parton final state. Finally both types of events are showered with HERWIG.

3.1.3 Hadronisation

The next step after completing the showering of a collision event is the formation of colour-neutral hadrons out of all the coloured partons produced by the showering algorithm. This process is called hadronisation and happens at small scales where the perturbative approach of QCD is not valid anymore, meaning that this can not be calculated from first principles and phenomenological models have to be used.

The model implemented within PYTHIA is the Lund string model [86, 98], which uses the fundamental property that the potential V of a QCD colour field or colour string between a quark q and an antiquark $\bar{q}$ grows linearly with the separation r between these quarks:

$$V(r) = \kappa r, \quad \text{with} \quad \kappa \sim 1 \ \frac{\text{GeV}}{\text{fm}}. \tag{3.8}$$

So when q and $\bar{q}$ move apart, their kinetic energy is gradually converted into potential energy stored in the string spanned between them. This continues until a quark-antiquark fluctuation in the string produces two real particles q' and $\bar{q}'$ by absorbing part of the energy of the string: $(q\bar{q}) \rightarrow (q\bar{q}') + (q'\bar{q})$. The production of such a new $q\bar{q}$ pair with quark mass m_q and quark transverse momentum $p_{\perp q}$ is based on quantum tunnelling and proportional to

$$\exp\left(\frac{-\pi m_q^2}{\kappa}\right) \ \exp\left(\frac{-\pi p_{\perp q}^2}{\kappa}\right). \tag{3.9}$$

The presence of the quark mass m_q in this equation means that only u, d and s quarks will be produced, since other quarks have a much larger mass.

One of these newly created quark pairs will form a meson, while subsequent splittings of the colour string between the other quark pair can happen. Baryons on the other hand are created within this model by the production of an diquark-antidiquark pair instead of a quark-antiquark pair during string breaking. The probability to obtain a resulting hadron h with mass m_h, transverse momentum $p_{\perp h}$ and longitudinal momentum fraction z is determined from the Lund symmetric fragmentation function

$$f(z) \propto \frac{1}{z}(1-z)^a \exp\left(\frac{-b\ (m_h^2 + p_{\perp h}^2)}{z}\right), \tag{3.10}$$

where a and b are tunable parameters. For the significantly heavier c and b quarks an additional factor $1/z^{b\ m_q^2}$ is included in the fragmentation function, as suggested by experimental data.

An alternative hadronisation algorithm is the cluster fragmentation model [99] which is implemented in HERWIG. Here all gluons produced during the showering are first split into quark-antiquark pairs. Afterwards all colour connected quarks are grouped into clusters from which hadrons are produced.

The final collection of particles produced by event generators contains a significant fraction of unstable particles. Their decay takes place after the hadronisation is finished. For the decay of some particles external packages can be used which give a more precise description of these decays, like for example TAUOLA [100] which is used for the decay of τ leptons and takes the spin information properly into account.

3.1.4 Underlying event

The two partons undergoing the hard interaction carry a colour charge, which means that the remainder of the incoming protons – the beam remnant – is also colour charged and will hadronise as well. Apart from the hard interaction the other partons of the two incoming protons can also interact, producing multiple parton interactions (MPI). Both the hadronisation of the beam remnant and the multiple parton interactions produce additional soft hadrons resulting in the so-called underlying event [84, 86].

For the modelling of multiple parton interactions within PYTHIA the incoming protons are viewed as a collection of incoming partons which are transversely distributed according to a double Gaussian density

$$\rho(r) = N_1 \exp\left(\frac{-r^2}{r_1^2}\right) + N_2 \exp\left(\frac{-r^2}{r_2^2}\right), \tag{3.11}$$

where N_2/N_1 and r_2/r_1 are tunable parameters. This causes the average number of partonic interactions to depend on the impact parameter between the two colliding protons, so on average central collisions will have a higher number of partonic interactions thus they are more likely to contain at least one hard scattering. These central collisions produce much more activity than peripheral collisions, which also means that hard scattering events like the $t\bar{t}$ events considered in this thesis are sitting on top of a higher MPI background than events without a hard scattering.

The colour structure inside each of the partonic interactions is described by the matrix elements and parton shower as explained before, but since each of these systems carries

away colour charge of the colliding protons some colour connection between these systems, including the proton remnants, must exist. These colour strings can even connect two partonic systems going into opposite directions and will give rise to a significant number of soft particles during hadronisation.

Another important aspect of the underlying event is colour reconnection. This is the interaction between different colour strings that can occur during hadronisation, which will modify the colour structure of the event and affect the hadronisation. Within PYTHIA the probability for a colour string to preserve its original colour connections depends on the total activity in the event, since this probability is lower in very busy events with lots of overlapping colour strings. The modelling of colour reconnection is especially relevant within measurements of the top-quark mass since reconnections can occur between the colour strings belonging to the hadronically decaying W boson and the colour strings between the b quarks and the remainder of the event. This can alter the measured top-quark mass via the measured momenta of the b quark and the W boson [101].

To obtain a good description of the hadronisation and underlying event in real collision events the free parameters of these phenomenological models need to be tuned to data. Within this thesis different tunes are used. Most simulated event samples at a centre-of-mass energy of 7 TeV are produced with the Z2 tune, which is identical to the Z1 tune [102] based on tunes to collision data from previous experiments and taking some new LHC results into account, but using a different set of parton density functions. The event samples at a centre-of-mass energy of 8 TeV are produced with the Z2* tune, which is a slightly updated version of the Z2 tune. Additional event samples with different tunes are also produced to allow a comparison between different tunes [103]. For this the P11 tune is used as a central reference, while the P11mpiHi and P11TeV tunes are, respectively, variations with more and less underlying event activity. An additional variant of the P11 tune where colour reconnections are disabled is also used: the P11noCR tune.

Some properties of the different tunes used for the 8 TeV analysis are shown in Figure 3.3. Here the number of charged generator-level particles with transverse momentum

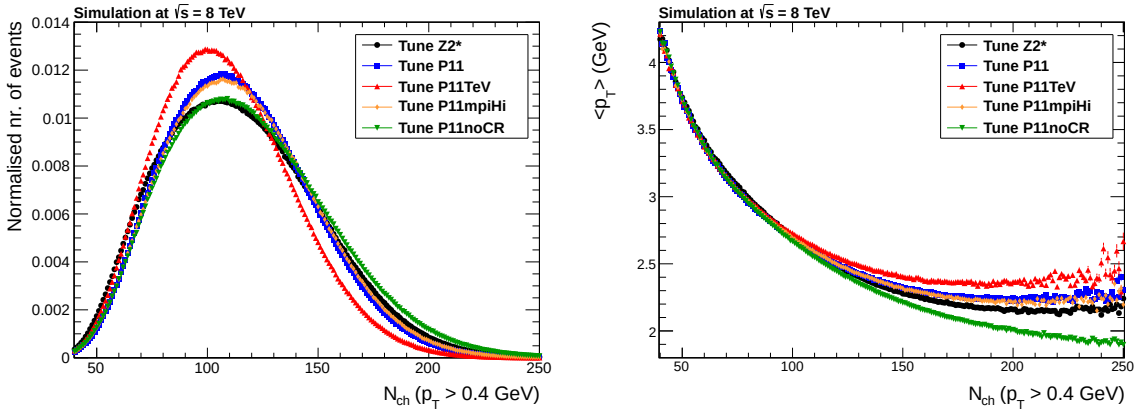

Figure 3.3: Comparison between different PYTHIA tunes of the number of charged particles N_{ch} with $p_T > 0.4$ GeV (left) and of the average charged particle p_T as a function of the number of charged particles (right).

$p_T > 0.4$ GeV is shown, together with the dependence of the average p_T of charged

particles on the number of charged particles. These plots show the decreased and increased event activity of the P11TeV and P11mpiHi tunes with respect to the central P11 tune. The effect of disabling colour reconnections can also be seen by comparing the P11noCR tune with the P11 tune. For events with a high number of charged particles this disabling of colour reconnections results in an increased number of charged particles but with a lower average p_T.

3.2 CMS detector simulation

After being generated the proton-proton collision events are passed on to the CMS detector simulation [76], which is implemented within the CMS software package CMSSW. It is based on the GEANT4 [104] simulation toolkit to provide a description of the interaction between the detector material and the traversing particles. Therefore the geometry of the detector and all its subcomponents is implemented within this software package, together with a precise map of the magnetic field.

All particles emerging from a generated proton-proton collision are propagated to the different layers of the CMS detector and at each layer the interaction between these particles and the detector is simulated. Afterwards the electronic signals produced by the active detector materials and the detector electronics are simulated, resulting in a data stream similar to the output of the real detector. At this point the effect of additional proton-proton interactions (pileup) is also included by adding the detector hits of generated proton-proton interactions on top of the already existing hits resulting from the main interaction. The effects of pileup occurring during the same bunch crossing or those immediately preceding or following the primary crossing are both included. The main disadvantage of this very detailed FULLSIM simulation is that it is quite slow, taking about 3 minutes of CPU time per event. Therefore a more simplified simulation called FASTSIM is also available which is about a factor 100 faster.

Most of the simulated event samples used in this thesis are centrally produced with FULLSIM and used by many analyses within the CMS collaboration. A couple of samples at $\sqrt{s} = 7$ TeV are privately produced using FASTSIM. All the simulated event samples used within this thesis are listed in Table 3.1 for $\sqrt{s} = 7$ TeV, and Tables 3.2 and 3.3 for $\sqrt{s} = 8$ TeV.

3.3 Physics object reconstruction

The reconstruction algorithms, which translate electronic hits in the different subdetectors into physical objects like leptons and jets, are described here. Due to its superior performance the particle-flow event reconstruction is used, which aims at reconstructing and identifying individual particles by combining the information of the different subdetectors. These individual particles can then be used to construct higher-level objects like jets and missing transverse energy.

Since the particle-flow event reconstruction uses the standard muon and electron reconstruction algorithms, a description of these is given first in Sections 3.3.1 and 3.3.2. Afterwards the different parts of the particle-flow algorithm are explained in Section 3.3.3.

Sample	Generator	σ (pb)	# events	$\mathcal{L}$ (fb^{-1})
$t\bar{t}$ + jets	MADGRAPH, tune Z2			
$m_t = 172.5$ GeV		172.0	58 M	337.4
$W \to \ell\nu_\ell$ + jets	MADGRAPH, tune Z2			
W + 4 jets		194.6	13 M	66.3
W + 3 jets		343.0	6 M	19.4
W + 2 jets		1618.1	25 M	15.5
$Z/\gamma^* \to \ell^+\ell^-$ + jets	MADGRAPH, tune Z2	3048	36 M	11.8
single top	POWHEG, tune Z2			
t-channel t		41.92	3.9 M	92.8
t-channel $\bar{t}$		22.65	1.9 M	85.6
tW-channel t		7.87	812 k	103.2
tW-channel $\bar{t}$		7.87	808 k	102.6
$t\bar{t}$ + jets variations	MADGRAPH, tune Z2			
$m_t = 161.5$ GeV		172.0	1.6 M	9.1
$m_t = 163.5$ GeV		172.0	1.6 M	9.3
$m_t = 166.5$ GeV		172.0	1.6 M	9.5
$m_t = 169.5$ GeV		172.0	1.6 M	9.2
$m_t = 175.5$ GeV		172.0	1.5 M	8.8
$m_t = 178.5$ GeV		172.0	1.6 M	9.4
$m_t = 181.5$ GeV		172.0	1.6 M	9.5
$m_t = 184.5$ GeV		172.0	1.6 M	9.5
Q^2 up, less ISR/FSR		172.0	3.2 M	18.9
Q^2 down, more ISR/FSR		172.0	3.6 M	20.8
matching up		172.0	3.9 M	8.8
matching down		172.0	1.5 M	22.9
$t\bar{t}$ + jets tune variations	MADGRAPH			
	tune P11	172.0	8.5 M	49.1
	tune P11noCR	172.0	8.7 M	50.4
$t\bar{t}$ + jets FASTSIM	MADGRAPH			
	tune P11	172.0	8.4 M	48.6
	tune P11mpiHi	172.0	8.4 M	48.6
	tune P11TeV	172.0	8.4 M	49.1

Table 3.1: Overview of the simulated event samples at $\sqrt{s} = 7$ TeV. Showering and hadronisation is performed with PYTHIA for all these samples.

Sample	Generator	σ (pb)	# events	$\mathcal{L}$ (fb^{-1})
$t\bar{t}$ + jets, $m_t = 172.5$ GeV	MADGRAPH, tune Z2*			
ℓ + jets		107.7	25.0 M	231.8
dilepton		25.8	12.0 M	465.5
all-hadronic		112.3	31.2 M	277.4
$W \to \ell\nu_\ell$ + jets	MADGRAPH, tune Z2*			
W + 4 jets		264.0	13.4 M	50.7
W + 3 jets		640.4	15.5 M	24.2
W + 2 jets		2159.2	34.0 M	15.8
W + 1 jet		6662.8	23.1 M	3.5
$Z/\gamma^* \to \ell^+\ell^-$ + jets	MADGRAPH, tune Z2*			
Z/γ^* + 4 jets		27.4	6.2 M	22.8
Z/γ^* + 3 jets		60.7	10.6 M	17.5
Z/γ^* + 2 jets		215.0	2.3 M	10.7
Z/γ^* + 1 jet		666.3	24.0 M	36.0
single top	POWHEG, tune Z2*			
t-channel t		56.4	3.7 M	66.0
t-channel $\bar{t}$		30.7	1.9 M	62.1
tW-channel t		11.1	494 k	44.5
tW-channel $\bar{t}$		11.1	493 k	44.5
$t\bar{t}$ + jets variations	MADGRAPH, tune Z2*			
$m_t = 161.5$ GeV		245.8	5.4 M	21.8
$m_t = 163.5$ GeV		245.8	5.4 M	21.8
$m_t = 166.5$ GeV		245.8	4.4 M	18.1
$m_t = 169.5$ GeV		245.8	5.2 M	21.2
$m_t = 175.5$ GeV		245.8	5.2 M	21.1
$m_t = 178.5$ GeV		245.8	4.7 M	19.2
$m_t = 181.5$ GeV		245.8	5.1 M	20.9
$m_t = 184.5$ GeV		245.8	5.2 M	21.3
Q^2 up, less ISR/FSR		245.8	5.0 M	20.3
Q^2 down, more ISR/FSR		245.8	5.4 M	21.9
matching up		245.8	5.4 M	21.9
matching down		245.8	5.5 M	22.3

Table 3.2: Overview of the simulated event samples at $\sqrt{s} = 8$ TeV, except some samples used for systematic studies. Showering and hadronisation is performed with PYTHIA for all these samples.

Sample	Generator	σ (pb)	# events	$\mathcal{L}$ (fb^{-1})
$t\bar{t}$ + jets, $m_t = 172.5$ GeV	MADGRAPH, tune P11			
ℓ + jets		107.7	12.0 M	111.3
dilepton		25.8	5.8 M	225.4
all-hadronic		112.3	11.7 M	103.7
$t\bar{t}$ + jets, $m_t = 172.5$ GeV	MADGRAPH, tune P11TeV			
ℓ + jets		107.7	7.8 M	72.7
dilepton		25.8	4.0 M	154.3
all-hadronic		112.3	7.9 M	70.7
$t\bar{t}$ + jets, $m_t = 172.5$ GeV	MADGRAPH, tune P11mpiHi			
ℓ + jets		107.7	8.0 M	73.9
dilepton		25.8	4.0 M	154.2
all-hadronic		112.3	8.0 M	70.8
$t\bar{t}$ + jets, $m_t = 172.5$ GeV	MADGRAPH, tune P11noCR			
ℓ + jets		107.7	12.0 M	111.6
dilepton		25.8	5.9 M	227.6
all-hadronic		112.3	11.6 M	102.8
$t\bar{t}$ + jets, $m_t = 172.5$ GeV	MC@NLO	245.8	32.6 M	132.7

Table 3.3: Overview of the simulated event samples at $\sqrt{s} = 8$ TeV used for systematic studies. Showering and hadronisation is performed with PYTHIA for all these samples, except the MC@NLO sample where HERWIG is used.

The reconstruction of jets, the identification of b-quark jets, and the reconstruction of the missing transverse energy is finally described in Sections 3.3.4, 3.3.5 and 3.3.6, respectively.

3.3.1 Muon reconstruction

Muons play a central role in the identification of $t\bar{t} \rightarrow b\bar{b}q\bar{q}\mu\nu_\mu$ events so an efficient reconstruction of these muons is necessary. Since they can easily traverse thick layers of matter they are the only detectable particles escaping the calorimeters and producing hits in the different muon detectors. The muon reconstruction consists of several steps [76, 105, 106].

During the first step, the local reconstruction, hits within individual DT and CSC chambers are combined into track segments. The track segments of the innermost chambers are then used to generate seeds, which serve as starting point for the next step: the standalone muon reconstruction. Like in the case of the track reconstruction (Section 2.2.1) Kalman filtering is used to propagate these tracks outwards to the next layers where compatible track segments or hits are included in the track fit. After reaching the outermost layer an outside-in Kalman filter is applied using the compatible track segments, hits and an additional beam-spot constraint for the final estimate of the track parameters. This results in so-called *standalone muons*.

Afterwards a matching charged-particle track in the inner tracking system is identified for each standalone muon track by comparing parameters of the two tracks propagated onto a common surface. Starting from the standalone muon track an outside-in Kalman filter is applied where the corresponding hits in the central tracking system are included in the fit, resulting in *global muons*.

Instead of the outside-in approach, which starts from standalone muon tracks, an inside-out approach is also used to reconstruct muons. In this case the starting point are all the reconstructed tracker tracks with transverse momentum $p_T > 0.5$ GeV and total momentum $p > 2.5$ GeV. They are all extrapolated to the muon system and if at least one track segment in the muon system matches the extrapolated track, the corresponding tracker track is identified as a *tracker muon*.

With this combined way of reconstructing muons a very high muon reconstruction efficiency is obtained. Muons with low momenta are very efficiently reconstructed via the tracker muon approach, while the global muon reconstruction is designed to have a high efficiency for muons penetrating multiple layers of the muon system. The muon reconstruction and identification efficiency has been measured in data to be larger than 98% for muons with $p_T > 5$ GeV [105].

Apart from the identification of muons and the precise measurement of their momenta, a correct assignment of their electrical charge is also of vital importance within this thesis. Since positively and negatively charged muons will bend in other directions under the influence of the magnetic field within the CMS detector, their charge is measured via the direction of curvature of their reconstructed tracks. The charge misidentification probability is measured in a data sample of cosmic ray muons where one cosmic ray muon traverses the entire detector but is reconstructed as two muon candidates with opposite charge [106]. The measured charge misidentification probability as a function of the transverse momentum of the cosmic ray muon is shown in Figure 3.4. Most of the muons produced in $t\bar{t}$

decay have $p_T < 100$ GeV, where the charge misidentification probability of the global muons is smaller than 10^{-3}.

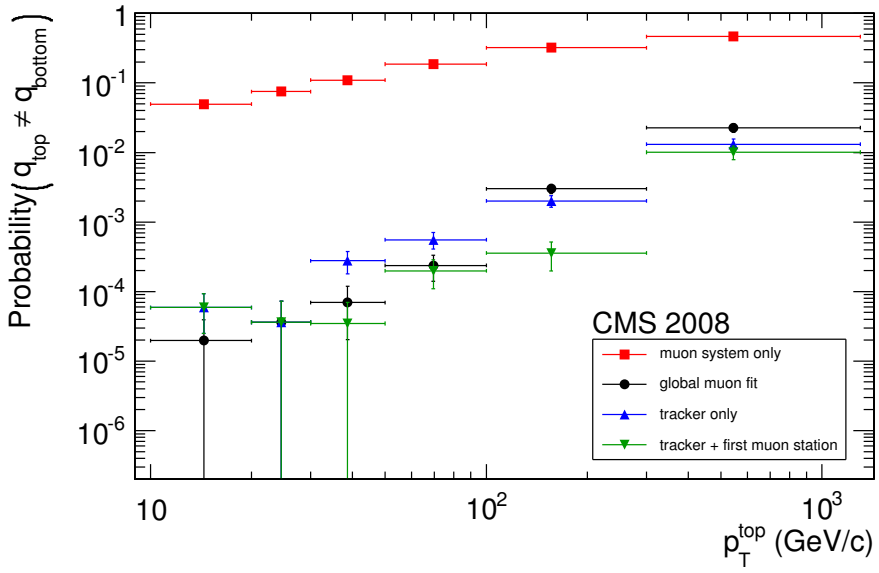

Figure 3.4: Muon charge misidentification probability as measured in a data sample of cosmic ray muons detected with CMS.

3.3.2 Electron reconstruction

As was the case for muons, an important role in the identification of $t\bar{t} \to b\bar{b}q\bar{q}'e\nu_e$ events is played by the electron reconstruction algorithm. Two closely related variants are available; one which was developed with as main goal an efficient reconstruction of isolated electrons and another one as part of the particle-flow algorithm where additional optimisations were done to include also an efficient reconstruction of non-isolated electrons produced within jets. A description of the former will be given here [76, 107], while more information about the latter can be found in Section 3.3.3. Since the electrons produced at the centre of the CMS detector have to traverse a significant amount of material before they reach the ECAL, Bremsstrahlung can be emitted by these electrons. Together with the strong magnetic field this will cause a spread of the initial electron energy in the ϕ direction, which has to be taken into account during their reconstruction.

The electron reconstruction is started by looking for electron track seeds to be used as initial estimates of the electron track parameters during the track reconstruction. Two complementary seeding algorithms are used: *ECAL-driven* seeding, which is more suitable for high p_T and isolated electrons, and *tracker-driven* seeding, optimised for low p_T electrons and electrons inside jets. The ECAL-driven seeding is started by clustering energy deposits in the ECAL subdetector within a region which is quite narrow in η but much larger in ϕ into *superclusters*. Afterwards the energy-weighted average impact point of the electron and its associated Bremsstrahlung photons is propagated towards the pixel detector, where compatible hits are identified and pairs of such hits are used as seeds. The tracker-driven

seeding algorithm, on the other hand, starts from the charged-particle tracks reconstructed with the iterative-tracking algorithm, as explained in Section 2.2.1. It identifies those tracks compatible with the electron particle hypothesis, either by comparing the parameters of the track and the matched supercluster or via a multivariate discriminator based on the tracks parameters. These tracks are then used as seeds for the electron track reconstruction.

The reconstruction of electron tracks based on the previously determined seeds is performed with the Gaussian-sum filter (GSF) [108], which uses a similar approach as the Kalman filter from the standard track reconstruction explained in Section 2.2.1. The main difference is that the Kalman filter assumes that the energy loss distribution of a charged particle traversing a thin layer of material is Gaussian, which is not the case for the emission of Bremsstrahlung by electrons, while the GSF algorithm assumes this energy loss distribution of electrons to be described by a Gaussian mixture (a weighted sum of Gaussian distributions). Like the standard tracking algorithm, the GSF algorithm starts from the track seed which is propagated to the next tracking layer where compatible hits are identified and included in the track fit. These steps are then repeated until the outermost tracker layer is reached.

With this combined ECAL+tracker-based algorithm electrons can be reconstructed very efficiently. An efficiency larger than 90% was measured for electrons with $p_T > 20$ GeV [109]. The assignment of their electrical charge is determined in the same way as for the muons, via the direction of curvature of their reconstructed tracks. The charge misassignment probability of electrons has also been measured in data [110]. Using a $Z/\gamma^* \rightarrow e^+e^-$ data sample the charge misassignment probability was measured to be within 0.1 – 0.4% and increasing with the electron's pseudorapidity.

3.3.3 The particle-flow event reconstruction

The main idea behind the particle-flow (PF) event reconstruction is to combine all the information of the different CMS subdetectors in order to reconstruct and identify all stable particles traversing the CMS detector. Afterwards the list of individual particles can be used to determine the missing transverse energy, to reconstruct jets, etc. The main advantage of this technique is that the momentum of a particle is estimated much more precisely since the information from multiple subdetectors is combined.

The algorithm consists of several steps [111, 112]. It starts with the reconstruction of its fundamental elements, the calorimeter clusters and the charged-particle tracks, which are then linked together into blocks. In the final step these blocks are identified as the different particles produced during proton-proton collisions: electrons, muons, photons, charged hadrons and neutral hadrons.

Building and linking of the fundamental elements

Since the momenta of charged particles are measured much more precisely in the tracker than in the calorimeters for transverse momenta up to several 100 GeV, a very efficient track reconstruction is necessary. The iterative-tracking algorithm as explained in Section 2.2.1 was developed for this purpose, since this algorithm has a high tracking efficiency and a low fake rate.

A clustering of the individual calorimeter hits is also performed separately in each of the calorimeter subdetectors. The algorithm starts by identifying seeds, which are defined as calorimeter hits with an energy above a certain threshold. Then topological clusters are built starting from these seeds by adding hits in calorimeter cells with at least one side in common with the already clustered cells. Finally particle-flow clusters are constructed within these topological clusters. Hereby each seed gives rise to exactly one PF cluster, where the energy of all the cells within the topological cluster is shared between the PF clusters based on the cell-cluster distance.

These charged-particle tracks and PF clusters are used to build electron candidates using the algorithm explained in Section 3.3.2. The resulting GSF tracks are used as electron candidates. Muon candidates are reconstructed with the standard algorithms from Section 3.3.1.

Since every particle traversing the CMS detector will give rise to several components in different subdetectors a linking algorithm is applied which identifies all components belonging to the same particle and combines them into blocks. Charged-particle tracks and PF clusters are matched by extrapolating the track to the relevant calorimeter subdetector and checking whether its extrapolated position is within the cluster boundaries. Linking an electron candidate track with a PF cluster is done in the same way as for other charged particles, but additional clusters caused by Bremsstrahlung photons are also added by extrapolating tangents of the GSF track to the ECAL surface. Two calorimeter clusters are linked together when the cluster position in the more granular calorimeter is within the cluster envelope of the less granular one. Charged-particle tracks in the inner tracking system and tracks in the muon system are matched based on the χ^2 of a global muon fit of the two tracks.

Particle identification

The identification of different particles starting from the linked subcomponents consists of several steps. First of all, muons and electrons are identified. Afterwards the algorithm looks for charged hadrons and finally the list of neutral hadrons and photons is constructed.

Muon candidates can be promoted to PF muons via three different selections: isolated, PF-tight and PF-loose. In the isolated selection only global muons are considered and they need to pass a loose isolation requirement. The remaining muons are passed to the PF-tight and PF-loose selections which both aim at identifying muons in jets. In the PF-tight selection information from the muon system and the calorimeters is used to reject hadronic punch-through. The muons selected by PF-loose criteria are used during charge hadron identification. There they will be promoted to PF muons if their track momentum is significantly larger than the corresponding energy deposit, which is incompatible with a charged hadron hypothesis.

The identification of PF electrons starts from the collection of electron candidates and is based on the most sensitive observables to separate electrons from charged hadrons: variables based on the energy matching between the tracks and the ECAL deposits, calorimeter-based and shower-shape variables, and pure tracking variables. A multivariate method is used to combine all these variables into a single discriminator, which is then used to identify PF electrons. Their momentum is obtained via a weighted combination of the electron track and the PF cluster momenta.

The remaining charged-particle tracks and PF clusters are used to build charged hadrons, photons and neutral hadrons. For tracks connected to PF clusters their momentum is compared to the PF cluster momentum. If the track momentum is similar to the PF cluster momentum a charged PF hadron is identified; if the track momentum is significantly smaller than the PF cluster momentum a charged PF hadron is identified together with a neutral PF particle, a PF photon or a neutral PF hadron; and if the track momentum is significantly larger than the PF cluster momentum additional muons are searched for starting from the muons selected by the PF-loose criteria. The momentum of a charged PF hadron is either taken from the charged-particle track or from a weighted average of the track and PF cluster momenta if these are compatible within uncertainties. The remaining PF clusters in the ECAL and HCAL give rise to PF photons and neutral PF hadrons, respectively.

Top projections

An additional interesting consequence of reconstructing individual particles with the PF algorithm is the so-called top projection [113]. This means that PF particles passing or failing certain identification criteria can be removed either from the entire event or from the reconstruction of certain objects. This option is used within this thesis to project out two types of particles: charged PF particles originating from pileup and isolated high p_T leptons from W-boson decay.

The identification of the charged PF particles originating from pileup collisions uses their charged-particle tracks, and the primary vertices reconstructed from these tracks as explained in Section 2.2.1. The main primary vertex is identified as the one with the highest $\sum p_T^2$ of its associated tracks. All PF particles originating from other primary vertices than the main primary vertex are identified as pileup PF particles and are discarded from further event reconstruction. This technique is called charged hadron subtraction (CHS) and is used within this thesis.

Isolated high p_T leptons are also identified during the event reconstruction so that they can be excluded from the list of particles used as input of the jet clustering. The starting point is the list of PF muons and PF electrons. In the 7 TeV analysis only PF muons with $p_T > 10$ GeV and $|\eta| < 2.5$ and PF electrons with transverse energy $E_T > 15$ GeV and $|\eta| < 2.5$ are considered. The relative isolation of each PF lepton ℓ is calculated as

$$I_{rel} = \frac{\sum_{CH} p_T^{CH} + \sum_{NH} p_T^{NH} + \sum_{\gamma} p_T^{\gamma}}{p_T^{\ell}}, \tag{3.12}$$

where the three sums run over all charged PF hadrons (CH) after charged hadron subtraction, all neutral PF hadrons (NH) and all photons (γ) still present within a cone of radius $\Delta R = 0.4$ around the lepton. The leptons are identified as isolated if $I_{rel} < 0.2$, which means they are not considered during jet reconstruction. Since in the 8 TeV analysis isolated PF electrons are not projected out before jet reconstruction, as explained in Section 4.1, only isolated PF muons need to be identified. The same kinematic criteria as at 7 TeV are used ($p_T > 10$ GeV and $|\eta| < 2.5$), but a modified definition of the relative isolation is used, which is more robust with respect to pileup:

$$I_{rel}^{\Delta\beta} = \frac{\sum_{CH} p_T^{CH} + \max\left(0, \sum_{NH} p_T^{NH} + \sum_{\gamma} p_T^{\gamma} - 0.5 \sum_{PU} p_T^{PU}\right)}{p_T^{\ell}}. \tag{3.13}$$

Here all sums use the same definition as before. The sum over all pileup PF particles (PU) runs over all charged PF particles identified as pileup by the CHS algorithm within a cone of radius $\Delta R = 0.4$ around the lepton and aims at removing the effect of neutral pileup energy deposits within this cone. The factor 0.5 in front of this sum is motivated by the fact that jets contain on average two times more charged PF particles than neutral PF particles [114]. PF muons are considered to be isolated and not used during the jet reconstruction if $I_{rel}^{\Delta\beta} < 0.2$.

3.3.4 Jet reconstruction

The b and light quarks produced during the $t\bar{t} \to b\bar{b}q\bar{q}'\ell\nu_\ell$ decay are not detected directly, since they will shower and hadronise resulting in narrow cones of particles called *jets*. Jet clustering algorithms are used to combine all these final-state particles into jets so that their momenta can be interpreted as the initial quarks' momenta. Since the energy response of any particle detector, including the CMS detector, is never perfect, additional *jet energy scale corrections* need to be applied to these reconstructed jets.

Jet clustering

A large variety of jet clustering algorithms is available in the literature [115]. They all start from a set of particles (or calorimeter deposits) and cluster them together in jets. This usually involves some parameters which need to be chosen, like the radius R of the jets. When choosing a jet algorithm two important theoretical properties of the algorithm need to be considered. It needs to be insensitive to additional soft radiation (infrared safe) and to the nearly collinear splitting of a hard particle (collinear safe) by yielding the same collection of jets independent of the occurrence of soft radiation or collinear splitting.

The jet algorithms used in this thesis are the $k_\perp$-like algorithms [116], which are infrared and collinear safe. They are all based on a similar distance definition:

$$d_{ij} = \min\left(k_{\perp i}^{2p}, k_{\perp j}^{2p}\right)\frac{\Delta R_{ij}^2}{R^2} \quad \text{and} \quad d_{iB} = k_{\perp i}^{2p}. \tag{3.14}$$

Here d_{ij} is the distance between two particles i and j, d_{iB} is the distance between a particle i and the beam B, $\Delta R_{ij} = \sqrt{\Delta y_{ij}^2 + \Delta\phi_{ij}^2}$ is the (y, ϕ)-based distance between two particles i and j, R is the radius parameter, and p is a parameter that governs the relative power of the energy versus geometrical scales. The clustering starts by looking for the minimum distance among all distances d_{ij} and d_{iB}. If this minimum is d_{ij} the momenta of particles i and j is summed and used to form a new object which replaces particles i and j in the list of input objects. If this minimum is d_{iB} the object i is removed from the list of input objects and promoted to the list of jets. Afterwards the distances are recalculated and the procedure is repeated until the list of input objects is empty.

The way in which the addition of the four-momenta in the clustering process is performed is called the *recombination scheme*. The most commonly used recombination schemes are the E_T and the E schemes. In the E_T scheme the total transverse energy of a jet is calculated via a scalar sum of the transverse energies of its constituents $E_T = \sum_j E_{T,j}$. The pseudorapidity η and azimuthal angle ϕ of a jet are calculated via an E_T-weighted

over its constituents: $\eta = \sum_j E_{T,j}\,\eta_j / \sum_j E_{T,j}$ and $\phi = \sum_j E_{T,j}\,\phi_j / \sum_j E_{T,j}$. In the E scheme, on the other hand, the four-momenta of the jets are calculated by the addition of the four-momenta of its constituents. The energy and momentum of all the jets within this thesis are calculated via the E recombination scheme.

The two $k_\perp$-like algorithms used in this thesis are the variants with the parameter p of Equation 3.14 equal to 1 (the $k_\perp$ algorithm) and to -1 (the anti-k_t algorithm). The former is used in the matching between the matrix element and the parton shower as explained in Section 3.1.2; while the latter is used within CMS to reconstruct either *PF jets* starting from the list of PF particles, or *generator-level jets* (*GenJets*) starting from the list of particles generated by PYTHIA after showering and hadronisation but before they enter the CMS detector simulation. The additional particles produced by pileup are not included in these generator-level jets, so they can also be used to correct for the presence of pileup. Apart from its theoretical properties other important reasons for choosing this algorithm are its execution speed and the fact that it produces conical-like jets with more robust boundaries with respect to soft particles, like for example the additional particles produced by pileup.

Jet energy scale corrections

Due to both the additional energy deposits caused by pileup and the detector imperfections like its limited energy resolution and small dead regions without any detection, the energy of a reconstructed PF jet will, on average, be lower than that of the corresponding generator-level jet. To correct for these effects jet energy scale (JES) corrections have been developed by CMS [117–119] using a factorised approach where each step corrects for a specific set of effects. An overview of the different jet energy correction levels used within this thesis is given in Figure 3.5. The first level applied to raw (uncorrected) PF jets is the L1 correction,

Figure 3.5: Graphical overview of the different jet energy corrections to be applied to data and simulation (MC).

which is different for data and simulation, and corrects for the additional energy deposits caused by pileup. Afterwards the simulation-based L2L3 correction is applied. Then only the data is corrected using the L2L3Residual correction which is estimated from data and aims at making the p_T-response in data equal to the response in simulation. Finally the optional L5L7 correction is applied to certain jets as explained in more detail in Section 4.2.1. These corrections and the corresponding jet energy scale uncertainties will be discussed here. The

JES uncertainties are mainly focused at covering the possible differences between data and simulation in the jet response. This is exactly what is needed within this thesis since the top-quark mass measurement is explicitly calibrated with simulation as explained in Chapter 5.

For the derivation of these corrections from simulated event samples a matching between PF jets and generator-level jets is needed, which is based on the ΔR distance in (η, ϕ)-space. Starting from the highest-p_T PF jet the closest generator-level jet within $\Delta R(\text{PF jet, GenJet}) < 0.4$ is identified. Both jets are matched to each other and not considered anymore in further matching. These steps are then repeated until all PF jets have been considered for matching. From these matched pairs of jets the jet p_T response is defined as the average of p_T^{PFjet}/p_T^{GenJet}.

L1 pileup correction

The need to correct the energy of a jet for the presence of pileup can clearly be seen in Figure 3.6. Here the p_T response of PF jets is shown for different

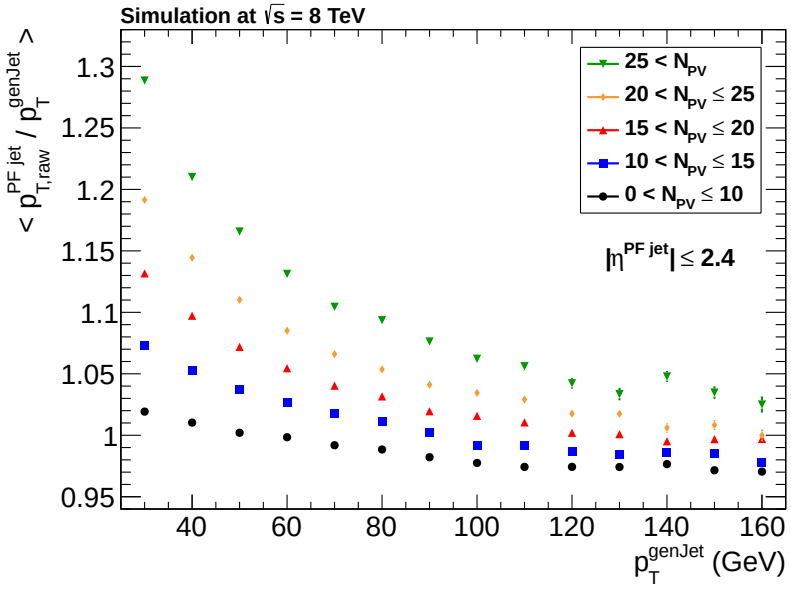

Figure 3.6: PF jet p_T response before any JES corrections for different primary vertex multiplicities, calculated using $t\bar{t}$ ℓ+jets events.

numbers of primary vertices, since this number is directly related to the amount of pileup. A clear dependency of the jet response on the primary vertex multiplicity is observed, caused by the additional particles produced by pileup interactions which are clustered also into the PF jets.

The L1 pileup correction [120, 121] aims at subtracting these additional energy deposits from the uncorrected PF jet transverse momentum $p_{T,raw}^{PFjet}$, resulting in the L1-corrected transverse momentum $p_{T,L1}^{PFjet}$, based on the following equation:

$$p_{T,L1}^{PFjet} = p_{T,raw}^{PFjet} - Offset(\rho, A_{PFjet}, \eta^{PFjet}, p_{T,raw}^{PFjet}). \tag{3.15}$$

Here A_{PFjet} is the area of the PF jet, which is calculated via artificially including a very large number of infinitely soft particles in the (y, ϕ)-space just before jet

clustering. The area in the (y, ϕ)-space occupied by these particles in each jet defines the jet area [122]. The p_T density ρ of each event is calculated with the $k_\perp$ algorithm with distance parameter $R = 0.6$. It is defined as the median of the p_T^j / A_j distribution, where j runs over all jets in the event.

The offset in Equation 3.15 for the 7 TeV analysis depends only on the jet area A_j and on the event p_T density ρ, and is defined as follows:

$$Offset^{7\ TeV} = A_j \left(p_0(\eta) + p_1(\eta) \cdot N_{PV}(\rho) + p_2(\eta) \cdot (N_{PV}(\rho))^2 \right). \tag{3.16}$$

Here $N_{PV}(\rho)$ is the dependency of the primary vertex multiplicity N_{PV} on the p_T density ρ, and $p_0(\eta)$, $p_1(\eta)$, and $p_2(\eta)$ are parameters obtained in bins of η^{PFjet}. These parameters are estimated via the p_T offset measured inside a cone of radius $R = 0.5$ in a randomly triggered event sample (*zero bias event sample*), and afterwards the method is calibrated on a simulated QCD multijet sample.

The pileup related jet energy scale uncertainty for the 7 TeV analysis consists of several sources added in quadrature, mostly covering some observed pileup-related differences between data and simulation. It includes the simulation-based residual bias on the p_T offset measurement, the observed difference between data and simulation of the measured p_T offset, the uncertainty on the effect of out-of-time pileup, and an uncertainty covering the observed jet rate variation versus N_{PV}.

For the 8 TeV analysis the offset in Equation 3.15 is defined in a different way, since for the high amount of pileup in the 8 TeV dataset a better performing approach was developed. In this case the offset is calculated from two simulated samples of QCD multijet events containing exactly the same events; once with pileup and once without pileup. A jet-by-jet matching between these two sample was performed and used to calculate the correction as follows:

$$Offset^{8\ TeV} = A_j \left(p_0(\eta) + p_1(\eta) \cdot \rho \cdot \left(1 + p_2(\eta) \log \left(p_{T,raw}^{PFjet} \right) \right) \right) \tag{3.17}$$

Here $p_0(\eta)$, $p_1(\eta)$, and $p_2(\eta)$ are parameters obtained in bins of η^{PFjet} by fitting the offset between the matched jets with and without pileup. This equation is not explicitly dependent anymore on N_{PV} as was the case for the 7 TeV correction, but the dependence on the amount of pileup is now included via the p_T density ρ.

This offset correction is applied to both simulation and data, and a residual correction is applied additionally to the data. This additional correction is calculated from the residual p_T offset observed in randomly triggered data events after applying the simulation-based pileup correction. The effect of these corrections on the jet p_T response is shown in Figure 3.7, where the large pileup dependency of the jet response from Figure 3.6 has almost completely vanished.

For the 8 TeV analysis the pileup related JES uncertainty consists of two sources added in quadrature, which are similar to some of the 7 TeV uncertainty sources. It includes the residual bias of the method to measure the p_T offset as observed in simulation, and 20% of the observed difference between data and simulation of the measured p_T offset (since separate corrections are applied to data and simulation).

L2L3 simulation-based correction

Although the jet p_T response is pileup independent after the L1 correction, Figure 3.7 clearly shows that additional corrections are necessary to make the response

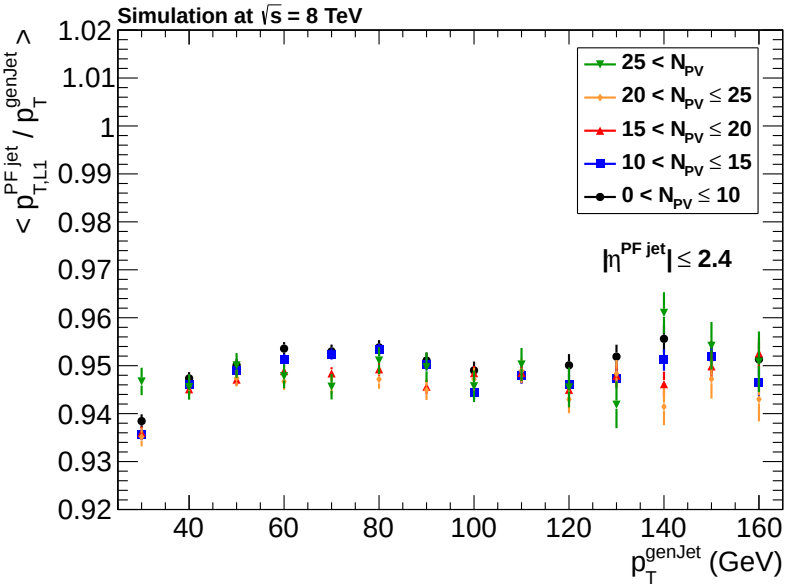

Figure 3.7: PF jet p_T response after the L1 pileup JES correction for different primary vertex multiplicities, calculated using $t\bar{t}$ ℓ+jets events.

equal to one. This is achieved with the L2L3 correction, which also makes the response flat in η^{PFjet}. The correction happens by dividing the L1-corrected p_T by the p_T- and η-dependent jet response which is obtained from a simulated QCD multijet sample:

$$p_{T,L1L2L3}^{PFjet} = p_{T,L1}^{PFjet} \cdot \frac{1}{\left\langle \frac{p_{T,L1}^{PFjet}}{p_T^{GenJet}} \right\rangle \left(p_{T,L1}^{PFjet}, \eta^{PFjet} \right)}. \tag{3.18}$$

The effect of this correction is illustrated in Figure 3.8, where the jet p_T response is shown after the entire L1L2L3 correction. A clear improvement in the jet response is visible compared to Figure 3.7. At low p_T the response is slightly too high, which is a direct consequence of the fact that these L2L3 corrections are derived from a QCD multijet samples, which is dominated by gluon jets, while the $t\bar{t}$ events of Figure 3.8 are dominated by quark jets. These quark jets have a higher p_T response than gluon jets as shown in Figure 3.9, so they are over-corrected by the standard L2L3 corrections.

Data-based relative correction (L2Residual)

Since the L2L3 correction is completely based on simulation additional data-driven corrections are estimated to be applied to the data. The L2Residual correction [123] aims at making the jet p_T response in data constant as a function of η^{PFjet}, taking the response in the barrel part of the CMS detector as reference. For this purpose QCD dijet events with at least one jet in the barrel ($|\eta| < 1.3$) are used, where the deviation from the assumed p_T-balance between those two jets is used to derive corrections in bins of η^{PFjet}.

Several sources of uncertainty on these corrections are again considered. For both the 7 and 8 TeV measurements the statistical uncertainties on the measured

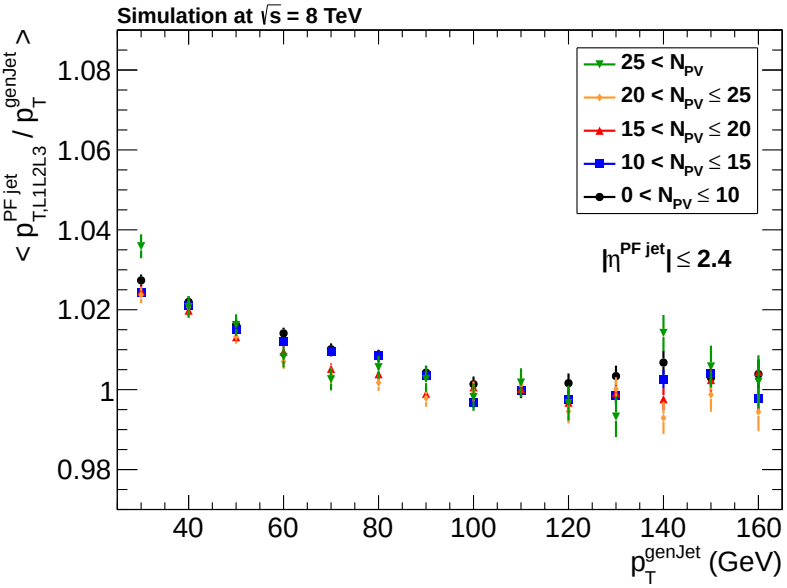

Figure 3.8: PF jet p_T response after the L1L2L3 JES corrections for different primary vertex multiplicities, calculated using $t\bar{t}$ ℓ+jets events.

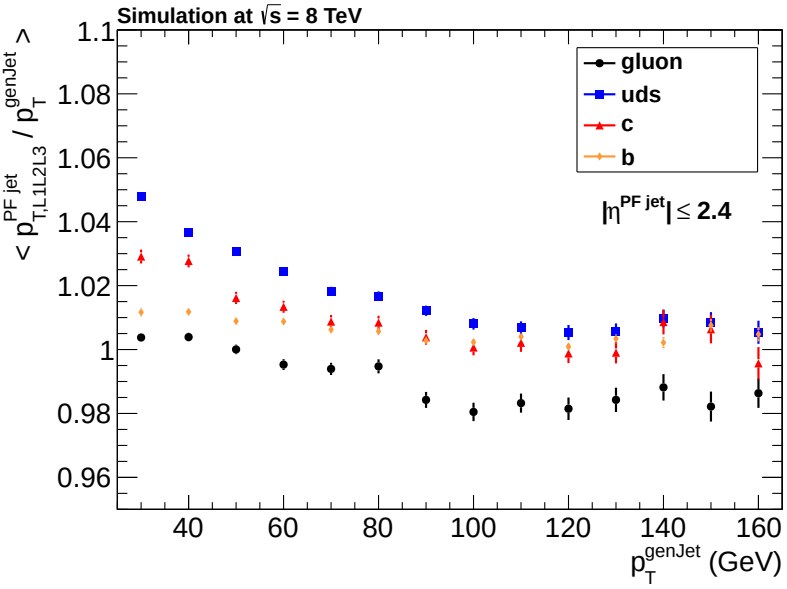

Figure 3.9: PF jet p_T response after the L1L2L3 JES corrections for different jet flavours, calculated using $t\bar{t}$ ℓ+jets events.

corrections, the systematic uncertainties due to the uncertainty on the measured jet energy resolution, and uncertainties on the effect of final-state radiation are taken into account. In the 8 TeV measurement a slight dependence of these corrections on the p_T of the jets was observed and additional uncertainties are added to cover this effect.

Data-based absolute correction (L3Residual)

The measurement of the absolute jet energy scale [123] is performed for jets inside

the barrel part of the CMS detector, since these numbers can be safely used outside the barrel after the L2Residual correction. Therefore γ+jet and Z+jet events (with $Z \to \mu^+\mu^-$ or $Z \to e^+e^-$) are selected which contain a barrel jet ($|\eta| < 1.3$). Since photons, electrons and muons are measured much more accurately than jets, the assumed p_T balance between the jet and the photon or Z-boson is used to measure an absolute jet energy scale correction to be applied to all jets in data.

The uncertainties on this measured absolute jet energy scale are similar for the 7 and 8 TeV measurements. The energy scale uncertainty on the photons and leptons used as reference objects in the p_T balancing is propagated to the measured corrections. Since the corrections are measured for a fixed jet p_T ($\sim$ 100 GeV at 7 TeV and $\sim$ 200 GeV at 8 TeV) and assumed to be p_T-independent, there is also an uncertainty covering the possible p_T-dependence of these corrections. This p_T-dependent uncertainty is estimated via two components: the fragmentation and hadronisation uncertainty, which is estimated by comparing two fragmentation and hadronisation models (PYTHIA versus HERWIG); and the uncertainty on the single-particle response of the calorimeters, where the uncertainty is estimated by scaling the single-particle response in simulation within its measured uncertainty.

To obtain the total p_T- and η-dependent jet energy scale uncertainties all the individual uncertainties estimated on each of the corrections are added in quadrature. This covers the uncertainty on the pileup removal and on the relative and absolute residual corrections. Apart from these, two additional sources of uncertainty are also added in quadrature, namely the time-dependent and the flavour-dependent uncertainty.

The time-dependent jet energy scale uncertainty was included since the jet energy scale was observed to increase in data taken later during the year. This can be caused by the reduction of the energy response of the calorimeters due to the high amount of radiation which they have to absorb. For jets inside the barrel this uncertainty is negligible, but it grows when going more and more in the forward region. At $|\eta| = 2.4$ the corresponding uncertainty equals 1.5% for the 7 TeV analysis and 0.5% for the 8 TeV analysis.

To cover the difference in jet response between the different types of partons giving rise to a jet, the flavour-dependent uncertainty was introduced. These different parton types are grouped together in light quarks (*uds*), gluons (g), c quarks and b quarks; which all give a slightly different jet response. In simulated QCD multijet events it is observed that at $p_T = 30$ GeV the p_T response of light-quark jets is about 3% higher than the response of gluon jets, while the response of c- and b-quark jets is in-between the light-quark and gluon jets response. For high p_T jets this difference is reduced to less than 2%. The flavour-dependent jet energy scale uncertainty aims at covering this effect and is calculated in a different way for the 7 TeV and the 8 TeV analyses. In both cases the uncertainty is estimated by comparing two existing hadronisation models, namely PYTHIA and HERWIG.

The 7 TeV flavour-dependent uncertainty is estimated by comparing the jet p_T response $\mathcal{R}$ between two QCD multijet samples; one generated with PYTHIA and the other one with HERWIG. The uncertainty is defined as

$$\sigma^{7\,TeV}_{JES,Flavour} = \max\left(\left|\frac{\mathcal{R}_f^{\mathrm{PYTHIA}}}{\mathcal{R}_{QCD}^{\mathrm{PYTHIA}}} - \frac{\mathcal{R}_f^{\mathrm{HERWIG}}}{\mathcal{R}_{QCD}^{\mathrm{HERWIG}}}\right|, f \in \{uds, g, c, b\}\right), \tag{3.19}$$

where $\mathcal{R}_f$ is the p_T response of the jet flavour f and $\mathcal{R}_{QCD}$ is the response in the inclusive QCD multijet sample. In this way the maximum difference between PYTHIA and HERWIG is taken, relative to the flavour composition in the QCD multijet sample used to estimate the simulation-based L2L3 jet energy scale correction.

The flavour-dependent jet energy scale uncertainty for the 8 TeV analysis is calculated in a different way. An uncertainty is defined independently for each flavour f as

$$\sigma^{8\,TeV}_{JES,f} = \left| \frac{\mathcal{R}_f^{\mathrm{HERWIG}}}{\mathcal{R}_f^{\mathrm{PYTHIA}}} - \frac{\mathcal{R}_{Ref}^{\mathrm{HERWIG}}}{\mathcal{R}_{Ref}^{\mathrm{PYTHIA}}} \right|, \tag{3.20}$$

where $\mathcal{R}_f$ is the p_T response of the jet flavour f and $\mathcal{R}_{Ref}$ is the response obtained at a reference point. The (p_T, η) bin used for the absolute scale measurement with the Z+jet sample is currently taken as reference point. This definition of the flavour-dependent uncertainty results in different uncertainties corresponding to the different pure flavour samples, which are shown in the left plots of Figure 3.10. The largest uncertainty is observed for jets originating from gluons were a large difference between PYTHIA and HERWIG is observed.

Finally these individual flavour uncertainties are combined to obtain the flavour-dependent jet energy scale uncertainty of a specific sample by using the expected flavour composition of this sample. The flavour composition of the ℓ+jets sample obtained with the event selection of Section 4.1 is shown in the left plots of Figure 3.10. The flavour composition of jets inside the barrel part of the CMS detector is dominated by b jets, while a larger fraction of gluon jets and light-quark jets is observed when going more towards the forward direction.

The total jet energy scale uncertainty used for the 7 TeV analysis is shown in Figure 3.11. In the barrel part of the detector the jet energy scale at low p_T is dominated by the uncertainty on the measured absolute jet energy scale and the flavour-related jet energy scale uncertainty, while at higher p_T the uncertainty on the absolute jet energy scale dominates the total uncertainty. When going more towards the endcaps those two uncertainties remain almost constant in size, while the pileup uncertainty increases significantly at low p_T where it is the dominant uncertainty and the time-dependent uncertainty increases over the entire p_T range to become the largest uncertainty at higher p_T.

For the 8 TeV analysis the total jet energy scale uncertainty and its subcomponents are shown in Figure 3.12. In the barrel part a similar picture as for the 7 TeV analysis arises: the uncertainty is again dominated by the absolute and the flavour components at low p_T, while at higher p_T the uncertainty on the absolute jet energy scale dominates although the flavour-related jet energy scale uncertainty is now not negligible at higher p_T. In the endcap a different picture arises since both the pileup and the time-dependent jet energy scale uncertainties are significantly reduced. Now almost all components contribute an equal amount to the total jet energy scale uncertainty.

Jet energy resolution

Another important aspect of the reconstructed PF jets is their energy or transverse momentum resolution. This has also been measured with the 7 TeV dataset by using the p_T balance in QCD dijet or γ+jet events [117]. The ratio of the measured resolution in data to the resolution in simulation is shown in Figure 3.13. Similar measurements have been

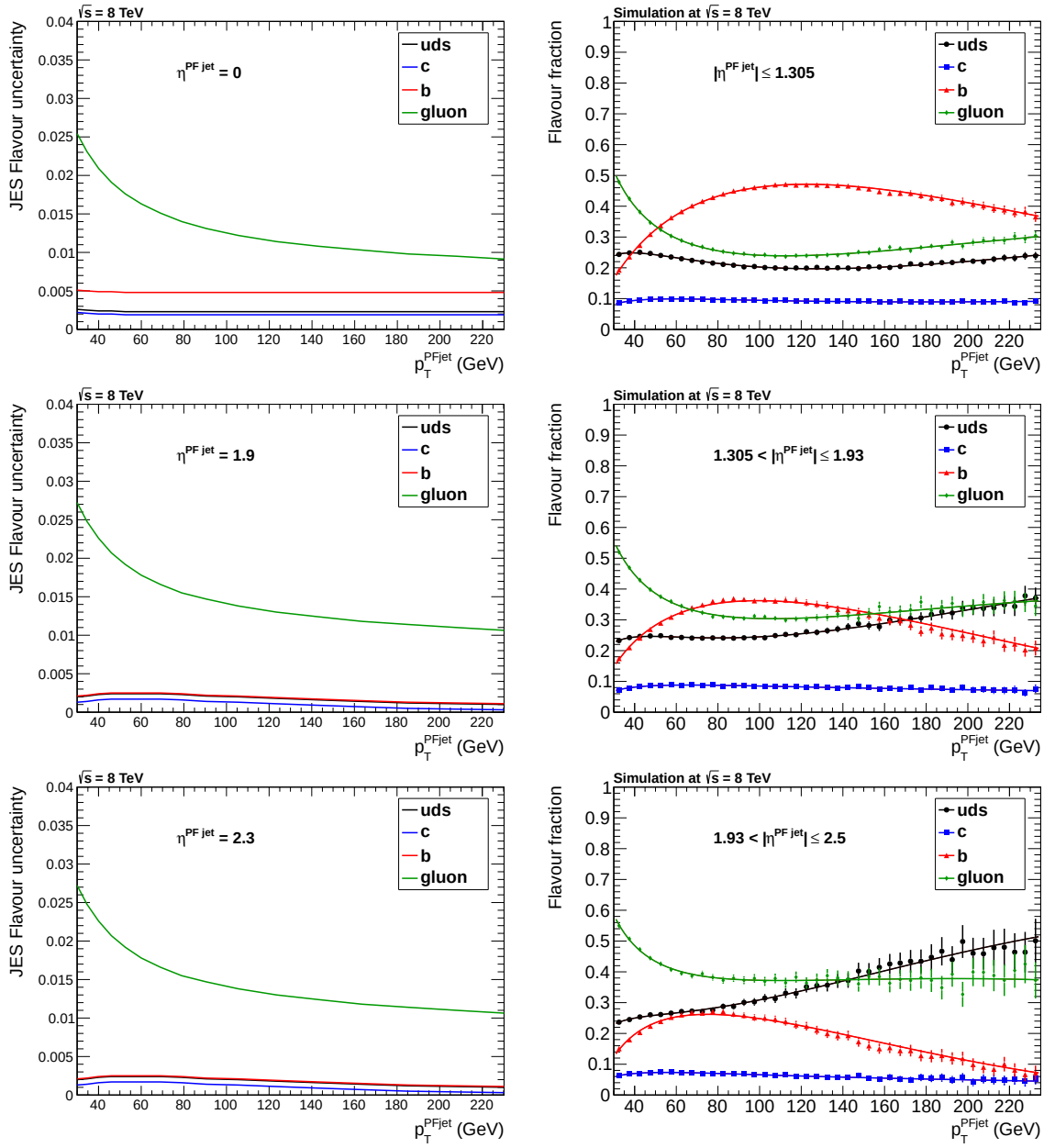

Figure 3.10: Individual jet energy scale uncertainties for each flavour (left) and the observed flavour composition after the ℓ+jets event selection of Section 4.1 (right).

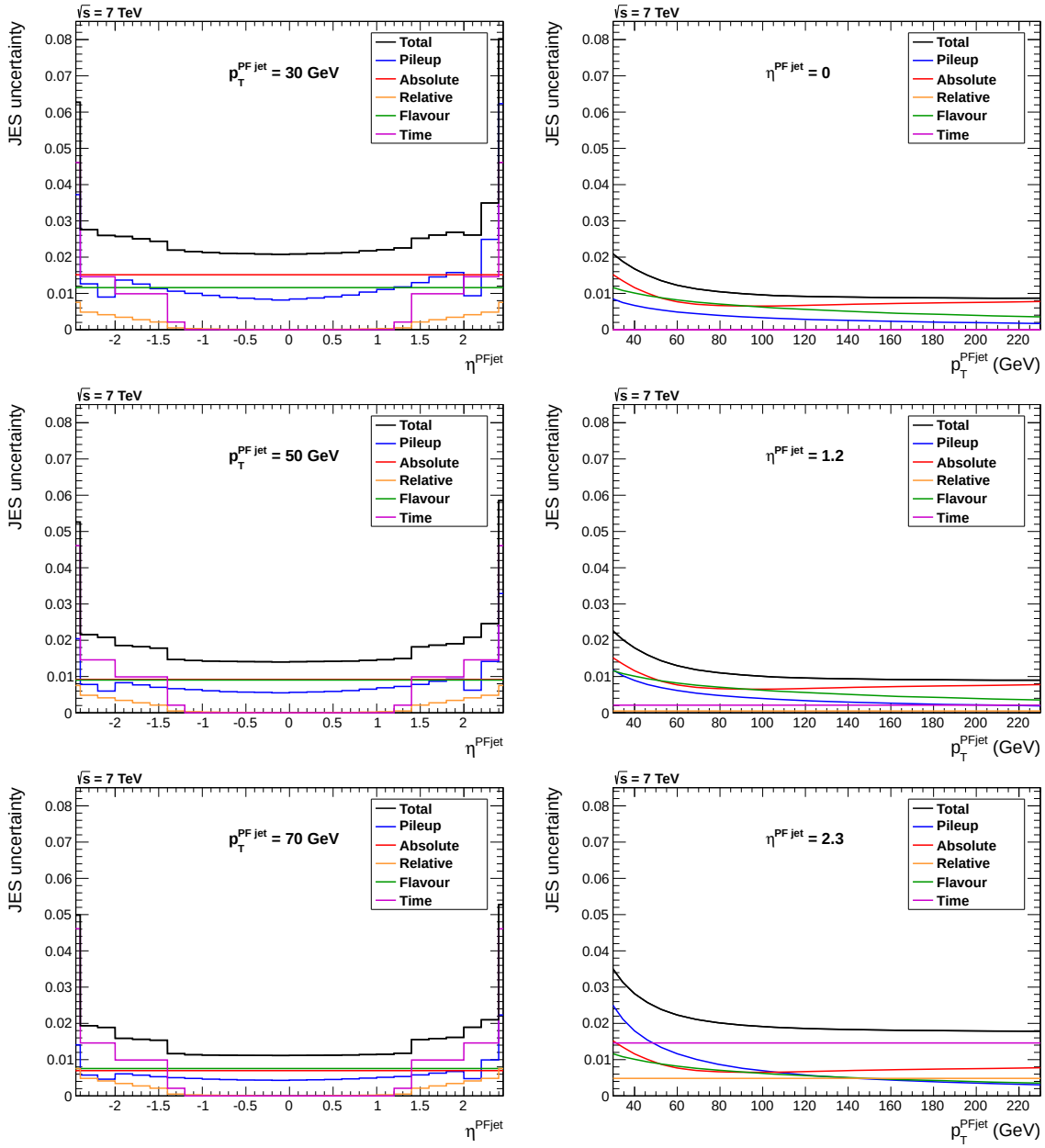

Figure 3.11: Total jet energy scale uncertainty and its subcomponents for the 7 TeV dataset, shown for different η and p_T regions.

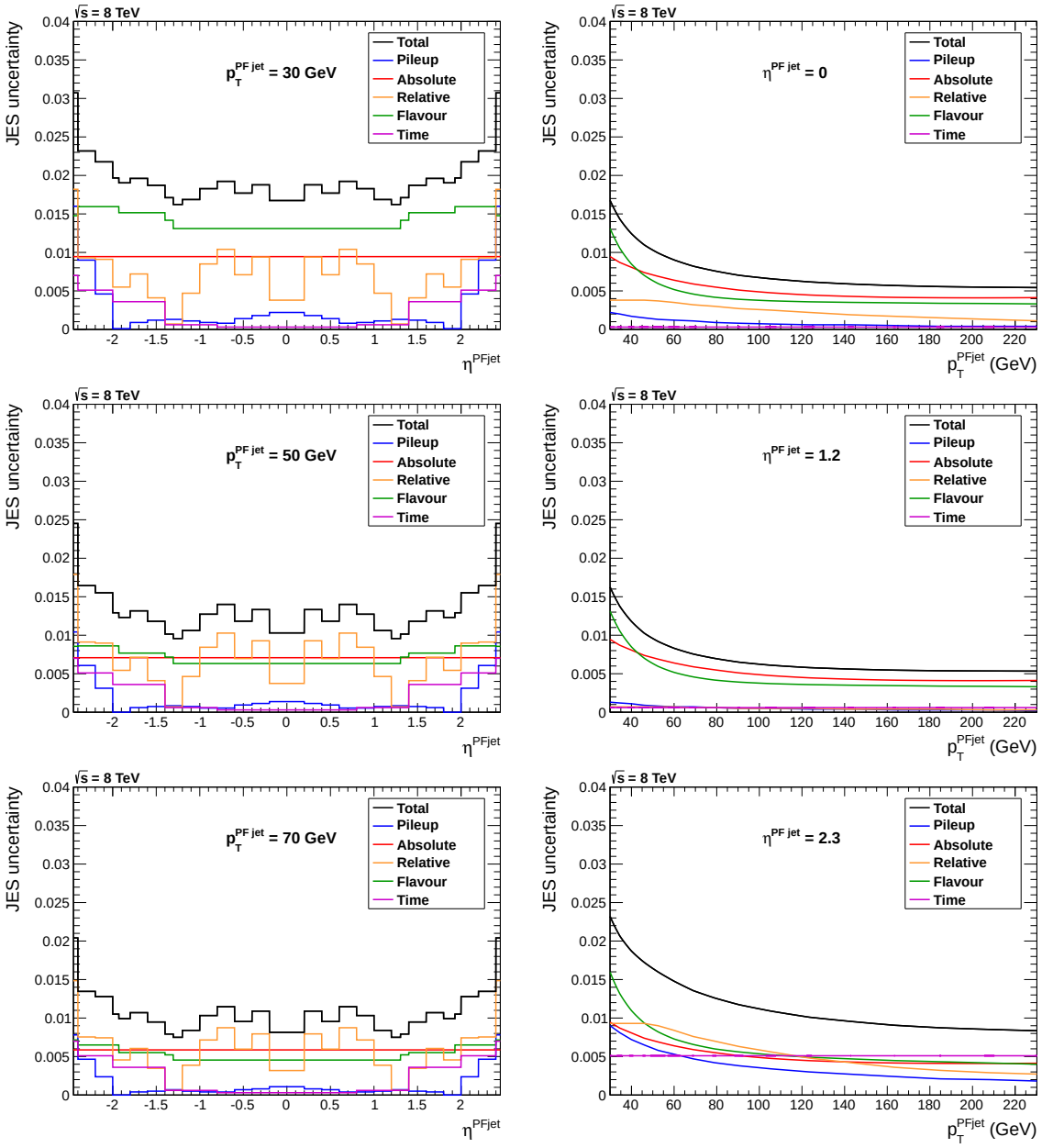

Figure 3.12: Total jet energy scale uncertainty and its subcomponents for the 8 TeV dataset, shown for different η and p_T regions.

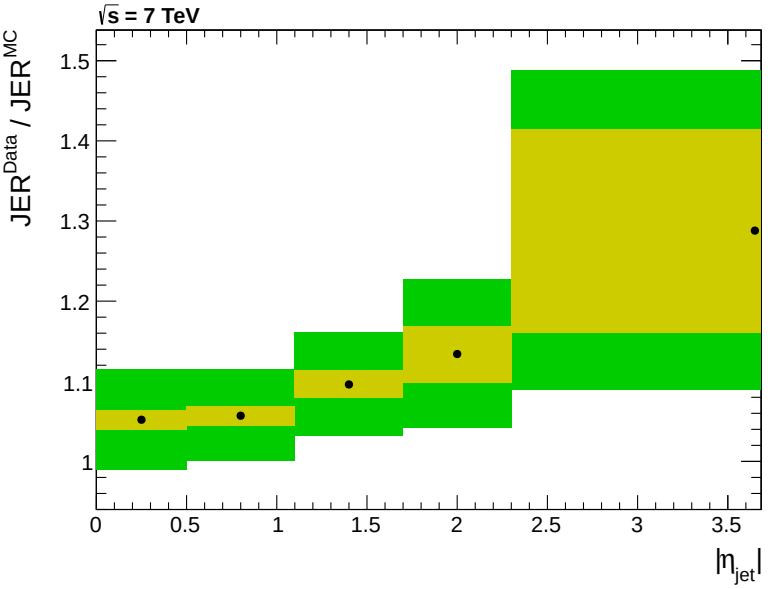

Figure 3.13: Measured data to simulation ratio of the jet energy resolution (JER), with statistical (inner yellow band) and systematic uncertainties (outer green band).

performed with the 8 TeV dataset and compatible results with a comparable precision were obtained.

The resolution measured in data is about 10% worse than the resolution in simulation and therefore a worsening of the jet energy resolution is applied to all the simulated events. This is done by comparing the energy and momenta of the reconstructed and corrected PF jets with the corresponding generator-level jets and applying a smearing to the energy and momenta of the PF jets using the measured data to simulation ratios from Figure 3.13. Additional information about the E_T and angular resolutions of jets can be found in Section 4.2.1.

3.3.5 b-jet identification

Since each $t\bar{t}$ event contains at least two b quarks, a correct identification of b jets is a powerful tool to either reject background events not containing b jets or to identify the correct b jets in the event topology. An efficient separation between b and non-b jets is possible due to a couple of distinct features of b jets. Since the b hadron which is formed inside a b jet has a reasonably long lifetime its decay will result in a couple of displaced tracks which are not originating from the primary collision vertex but from a secondary displaced vertex.

In this thesis the *Combined Secondary Vertex* algorithm (CSV) is used [71], where the identification of b jets consists of several steps. During the first step of the algorithm, charged-particle tracks are associated to the jet under study and used in the identification of secondary vertices which are displaced with respect to the primary vertex. For this purpose high-quality tracks within a cone of radius $\Delta R = 0.3$ around the jet axis are used to search for vertex candidates, where the jet axis is defined by the direction of the jet momentum vector. When a vertex is found its associated tracks are removed from the list of tracks and

the procedure is repeated until no vertex candidates are found. During the first iteration of this procedure the interaction region is used as a constraint in order to remove promptly produced tracks not originating from b-hadron decay.

In the next step several variables with discrimination power between b and non-b jets, like track impact-parameter significances or secondary vertex properties, are combined to obtain an optimal separation between both categories. This is done separately for three jet categories: jets where a secondary vertex was identified via its associated tracks, jets where no secondary vertex was identified but where a pseudo-vertex is found using less tight constraints, and jets where no secondary vertex and no pseudo-vertex is identified.

The resulting CSV discriminator is shown in Figure 3.14 separately for the different jet flavours. A clear separation between b jets and light-quark or gluon jets is obtained.

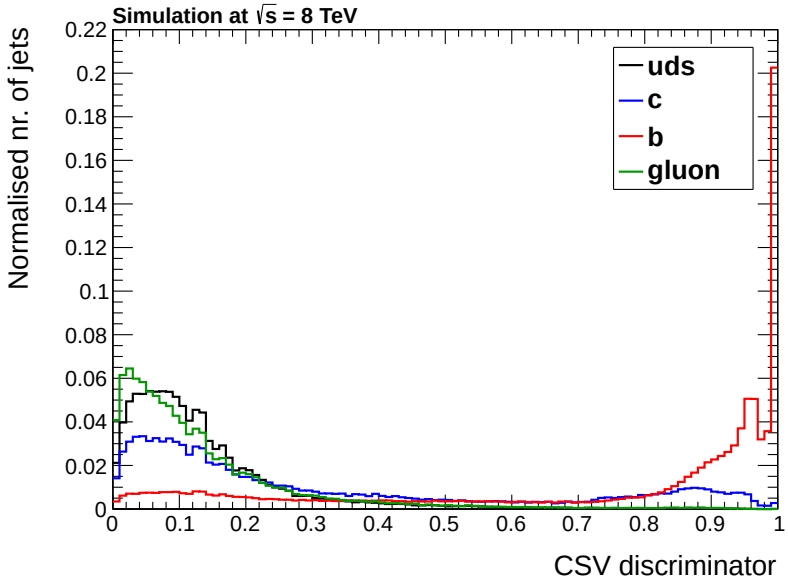

Figure 3.14: Combined Secondary Vertex b-tag discriminator for the different jet flavours.

Between b and c jets the separation is less optimal since c jets can also contain hadrons with a slightly longer lifetime. Both the 7 and the 8 TeV analysis use this algorithm at its medium working point. In the 7 TeV case this algorithm has an efficiency $\varepsilon_b = 70.8 \pm 1.9\%$ to correctly identify b jets, while the misidentification efficiency of non-b jets is $\varepsilon_l = 4.3\%$ For the 8 TeV analysis a b-tag efficiency $\varepsilon_b = 68.2 \pm 1.2\%$ is observed, with a mistag rate $\varepsilon_l = 4.2\%$.

3.3.6 Missing transverse energy

The neutrinos produced during the leptonic decay of the W-bosons originating from top quark decay are not detected by any of the CMS subdetectors, since they are known to traverse gigantic amounts of matter before they interact with it. This means that the measurement of their energy and momentum is a complex task; their presence can only be interfered indirectly by the apparent violation of momentum conservation in the transverse plane caused by the neutrinos escaping detection. Therefore the CMS detector was made as hermetic as possible around the interaction point to allow a measurement of the energy

and momentum of all particles produced in the transverse direction during proton-proton collisions.

The missing transverse energy [124] is calculated from the transverse momenta $\vec{p}_{T,i}$ of all the PF particles in the following way:

$$\vec{\not{E}}_{T,raw} = -\sum_{i} \vec{p}_{T,i}, \tag{3.21}$$

where the sum runs over all the reconstructed PF particles. Since the energy response of uncorrected jets is known to be slightly too small, as explained in Section 3.3.4, the L2L3 jet energy scale corrections are also propagated to the $\not{E}_T$:

$$\vec{\not{E}}_T = \vec{\not{E}}_{T,raw} - \sum_{PFjets} \left(\vec{p}^{\,PFjet}_{T,L1L2L3} - \vec{p}^{\,PFjet}_{T,L1} \right), \tag{3.22}$$

where the sum runs over all PF jets with $p^{PFjet}_{T,L1L2L3} > 10$ GeV.

Because of momentum conservation in the transverse plane an event without undetected particles will thus result in a low $\not{E}_T$ value, while an event containing an undetected particle will have a high $\not{E}_T$ value. In ℓ+jets $t\bar{t}$ events the $\not{E}_T$ resolution ranges between 40% and 15%, which is calculated by comparing the reconstructed $\not{E}_T$ with the transverse momentum of the neutrino produced in $t\bar{t}$ decay in simulation. Within this thesis the $\not{E}_T$ will only be used in the 8 TeV analysis to normalise the data-driven QCD multijet background as explained in Section 4.1.

Chapter 4

Selection and topology reconstruction of top quark pairs

Since only a tiny fraction of the proton-proton collisions produced by the LHC will contain top quark pairs, an optimised event selection is applied to the collision events. Hereby $t\bar{t} \to b\bar{b}W^+W^- \to b\bar{b}q\bar{q}'\ell\nu_\ell$ events (with ℓ either an electron or a muon) are selected with a high efficiency, while other processes not containing top quark pairs need to be efficiently rejected. The different steps of the $t\bar{t}$ ℓ+jets event selection are presented in Section 4.1.

Each $t\bar{t}$ event contains typically at least four jets in the final state originating from the showering and hadronisation of the four quarks of the $t\bar{t} \to b\bar{b}q\bar{q}'\ell\nu_\ell$ decay. The estimation of the mass of the hadronically decaying top quark ($t \to bq\bar{q}'$) out of these jets is performed with a kinematic fit, which significantly improves the resolution on this estimated top-quark mass. This is discussed in Section 4.2, together with the flavour-specific jet energy scale corrections which are applied to the jets before they are given to the kinematic fit.

4.1 Selection of ℓ+jets top quark pairs

The most striking properties of a $t\bar{t}$ pair decaying in the ℓ+jets channel are the isolated lepton and the four high-p_T jets. These are used to efficiently identify such events. Section 4.1.1 discusses the first step, which is the requirement that the event passes the online event selection (*trigger*) and some additional filters in order to reject events with lots of electronic noise in the detector. Afterwards the event is required to contain exactly one well-identified isolated lepton, as explained in Section 4.1.2. Finally each event needs to contain at least 4 jets from which at least one is b-tagged, which is explained in Section 4.1.3. An overview of the event selection results is given in Section 4.1.4.

4.1.1 Trigger requirement and event cleaning

Data events are only recorded for offline analysis if they are selected by the online event selection system described in Section 2.2.4, i.e. if they pass the selection of one or more triggers. Therefore the first selection step is the trigger requirement, which is applied to both data and simulation. The triggers used for both the 7 and the 8 TeV datasets are

chosen based on two criteria: they need to select events with the desired topology in an efficient way and they need to be the trigger paths with the lowest p_T thresholds on the corresponding objects (leptons or jets).

The data used for the 7 TeV analysis was selected with a trigger requiring at least one isolated electron or muon and at least three jets. The trigger-level electrons were required to have $E_T > 25$ GeV and to be within the tracker acceptance ($|\eta| < 2.5$), while for the trigger-level muons a minimum p_T of 17 GeV was required and they needed to be within the region with the highest muon trigger efficiency ($|\eta| < 2.1$). The trigger-level jets were required to have $p_T > 30$ GeV and had to be within $|\eta| < 2.4$. For most of the 2011 data the jets were reconstructed at trigger level using only calorimeter information, but at the end of the 2011 LHC run a fraction of the data ($\sim 18\%$) was recorded with PF jets at trigger level.

The 8 TeV data was recorded with a simpler trigger, which required only the presence of at least one isolated electron or muon. The trigger-level electrons needed to pass the $E_T > 27$ GeV and $|\eta| < 2.5$ GeV requirements, while the trigger-level muons had to be within $|\eta| < 2.1$ and had to have $p_T > 24$ GeV.

Since non-physical backgrounds like electronic noise can mimic the signatures looked for by the event selection, an efficient rejection of these backgrounds is needed. The first requirement, which is applied to both data and simulation, asks that the main primary vertex (cf. Section 2.2.1) is located within a cylinder of radius 2 cm and length 48 cm centred around the nominal interaction point. This ensures that a proton-proton collision took place. Apart from that, the data events need to pass some additional requirements. They need to be recorded during a period of time when the detector was known to be turned completely on and all subdetectors were functioning in a proper way. Additional event filters were also applied, which were developed to reject events containing electronic noise while not rejecting good collision events.

4.1.2 Lepton selection criteria

Since the expected final state contains exactly one high-p_T isolated lepton originating from the W-boson decay, only events with exactly one such lepton are selected. The p_T and $|\eta|$ requirements of these leptons follow closely the requirements applied in the trigger. Apart from the kinematic requirements some additional lepton identification is performed to enhance the purity of the lepton selection.

Electron identification criteria

The starting point of the electron identification used in the 7 TeV analysis [107] is the collection the PF electrons, which were discussed in Section 3.3.3. Additionally, electron candidates need to be either in the barrel ($|\eta_{supercluster}| < 1.4442$) or the endcap ($|\eta_{supercluster}| > 1.5560$) part of the ECAL subdetector, excluding the barrel-endcap transition region. Their transverse and longitudinal impact parameters with respect to the main primary vertex also need to pass $|d_0| < 0.02$ cm and $|\Delta z| < 1$ cm, in order to ensure that it is a prompt electron produced during the hard scattering.

Afterwards they are required to pass some standard additional selection criteria. Here the same identification variables are used for electron candidates in the barrel and in the

endcap, but with different cut-values. An overview of these variables is given in Table 4.1. The *VBTF WP80* identification criteria are used in the identification of signal electrons

Working Point	*WP*80		*WP*95	
Variable	Barrel	Endcap	Barrel	Endcap
$\sigma_{i\eta i\eta}$	< 0.01	< 0.03	< 0.01	< 0.03
H/E	< 0.04	< 0.025	< 0.15	< 0.07
$\Delta\phi$	< 0.06	< 0.03	< 0.8	< 0.7
$\Delta\eta$	< 0.004	< 0.007	< 0.007	< 0.01

Table 4.1: Electron identification criteria used for the 7 TeV analysis.

while the *WP95* criteria are used for rejecting dilepton events as explained further below. Here $\sigma_{i\eta i\eta}$ is the width of the electron shower in the η-direction as calculated from the 5×5 array of crystals around the electron energy deposit in the ECAL, H/E is the hadronic leakage variable calculated by taking the ratio of the compatible HCAL energy deposit over the ECAL energy deposit of the electron, and the $\Delta\phi$ and $\Delta\eta$ variables provide a measure of the spatial matching between the track and the supercluster.

To reject electrons from *b*-hadron decay and charged hadrons falsely identified as electrons, an additional isolation requirement is applied to the electron candidates. The same definition of the relative isolation I_{rel} as in Section 3.3.3 is used,

$$I_{rel} = \frac{\sum_{CH} p_T^{CH} + \sum_{NH} E_T^{NH} + \sum_{\gamma} E_T^{\gamma}}{p_T^e}, \tag{4.1}$$

where the three sums run over all charged PF hadrons after charged hadron subtraction (CH), all neutral PF hadrons (NH) and all photons (γ) still present within a cone of radius $\Delta R = 0.4$ around the electron. Electrons are considered to be isolated if $I_{rel} < 0.1$.

Another important source of background which needs to be rejected are photons which are converting into an electron-positron pair when they traverse the tracker material. This conversion rejection is applied in a later stage of the event selection as explained below. An electron is rejected when the most inner expected hit of the reconstructed electron track is missing. Alternatively, if a conversion partner-track to the electron tracks is found, the electron is also rejected. A track is identified as a conversion partner-track if $|\Delta \cot\theta| < 0.02$ and $|Dist| < 0.02$ cm, where $\Delta \cot\theta$ is the difference in $\cot\theta$ between the two tracks and *Dist* is the distance between the two tracks measured in the transverse plane at the point where both tracks are parallel.

The electron identification used for the 8 TeV analysis is somewhat different than the one used in the 7 TeV case [109]. Here the starting point is the collection of electrons reconstructed with the method explained in Section 3.3.2. The PF electrons were not used in the 8 TeV analysis since there was a slight loss of identification efficiency for endcap electrons in the high-pileup environment of the 2012 data. Initially the electron candidates need to pass the same transverse impact parameter criterion ($|d_0| < 0.02$ cm) and they also need to be either in the barrel ($|\eta_{supercluster}| < 1.4442$) or the endcap ($|\eta_{supercluster}| > 1.5560$).

In contrast to the 7 TeV analysis where a cut-based electron identification was applied, the electron identification for the 8 TeV analysis uses a multivariate approach where several

observables are combined into a single variable *mvaId*. Four different categories of input variables are used: pure tracking variables like the χ^2 of the GSF track fit or the number of hits of a Kalman-filter track fit starting from the same seed, pure ECAL variables like the shower widths in the η and ϕ directions or the ECAL energy in a 1×5 array of crystals divided by the ECAL energy in a 5×5 array of crystals, geometrical matching variables like $\Delta\phi$ and $\Delta\eta$ as used in the 7 TeV analysis, and energy matching variables like H/E and the ratio of the supercluster energy to the track momentum. The resulting combined *mvaId* variable is plotted in Figure 4.1 for reconstructed electron candidates with or without a matched generator-level electron. A clear separation between both categories is visible and

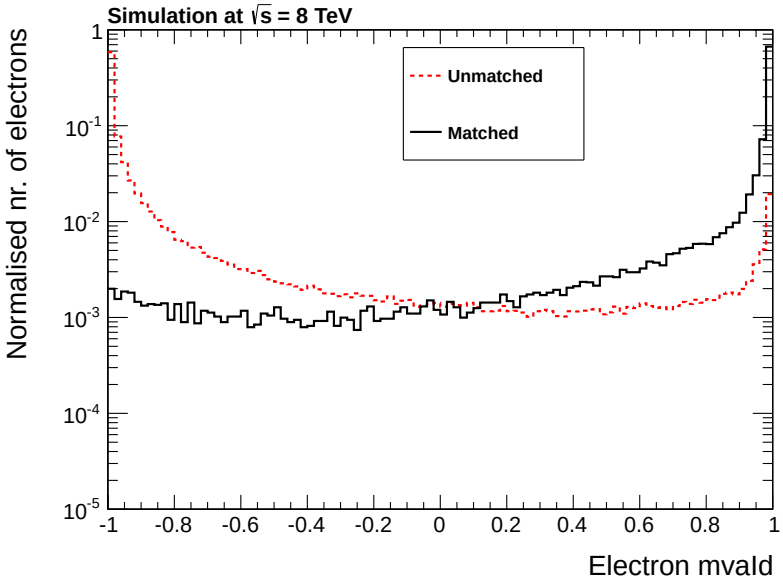

Figure 4.1: Electron *mvaId* variable for reconstructed electron candidates with or without a matched electron from W-boson decay at generator level within $\Delta R < 0.1$.

used to identify electrons via $mvaId > 0.5$.

Apart from the multivariate identification each electron also needs to pass the conversion veto, which is somewhat different from the 7 TeV conversion veto. An electron is again rejected when the most inner expected hit of the reconstructed electron track is missing. Alternatively, an explicit conversion vertex fit is performed based on the reconstructed charged particle tracks and the electron is also rejected if such a conversion vertex is found. These conversion vertices need to pass some requirements on their fit probability ($P_{fit} > 10^{-6}$), their transverse decay length ($l_{xy} > 2$ cm) and their number of hits of the corresponding track before the vertex position ($nHitsBeforeVtx = 0$).

Since the relative lepton isolation as defined in Equation 4.1 degrades significantly in the high pileup events of the 2012 data, the 8 TeV analysis uses a modified relative isolation variable which is more robust with respect to pileup. The idea is similar as what is done in Equation 3.13, where the expected contribution of pileup to the relative isolation is subtracted. The *effective area* corrected relative isolation is defined as

$$I_{rel}^{EA} = \frac{\sum_{CH} p_T^{CH} + \max\left(0, \sum_{NH} p_T^{NH} + \sum_{\gamma} p_T^{\gamma} - \rho \cdot A_{eff}\right)}{p_T^{\ell}}, \tag{4.2}$$

where the sums are defined as before but with a cone size of $\Delta R = 0.3$, and ρ is the p_T density of the event calculated in the same way as explained in Section 3.3.4 for the jet energy corrections. The effective areas A_{eff} in this equation are estimated from data in bins of $|\eta|$. They are defined as the areas that make the signal efficiency flat versus the number of pileup interactions. The relative isolation distribution for reconstructed electron candidates with or without a matched generator-level electron is shown in Figure 4.2, where a clear separation between both categories is visible. Electrons are isolated if they pass the

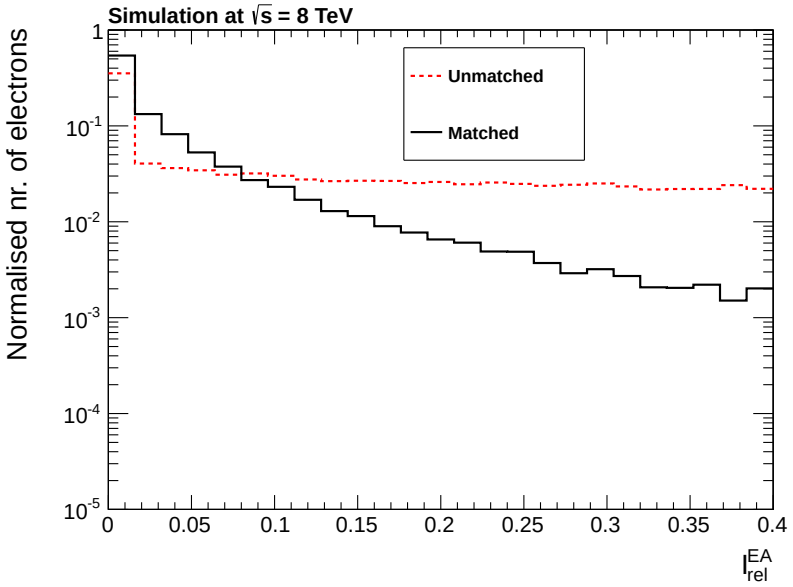

Figure 4.2: Electron relative isolation as defined in Equation 4.2 for reconstructed electron candidates with or without a matched electron from W-boson decay at generator level within $\Delta R < 0.1$.

$I_{rel}^{EA} < 0.1$ requirement.

Muon identification criteria

The identification of muons in both the 7 and the 8 TeV analyses starts from the collection of PF muons, as discussed in Section 3.3.3. On top of these, some additional identification and isolation requirements are applied which are slightly different between 7 and 8 TeV.

The muons from the 7 TeV dataset [105] need to be reconstructed as global muons and they are required to originate from the main primary vertex via their transverse and longitudinal impact parameters ($|d_0| < 0.02$ cm and $|\Delta z| < 1$ cm). Afterwards they need to pass the standard muon identification criteria on the normalised χ^2 of the global muon fit ($\chi^2/ndf < 10$), the number of hits in the muon system used for the global muon fit ($n_{MuonHits} > 0$), the number of hits in the inner tracker ($n_{TrackerHits} > 10$), the number of hits in the pixel system ($n_{PixelHits} > 0$), and the number of muon segments matched to the global muon ($n_{MatchedSegments} > 1$). Finally the muons are required to be isolated by asking $I_{rel} < 0.125$. The same definition of the relative isolation as for the electrons is used (Equation 4.1), also with a cone size of $\Delta R = 0.4$.

The identification of muons for the 8 TeV analysis is done in a similar way [125]. It starts by requiring that they originate from the main primary vertex ($|d_0| < 0.02$ cm and $|\Delta z| < 0.5$ cm) and afterwards similar muon identification criteria as in the 7 TeV case are applied. Exactly the same criteria on the χ^2/ndf, $n_{MuonHits}$, $n_{PixelHits}$ and $n_{MatchedSegments}$ variables are used while the criterion on the number of inner tracker hits is replaced by a requirement on the number of layers of the tracker system with a valid hit ($n_{TrackerLayers} > 5$). Finally an isolation requirement is imposed which is corrected for the presence of pileup. The technique used to remove the pileup dependence is slightly different with respect to what is done for the electrons, but it has a similar effect. The same $\Delta\beta$-corrected relative isolation as used in Section 3.3.3 (Equation 3.13) is used here

$$I_{rel}^{\Delta\beta} = \frac{\sum_{CH} p_T^{CH} + \max\left(0, \sum_{NH} p_T^{NH} + \sum_{\gamma} p_T^{\gamma} - 0.5 \sum_{PU} p_T^{PU}\right)}{p_T^{\ell}}. \tag{4.3}$$

A comparison of this isolation variable for PF muons with or without a matched generator-level muon is shown in Figure 4.3. Here the separation is less clear than for the electrons,

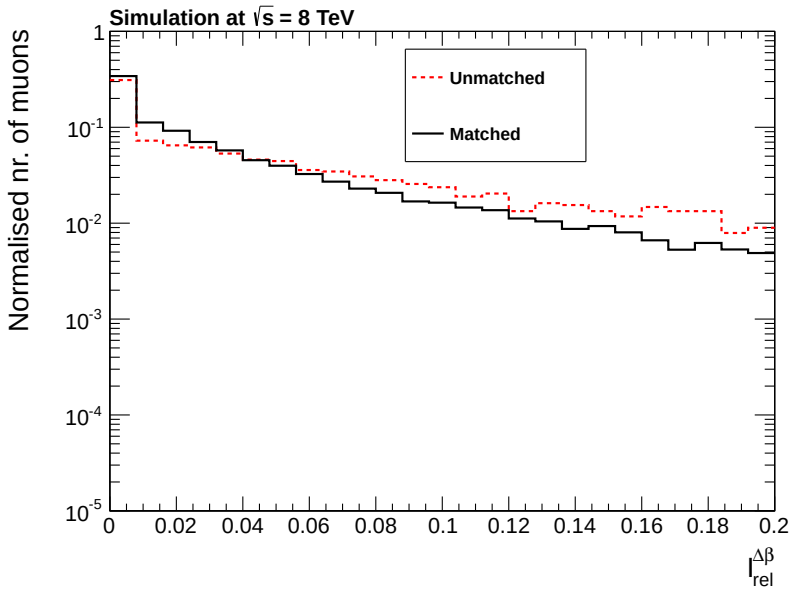

Figure 4.3: Muon relative isolation as defined in Equation 4.3 for PF muons with or without a matched generator-level muon from W-boson decay.

since this plot is made after the PF muon identification, which rejects most of the badly reconstructed muons, and by the requirement on the relative isolation applied during the top projections (Section 3.3.3). Muons are isolated if they pass the $I_{rel}^{\Delta\beta} < 0.12$ requirement.

Overview of the lepton selection

After the initial event selection requirements from Section 4.1.1, the lepton selection is applied to candidate events in the electron and muon channels. In both channels and for both centre-of-mass energies the selection requires exactly one isolated lepton passing the

identification criteria as explained before, while a veto on other more loosely identified isolated leptons is applied to reject dilepton $t\bar{t}$ and $Z/\gamma^* \to \ell^+\ell^-$ events.

Events in the 7 TeV electron channel need to contain exactly one isolated electron of $E_T > 30$ GeV within $|\eta| < 2.5$ surviving the *WP*80 identification criteria. They are rejected if a global PF muon passing $p_T > 10$ GeV, $|\eta| < 2.5$ and $I_{rel} < 0.2$ is found; or if an additional PF electron with $E_T > 20$ GeV, $|\eta| < 2.5$ and $I_{rel} < 0.2$ passing the *WP*95 identification criteria is found which gives a combined invariant mass with the signal electron within $76 < m_{ee} < 106$ GeV. Finally the signal electron is required to pass the photon conversion veto as explained before.

In the 7 TeV muon channel exactly one isolated muon with $p_T > 20$ GeV within $|\eta| < 2.1$ is required. Events containing additional global PF muons with $p_T > 10$ GeV, $|\eta| < 2.5$ and $I_{rel} < 0.2$ or containing PF electrons with $E_T > 15$ GeV, $|\eta| < 2.5$ and $I_{rel} < 0.2$ are rejected.

The events in the electron channel at 8 TeV are required to contain exactly one isolated and well-identified electron with $p_T > 32$ GeV and $|\eta| < 2.5$ which also passes the photon conversion veto. Similarly, the muon channel events used in the 8 TeV analysis need to contain exactly one isolated and well-identified muon with $p_T > 25$ GeV and $|\eta| < 2.1$. If any of the events in both decay channels contains an additional electron with $p_T > 20$ GeV, $|\eta| < 2.5$, $I_{rel}^{EA} < 0.15$ and $mvaId > 0$ or an additional PF muon with $p_T > 10$ GeV, $|\eta| < 2.5$ and $I_{rel}^{\Delta\beta} < 0.2$ which is a tracker muon or a global muon then events are rejected.

The trigger, identification and isolation efficiencies of electrons and muons in the 8 TeV data have been measured centrally within the CMS collaboration [125, 126]. Although the data-to-simulation ratios of these efficiencies are very close to one they are still used to rescale the efficiency in simulation so that it better agrees with the observed efficiency in data.

Apart from these lepton selection requirements the leptons are also used in a ΔR-based cleaning between the leptons and the jets to avoid the double-counting of some PF particles within the analysis. This is further explained in the next section.

4.1.3 Jet selection criteria

After passing the lepton selection criteria the jet selection is applied to the events, which is the same for the 7 and 8 TeV analyses. Events need to contain at least four L1L2L3 corrected PF jets, since the ℓ+jets $t\bar{t}$ decay has four quarks in the final state. These jets need to have $p_T > 30$ GeV and $|\eta| < 2.4$, and they need to be well separated from the lepton selected in the previous step ($\Delta R > 0.3$). They also need to pass the jet identification criteria which are designed to reject jets originating from electronic noise in the calorimeters. Therefore a selection is applied on the composition of and the energy sharing between the PF particles which make up the jet. This selection is applied on several variables: the jet energy fraction carried by neutral PF hadrons ($f_{NH} < 0.99$), the jet energy fraction carried by neutral electromagnetic PF particles ($f_{NE} < 0.99$), the jet energy fraction carried by charged PF hadrons ($f_{CH} > 0$), the jet energy fraction carried by charged electromagnetic PF particles ($f_{CE} < 0.99$), the number of constituents (PF particles) of the jet ($n_{PFparticles} > 1$), and the number of charged particles of the jet ($n_{charged} > 0$).

Since most of the non-$t\bar{t}$ backgrounds like W+jets, Z/γ^*+jets and QCD multijet are

not expected to contain a high number of b quarks, an additional b-tag requirement is applied to both the 7 and the 8 TeV datasets. In both cases one of the selected jets needs to be b-tagged using the standard requirement as explained in Section 3.3.5 (medium working point of the CSV algorithm).

4.1.4 Results after the full event selection

The entire event selection consisting of the trigger, event cleaning, lepton criteria and jet criteria is applied to all the simulated events and to the real data events collected with the relevant triggers. So before the actual analysis is carried out a comparison between data and simulation is performed to check how well the data is described by the simulation. This comparison is done after an additional event selection requirement based on the kinematic fit as explained below in Section 4.2.2 ($\chi^2/ndf < 10$).

As explained in Section 3.2 the effect of pileup is included in the simulation. The number of pileup collisions included in the simulation does not perfectly match with the amount of pileup present in the data, since the amount of pileup in the data is only known after the data has been taken. Therefore the simulated events are reweighted to match the pileup profile observed in data. The amount of pileup in data is estimated via the total number of pp collisions N_{pp}, which is obtained by using the measured instantaneous luminosity $\mathcal{L}$ of each bunch crossing and the total inelastic proton-proton collision cross-section σ_{pp} via

$$N_{pp} = \sigma_{pp} \int \mathcal{L} \, . \tag{4.4}$$

The performance of this reweighting can be checked afterwards by comparing the number of primary vertices between data and simulation.

Comparison between data and simulation at 7 TeV

During the 2011 data-taking period a total of 4.97 ± 0.11 fb^{-1} of data for offline analysis was recorded in both the e+jets and μ+jets channels. The effect of the pileup reweighting can clearly be seen from Figure 4.4, where the primary vertex multiplicity is shown with and without the reweighting after the entire event selection. The reweighting clearly improves the agreement, although it is still not completely perfect. A further discussion of the effect of pileup on the measurements can be found in the systematic uncertainties of the respective analysis chapters (Sections 5.3.2 and 6.3). The pileup reweighting shown here will be applied to all subsequent 7 TeV results.

An overview of the number of events after the full event selection for the different simulated event samples and as observed in data is shown in Table 4.2. In general a reasonable agreement is found, although the total number of events selected from simulation is always 2 to 4% higher when compared to the number of selected data events. Agreement between data and simulation in the overall event yield is less important for the analyses presented in this thesis than agreement in their kinematic distributions. Therefore all comparisons between data and simulation have the simulated signal and background events rescaled to match the number of events observed in data. This was also done for the plots shown in Figure 4.4.

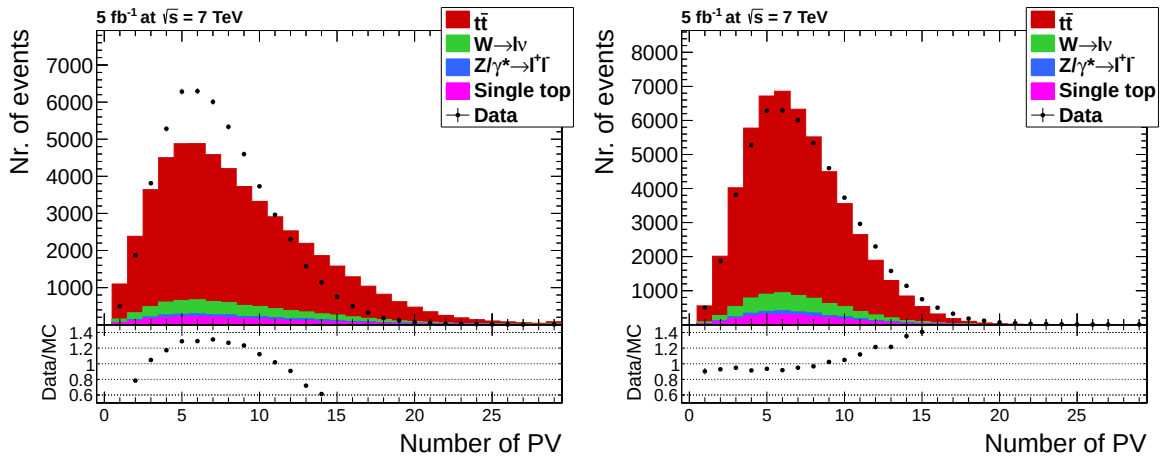

Figure 4.4: Comparison between data and simulation of the number of primary vertices before (left) and after (right) pileup reweighting.

Sample	e^++jets	e^-+jets	μ^++jets	μ^-+jets
$t\bar{t}$	10554 ± 12	10534 ± 12	13380 ± 14	13262 ± 14
W+jets	1116 ± 10	816 ± 8	1398 ± 11	979 ± 9
Z/γ^*+jets	249 ± 11	234 ± 10	195 ± 9	201 ± 9
Single top	560 ± 5	511 ± 5	714 ± 6	638 ± 6
Total	12479 ± 20	12095 ± 18	15687 ± 21	15080 ± 20
Observed	12169	11609	15258	14715

Table 4.2: Number of events passing the full selection in data and the expectation from simulation for the 7 TeV analysis. The uncertainties on the event counts are statistical, reflecting the limited size of the different samples.

The resulting agreement between data and simulation for the jet kinematics and the number of jets is shown in Figure 4.5, where a decent agreement is observed. These

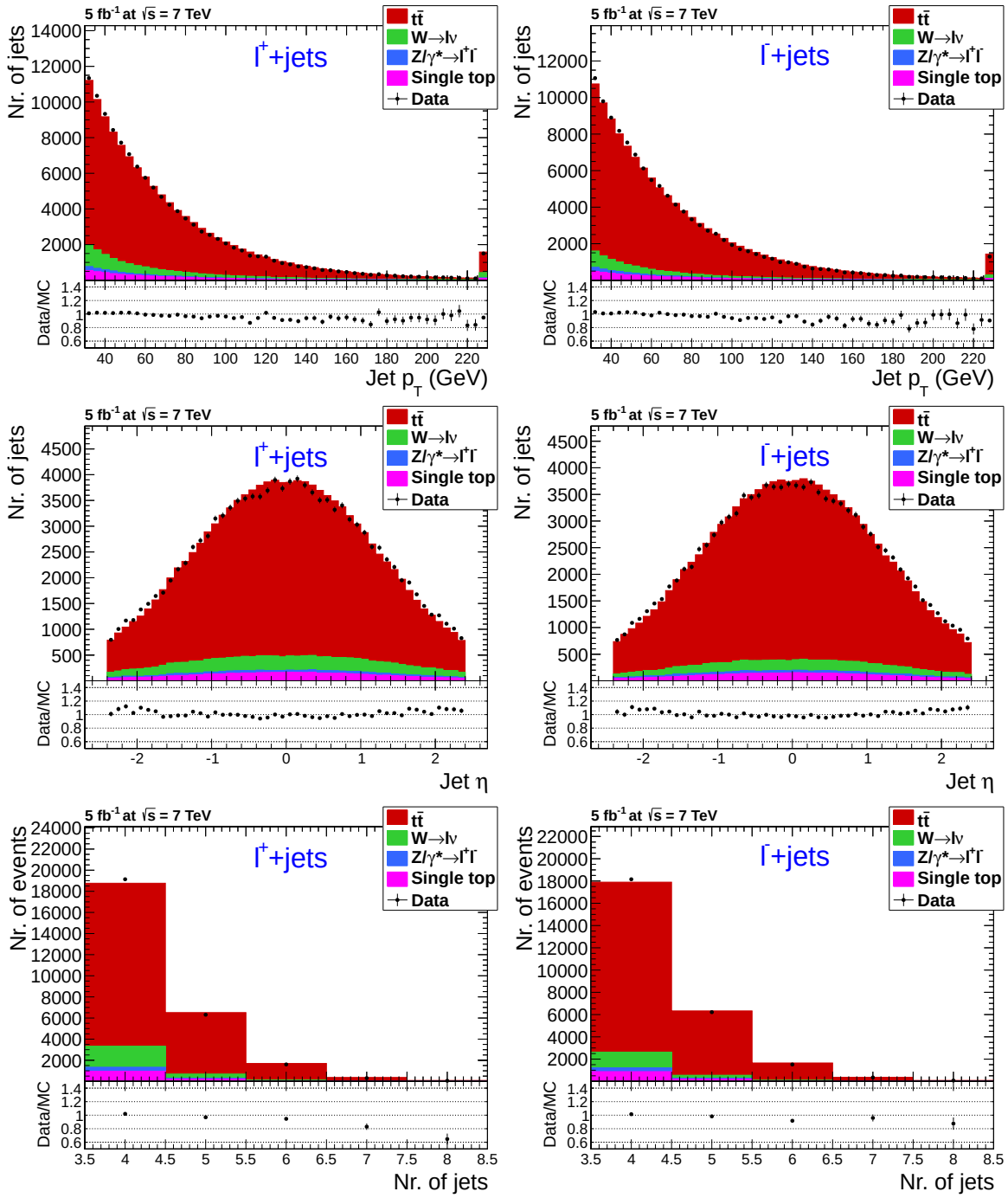

Figure 4.5: Comparison between data and simulation of the p_T (top) and η (middle) of the four leading jets, and the number of select jets (bottom); shown separately for ℓ^++jets (left) and ℓ^-+jets (right) events.

comparisons are done separately for ℓ^++jets and ℓ^-+jets events since both categories of events are used differently within the measurement of the mass difference between the

top and the antitop quark. Other relevant variables are the number of b-tagged jets and the lepton charge. Their data to simulation comparison is shown in Figure 4.6, where a reasonable agreement is also observed.

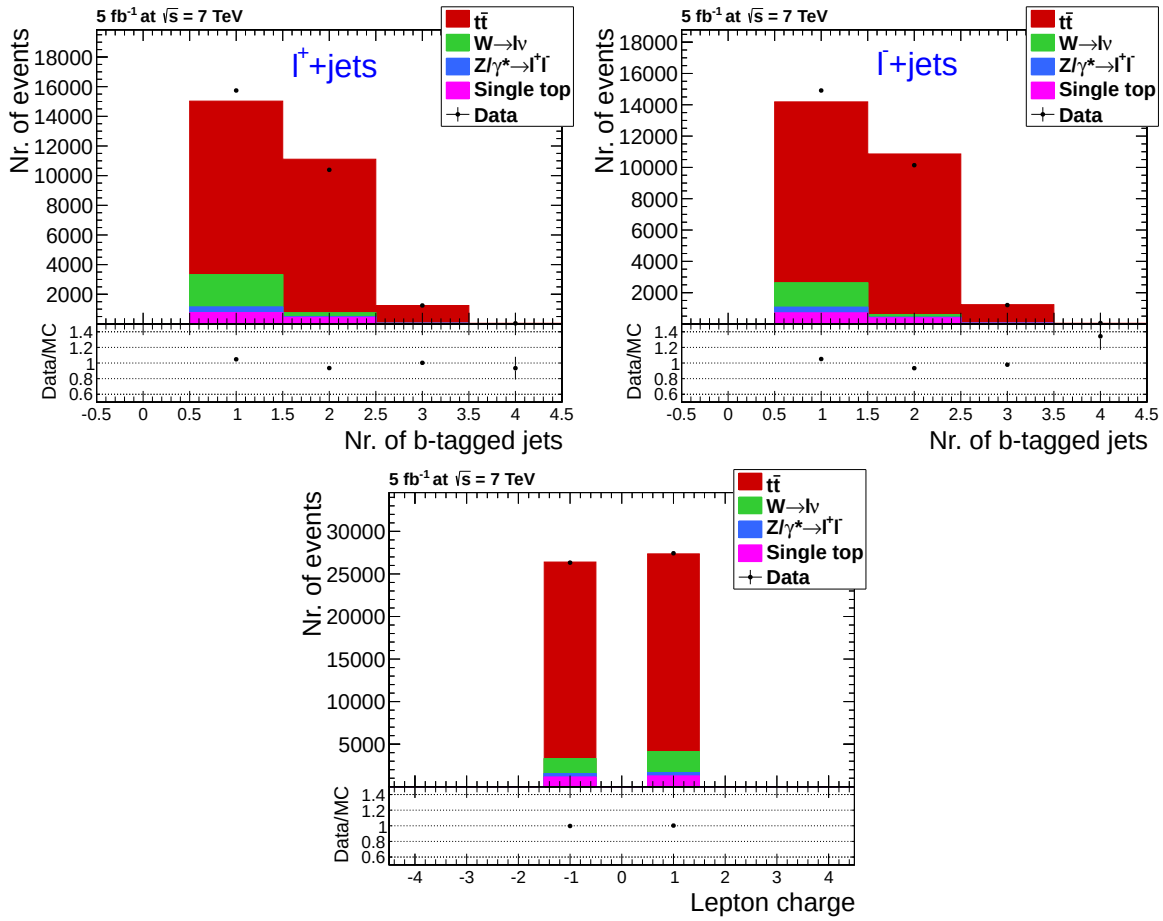

Figure 4.6: Comparison between data and simulation of the b-tag multiplicity separately for ℓ^++jets (top left) and ℓ^-+jets (top right) events, and of the lepton charge (bottom) .

Comparison between data and simulation at 8 TeV

The triggers based on a single isolated lepton, which were discussed before, recorded a total of 19.7 ± 0.9 fb^{-1} of data in both the electron and the muon channel during the 2012 data-taking period. Pileup reweighting was also applied and it has a similar effect on the vertex multiplicity as in 7 TeV, as shown in Figure 4.7. The agreement between data and simulation is again clearly improved by the reweighting.

For the 8 TeV analysis QCD multijet events are included as another background source in the analysis, mainly because the low $\not{E}_T$ region was not properly described by the simulation without including such events. Since the proper simulation of sufficient QCD multijet background events is very difficult due to the very high cross-section, the QCD multijet background is obtained in a data-driven way. One of the standard techniques to do this is by selecting a data sample which is highly enriched in QCD multijet events by inverting the lepton isolation criteria during the event selection.

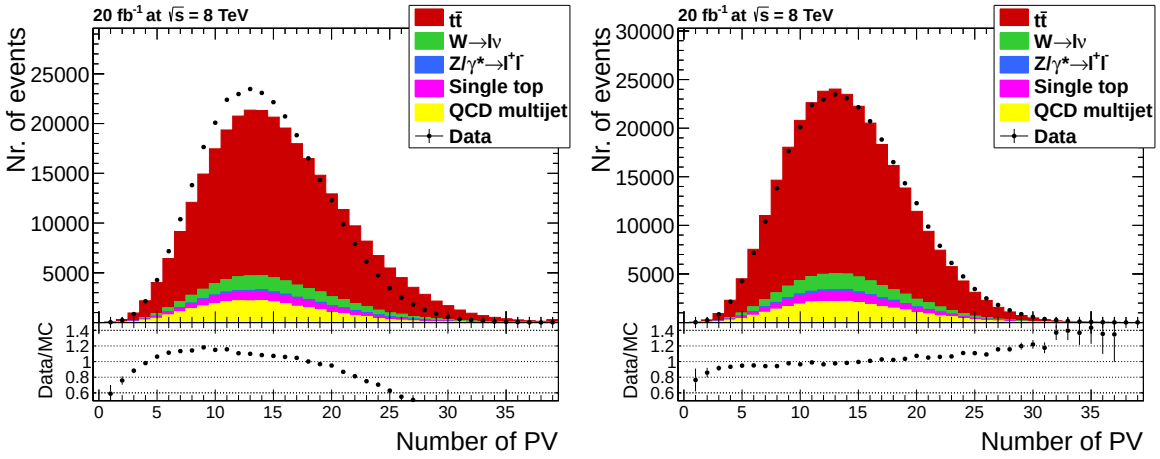

Figure 4.7: Comparison between data and simulation of the number of primary vertices before (left) and after (right) pileup reweighting.

The QCD-enriched e+jets sample was obtained by applying exactly the same selection criteria to real data events but with inverted identification and isolation criteria applied to the signal electron ($mvaId < 0.5$ and $I_{rel}^{EA} > 0.2$). For the QCD-enriched μ+jets sample a similar procedure was used but only the isolation criterion of the signal muon was inverted ($I_{rel}^{\Delta\beta} > 0.2$). The normalisation of these data-driven QCD multijet samples is done separately in the e+jets and μ+jets samples with a binned maximum-likelihood fit to the $\not\!\!E_T$ distribution observed in data in order to estimate the fraction of QCD multijet events. The $\not\!\!E_T$ distributions of the data-driven QCD multijet samples and of the combination of all other samples is given as input to the fit, while the fractions of QCD multijet events is left free and determined by the fit. The obtained results are plotted in Figure 4.8. The esti-

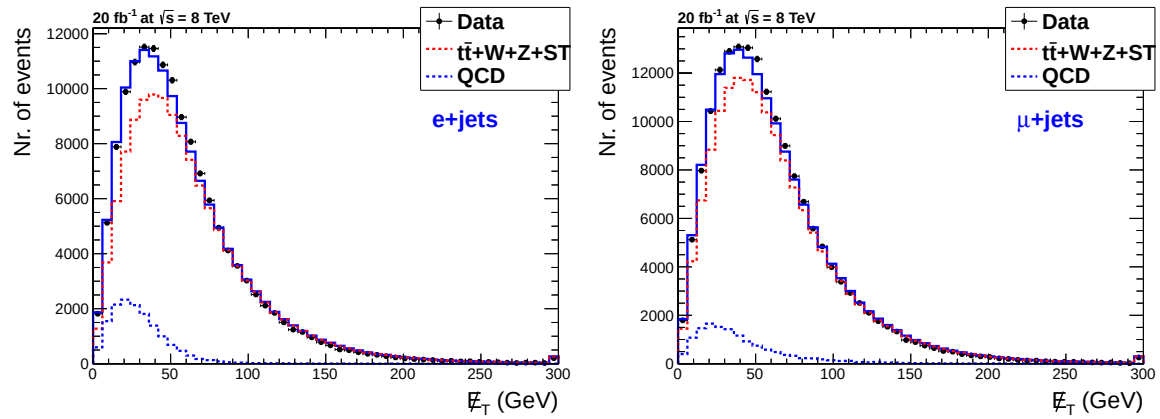

Figure 4.8: Results of the binned maximum-likelihood fit to estimate the fraction of QCD multijet events in the e+jets (left) and the μ+jets (right) channels.

mated fractions of QCD multijet events in the e+jets and μ+jets samples are, respectively, 10.6% and 8.0%, and are used to normalise the QCD multijet background in all subsequent 8 TeV results.

The number of events after the entire event selection is shown in Table 4.3 for the different decay channels. Again a reasonable agreement between data and simulation is

Sample	e^++jets	e^-+jets	μ^++jets	μ^-+jets
$t\bar{t}$	56185 $\pm$ 68	55737 $\pm$ 68	72358 $\pm$ 77	72433 $\pm$ 77
W+jets	5474 $\pm$ 50	4147 $\pm$ 45	7180 $\pm$ 59	5198 $\pm$ 51
Z/γ^*+jets	839 $\pm$ 12	804 $\pm$ 11	816 $\pm$ 12	820 $\pm$ 11
Single top	3122 $\pm$ 35	2671 $\pm$ 34	3926 $\pm$ 40	3428 $\pm$ 38
QCD multijet	7959 $\pm$ 89	7269 $\pm$ 85	7182 $\pm$ 85	7207 $\pm$ 85
Total	73579 $\pm$ 128	70628 $\pm$ 123	91462 $\pm$ 136	89086 $\pm$ 132
Observed	71952	70396	87039	84024

Table 4.3: Number of events passing the full event selection in data and the expectation from simulation for the 8 TeV analysis. The QCD multijet background is obtained in a data-driven way as explained in the text. The uncertainties on the event counts are statistical, reflecting the limited size of the different samples.

observed, since the differences in the number of selected events are below 2% in the e+jets channels and within 5 to 6% in the μ+jets channel. As in the 7 TeV analysis the agreement in the overall event yield is less important, hence the signal and background events are also rescaled to match the number of observed events in data to ease the comparison of the shapes between data and simulation.

The comparison between data and simulation for the jet kinematics and the jet multiplicity can be found in Figure 4.9 and the comparison for the b-tag multiplicity and the lepton charge is shown in Figure 4.10. In general a reasonable agreement between data and simulation is observed for all the observables studied here.

4.2 Top-quark mass reconstruction

All the events passing the event selection discussed previously can be used to estimate the top-quark mass. In this thesis only the mass of the hadronically decaying top quark will be estimated since all kinematics of this system are measured, in contrast to the leptonically decaying top quark where the non-interacting neutrino results in partly unmeasured kinematics. The four leading jets in each event are used for this purpose, which can be assigned to the four quarks from the $t\bar{t} \to b\bar{b}q\bar{q}'\ell\nu_\ell$ decay in 24 different ways (24 *jet combinations* or *jet permutations*). Since the interchange of the jets assigned to the two light quarks q and $\bar{q}'$ from W-boson decay will result in exactly the same reconstructed top-quark and W-boson masses, only 12 relevant jet combinations need to be considered.

Instead of choosing the most optimal jet combination per event, which is never 100% efficient, multiple combinations per event will be used as input to the Ideogram method in order to estimate the top-quark mass as explained in Chapter 5. Therefore the mass of the hadronically decaying top-quark is estimated for each of the 12 jet combinations. The estimation itself is performed via a kinematic fit with mass constraints as described in Section 4.2.2. Prior to this estimation, additional jet energy corrections are applied to the jets, depending on the flavour assigned to them in each particular jet combination as discussed in Section 4.2.1. The resolution on the jet kinematics is also used as input to the

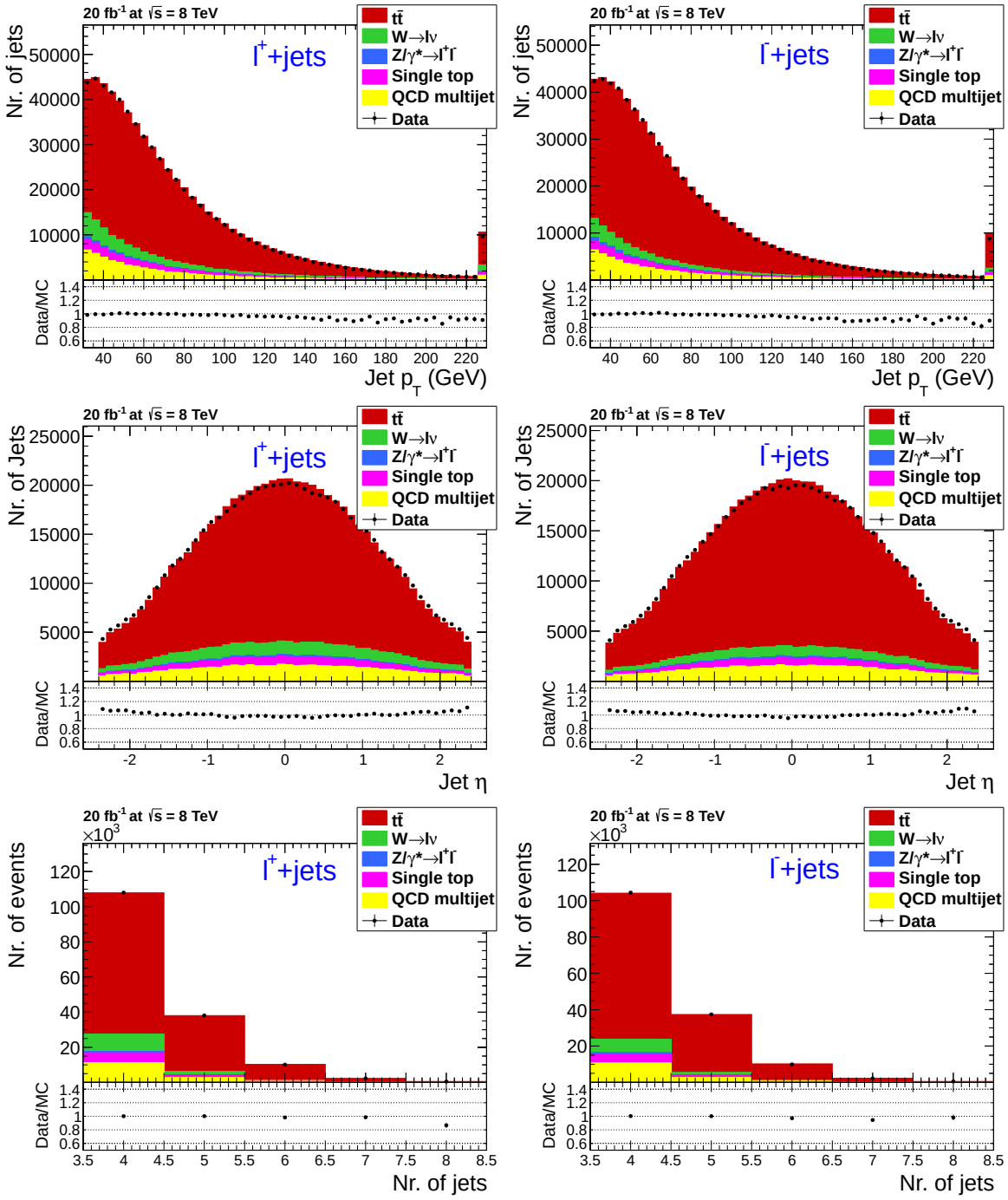

Figure 4.9: Comparison between data and simulation of the p_T (top) and η (middle) of the four leading jets, and the number of select jets (bottom); shown separately for ℓ^++jets (left) and ℓ^-+jets (right) events.

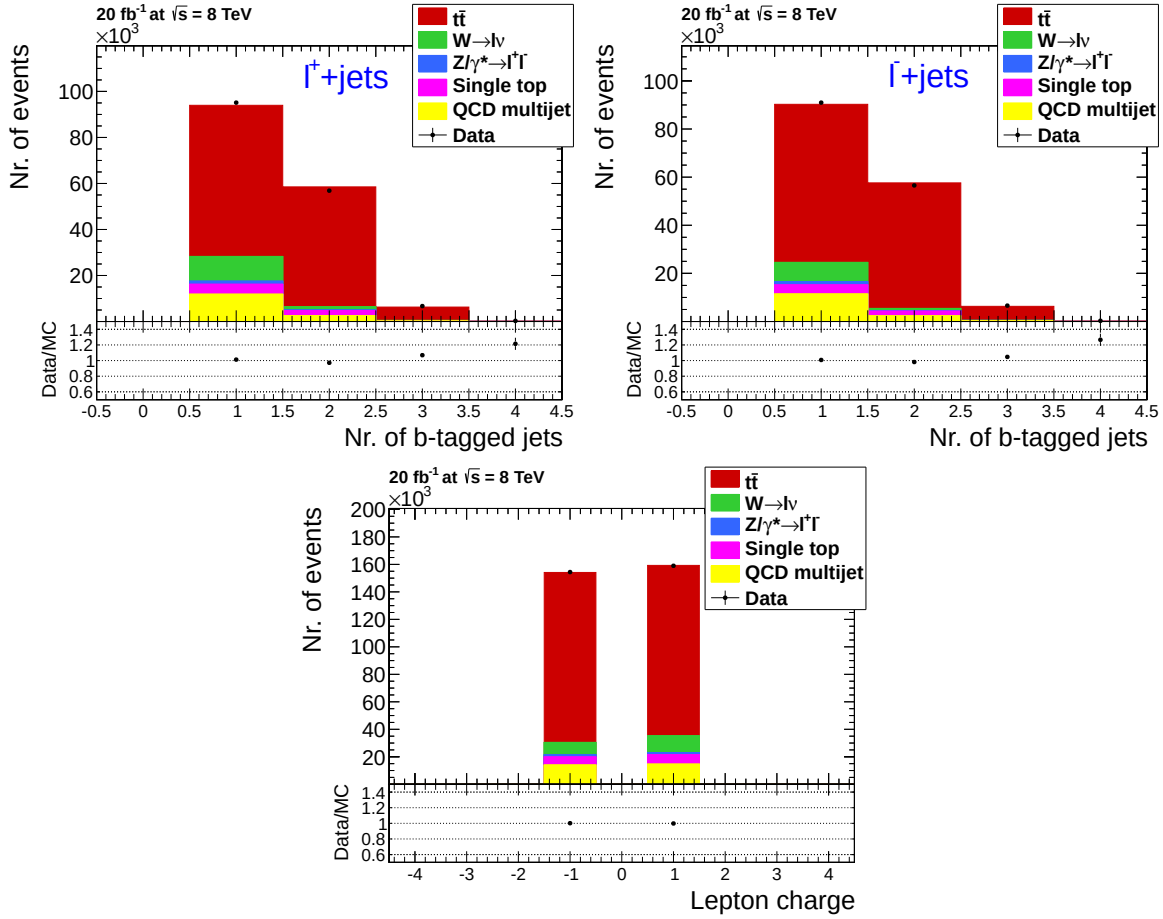

Figure 4.10: Comparison between data and simulation of the *b*-tag multiplicity separately for ℓ^++jets (top left) and ℓ^-+jets (top right) events, and of the lepton charge (bottom).

kinematic fit and their derivation is also discussed in this section. Finally, in Section 4.2.3 the results of the kinematic fit will be presented together with a comparison between data and simulation of these results.

4.2.1 Parton-level jet energy corrections and jet resolutions

Since the standard jet energy corrections used within CMS and discussed in Section 3.3.4 only provide flavour-inclusive corrections up to the level of generator jets, additional corrections estimated from simulated $t\bar{t}$ events in the ℓ+jets decay channel are applied to the jets before the top-quark mass estimation. These L5L7 corrections are estimated separately for the b jets from top-quark decays and for the light jets from W-boson decays by comparing the transverse energies E_T of the PF jets and the corresponding quarks. These quarks are matched to the jets using a ΔR matching requirement of $\Delta R < 0.3$.

In bins of p_T^{PFjet} and $|\eta^{PFjet}|$ the distribution of the relative E_T difference between the PF jet and the matched quark $\frac{E_T^{PFjet} - E_T^{quark}}{E_T^{PFjet}}$ is fitted with a Gaussian function. The resulting mean μ is used to obtain the parton-level jet energy correction in this (p_T^{PFjet}, $|\eta^{PFjet}|$) bin. All the results for each $|\eta^{PFjet}|$ bin are then fitted with suitable functions to obtain the jet energy corrections as a function of p_T^{PFjet} as shown in Figures 4.11 and 4.12 for 7 and 8 TeV, respectively. These plots show that light-quark jets are slightly over-corrected except at

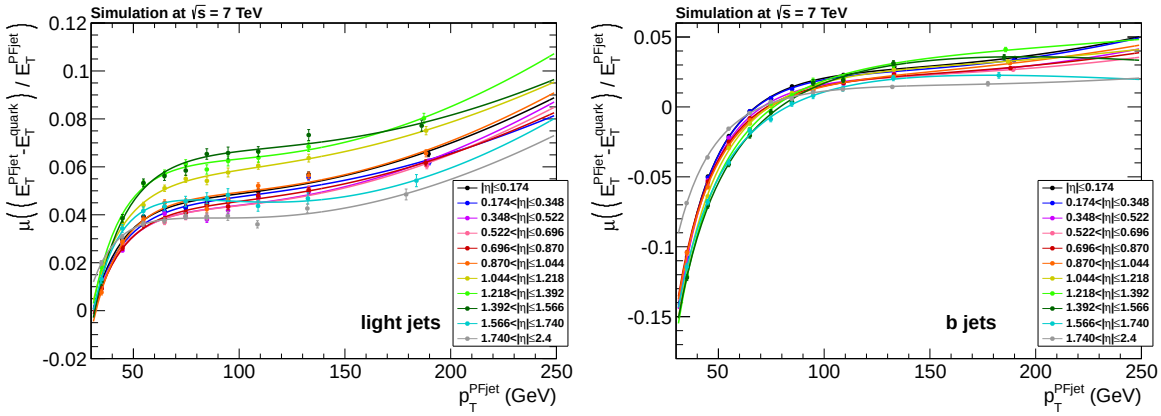

Figure 4.11: Fitted parton-level jet energy correction functions for light jets (left) and b jets (right) as estimated with 7 TeV $t\bar{t}$ ℓ+jets events.

very low p_T, while this effect is significantly smaller for b-quark jets. At very low p_T these b-quark jets are significantly under-corrected by the standard jet energy scale corrections. This is caused by the production of neutrinos during the b-hadron decay within these jets, which is not taken into account by the standard corrections.

The effect of these additional jet energy corrections on the reconstructed top-quark mass is shown in Figure 4.13, where the invariant mass of the three jets matched to the three quarks from the $t \to bq\bar{q}'$ decay is plotted before and after these combined L5L7 corrections. Both distributions are fitted with a Gaussian function and a comparison of the mean μ of both fits clearly shows that the top-quark mass moves closer to the generated top-quark mass of 172.5 GeV after the additional corrections are applied. The relative

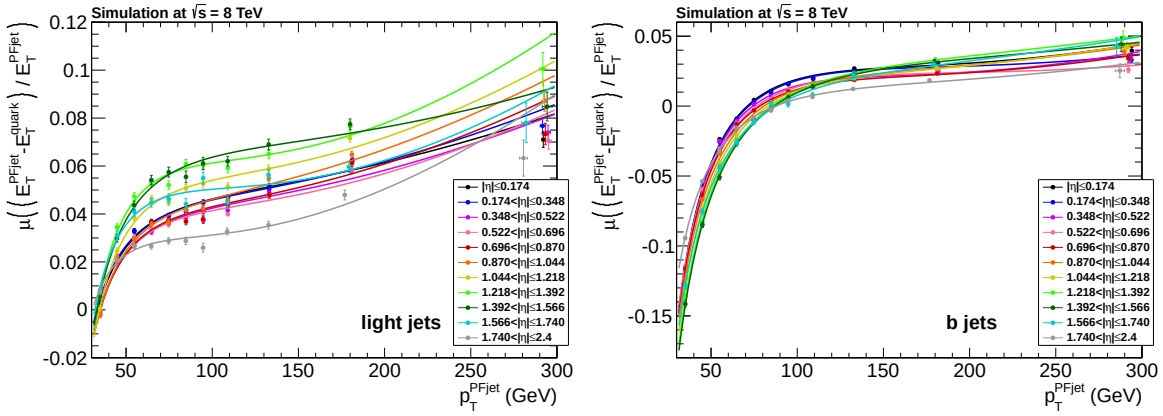

Figure 4.12: Fitted parton-level jet energy correction functions for light jets (left) and b jets (right) as estimated with 8 TeV $t\bar{t}$ ℓ+jets events.

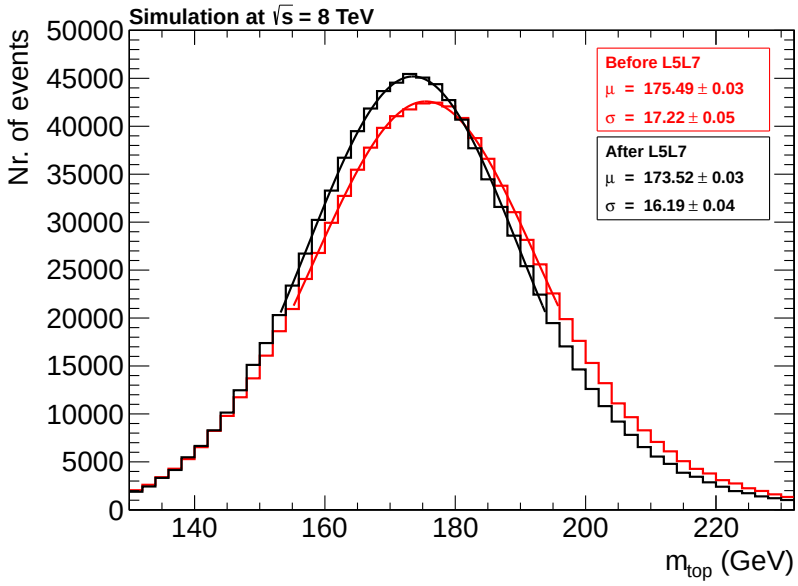

Figure 4.13: Top-quark mass as estimated by taking the invariant mass of the three jets matched to the three quarks from top-quark decay before and after the L5L7 jet energy corrections.

resolution on the top-quark mass which is calculated as σ/μ is also slightly improved from 9.81% to 9.33%.

The resolutions on the reconstructed E_T, θ and ϕ of the PF jets are an important input of the kinematic fit. Their calculation happens in a similar way as the L5L7 jet energy corrections. They are also estimated separately for b jets from top-quark decays and light jets from W-boson decays after applying the L5L7 jet energy corrections. In the same bins of p_T^{PFjet} and $|\eta^{PFjet}|$ the distributions of the absolute difference in E_T, θ and ϕ between the PF jets and the quarks are fitted with a Gaussian function and its width σ is taken as the estimated resolution. In each $|\eta^{PFjet}|$-bin these estimated resolutions are again fitted with suitable functions as shown in Figures 4.14 and 4.15. The E_T resolution of the PF jets shows a small dependency on p_T^{PFjet}, which results in an improved relative E_T resolution for increasing p_T^{PFjet}. This is caused by the fact that high-p_T jets tend to consist of more energetic particles and these are measured more precisely by the CMS detector. The improvement of the θ and ϕ resolutions for increasing p_T^{PFjet} is also caused by this effect. The improvement of the resolution for PF jets with higher $|\eta|$ is also a consequence of this effect since a jet with a certain p_T in the forward region is more energetic than a jet of the same p_T in the central region of the detector.

4.2.2 Kinematic fit

For each of the twelve possible jet combinations in an event the mass of the hadronically decaying top-quark candidate is estimated with a kinematic fit. In this procedure the measured jet kinematics, their corresponding resolutions and the kinematic constraints of the system are used in a combined way. The measured kinematics are slightly varied within their resolutions until they fulfil the imposed constraints, resulting in a significantly improved resolution on the reconstructed top-quark. The kinematic fit algorithm is described in great detail in Ref. [127]. Only the concepts relevant within this thesis will be discussed below.

Fitting event topologies with kinematic constraints

In general one typically has a system with n measured parameters $\vec{y}$ and m constraints $\vec{f}$. These constraints will be fulfilled for the true parameters $\bar{y}$ and are defined as

$$\begin{aligned} f_1(\bar{y}_1, \cdots, \bar{y}_n) &= 0 \\ &\vdots \\ f_m(\bar{y}_1, \cdots, \bar{y}_n) &= 0. \end{aligned} \tag{4.5}$$

The measured values $\vec{y}$ will not exactly fulfil these constraints so corrections $\Delta\vec{y}$ to these measured parameters need to be calculated. The corrected values $\vec{y}' = \vec{y} + \Delta\vec{y}$ will then fulfil these constraints. At the same time the weighted sum

$$S(\vec{y}) = \Delta\vec{y}^T \mathcal{V}^{-1} \Delta\vec{y} \tag{4.6}$$

should also be minimal, where $\mathcal{V}$ is the covariance matrix of the measured parameters.

The minimisation of this weighted sum taking the constraints of Equation 4.5 into account is performed by using *Lagrange Multipliers*. Therefore a new likelihood is defined

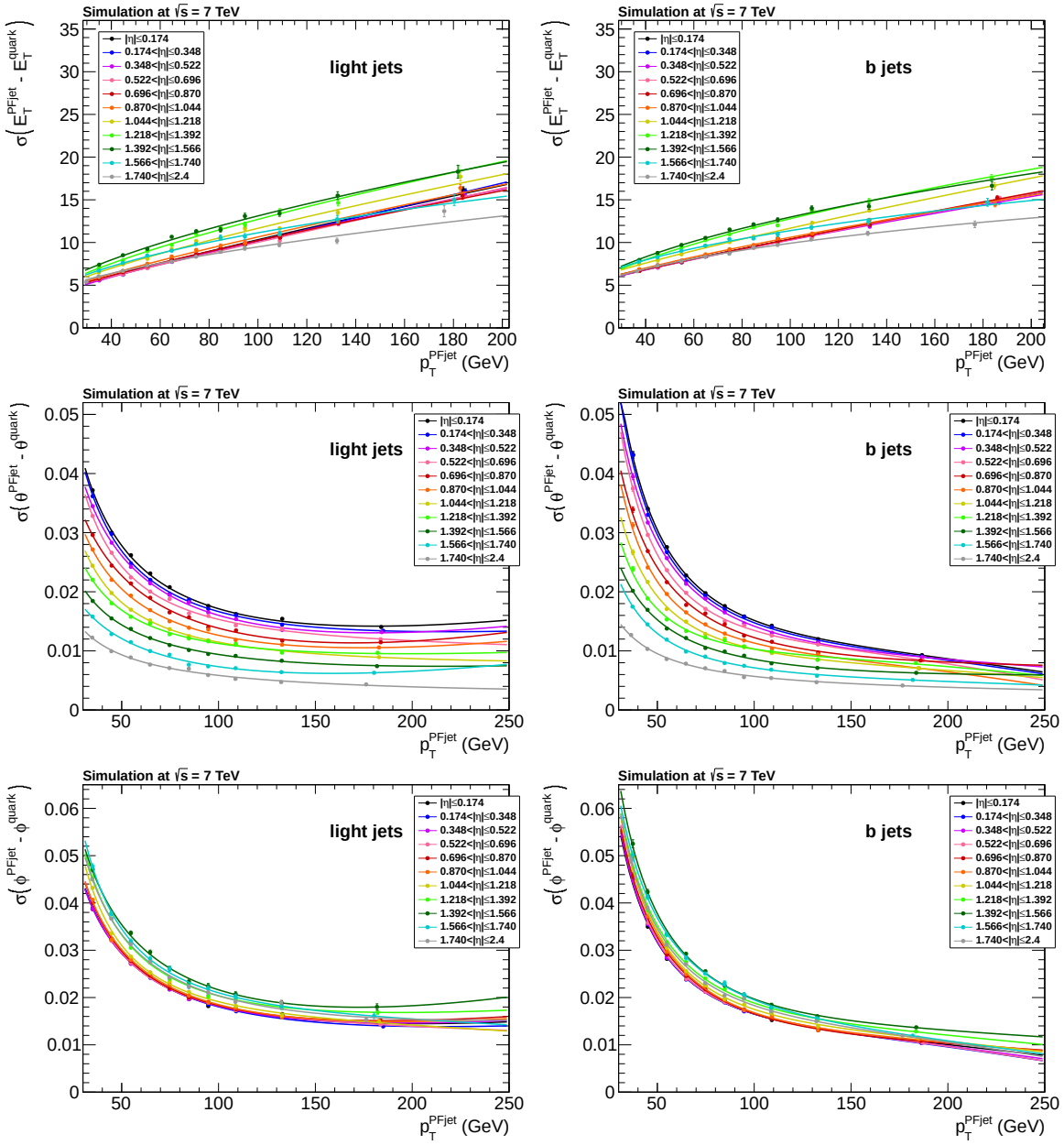

Figure 4.14: E_T (top), θ (middle) and ϕ (bottom) resolutions for reconstructed light jets (left) and b jets (right) as estimated with 8 TeV $t\bar{t}$ ℓ+jets events.

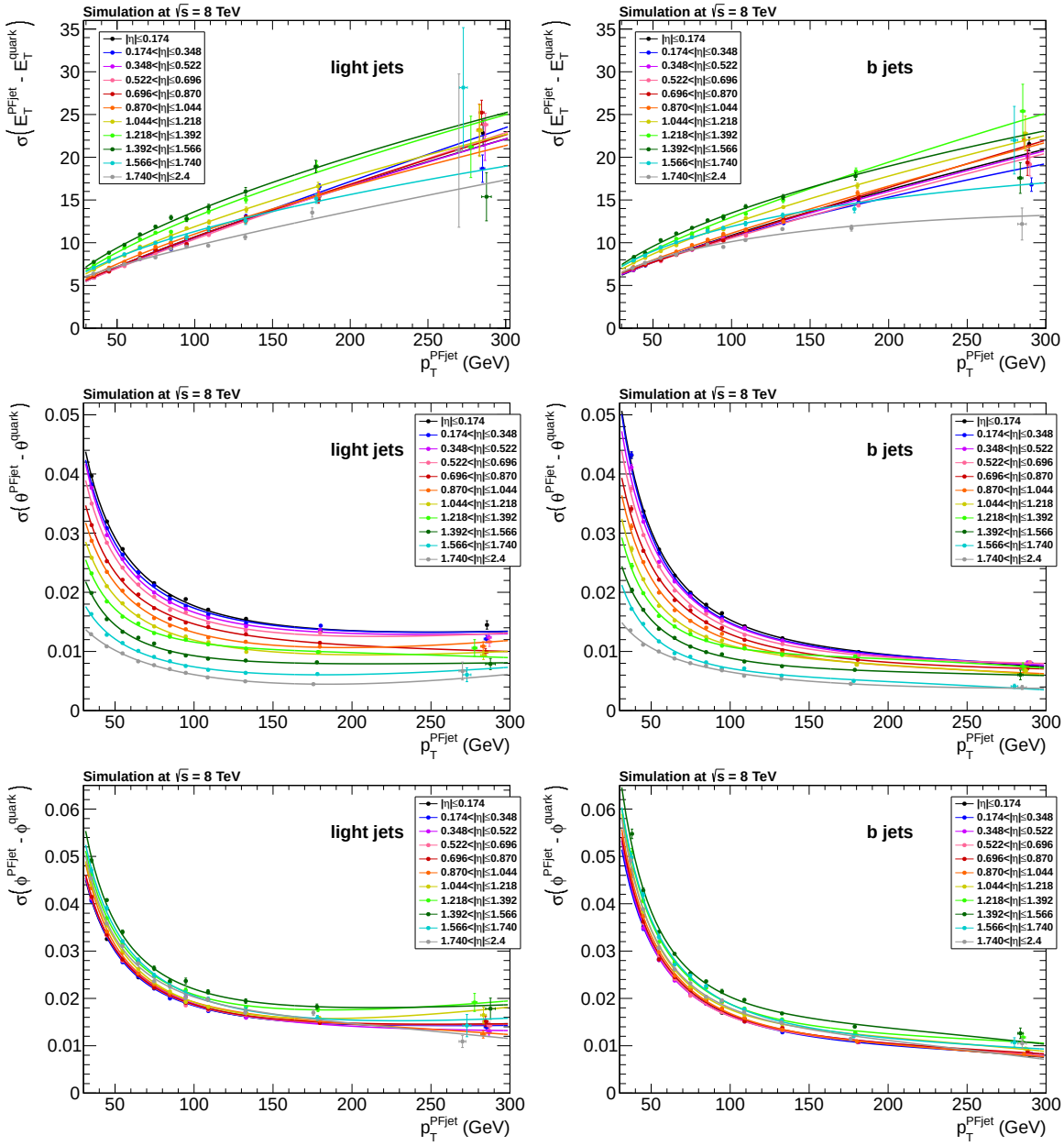

Figure 4.15: E_T (top), θ (middle) and ϕ (bottom) resolutions for reconstructed light jets (left) and b jets (right) as estimated with 8 TeV $t\bar{t}$ ℓ+jets events.

as

$$L(\vec{y}, \vec{\lambda}) = S(\vec{y}) + 2\sum_{k=1}^{m} \lambda_k f_k(\vec{y}), \tag{4.7}$$

where $\vec{\lambda}$ are the Lagrange Multipliers. This function has a local minimum when $S(\vec{y})$ is minimal and the constraints are exactly fulfilled ($f_k(\vec{y}) = 0$).

When the constraints are linear a solution can be found directly. Otherwise the minimisation is performed with an iterative approach where the constraints are linearised in each iteration as

$$f_k(\vec{y}') \approx f_k(\vec{y}^*) + \sum_{i=1}^{n} \left.\frac{\partial f_k}{\partial y_i}\right|_{y_i = y_i^*} (\Delta y_i - \Delta y_i^*)\,. \tag{4.8}$$

Here $\vec{y}$, $\vec{y}^*$ and $\vec{y}'$ are respectively the start value, the value after the previous iteration and the value after the current iteration of the measured parameters; and the differences Δy_i and Δy_i^* are defined by $\Delta y_i = y_i' - y_i$ and $\Delta y_i^* = y_i^* - y_i$. This procedure is repeated until the predefined convergence criteria are fulfilled

$$\frac{S_{n-1} - S_n}{ndf} < \epsilon_S \quad \text{and} \quad \sum_{k=1}^{m} \left| f_k^{(n)}(\vec{y}) \right| < \epsilon_F, \tag{4.9}$$

where n denotes the number of the iteration and ndf is the number of constraints. These criteria ensure that the change in χ^2 as calculated from Equation 4.6 is small and that the constraints f_k are properly fulfilled.

Top-quark mass estimation with a kinematic fit

The estimation of the mass of the hadronically decaying top-quark candidate in each jet combination is performed with the kinematic fit explained previously. The energies and momenta of the PF jets assigned to the b and light quarks from $t \to bq\bar{q}'$ decay within the jet combination are first corrected to the parton level using the flavour-dependent L5L7 jet energy corrections from Section 4.2.1. These corrected jets are parametrised via their (E_T, θ, ϕ), which are related to the energy and momentum via

$$\vec{p} = \begin{pmatrix} E_T \cos\phi \\ E_T \sin\phi \\ E_T \cot\theta \end{pmatrix} \quad \text{and} \quad E = \frac{E_T}{\sin\theta}, \tag{4.10}$$

and given as input to the kinematic fit together with their corresponding resolutions as discussed in Section 4.2.1.

In the kinematic fit two mass constraints are used. The invariant mass of the two jets identified as light jets from W-boson decay is constrained to the known W-boson mass of 80.4 GeV [9], while the invariant mass of these two jets and the hadronic b-jet candidate is constrained to the assumed top-quark mass. The top-quark mass constraint is varied in steps of 2 GeV within a range of $\pm$50 GeV around a central mass value. This central value is calculated by taking the invariant mass of the two light jets and the hadronic b jet but with the momenta of the two light jets rescaled with a common factor to make their invariant mass equal to the W-boson mass of 80.4 GeV.

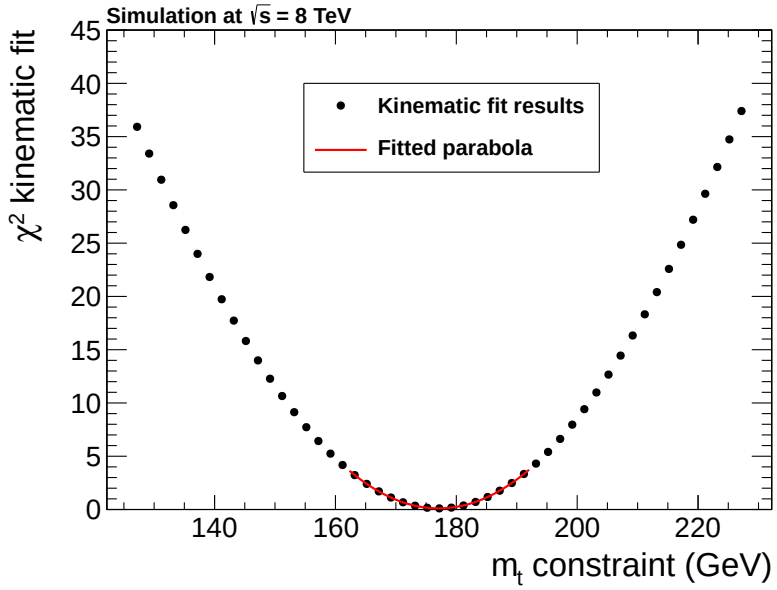

Figure 4.16: Example of the application of the kinematic fit for different top-quark mass constraints using a simulated $t\bar{t}$ ℓ+jets event.

The kinematic fit is performed for every top-quark mass constraint, as illustrated in Figure 4.16. This χ^2 as a function of the top-quark mass constraint is then fitted with a parabola of the form $a \cdot (x - m)^2 + h$ in a range of ± 15 GeV around the mass constraint with the smallest χ^2; where a, m and h are determined by the fit. The fitted value of m is taken as the fitted top-quark mass m_t^{Fit} and its uncertainty $\sigma(m_t^{Fit})$ is extracted from the parabola using $\chi^2 = \chi^2_{min,parabola} + 1$. Finally, the minimal χ^2_{min} returned by this entire fitting procedure is calculated by taking the χ^2 returned by the kinematic fit with the top-quark mass constraint equal to m_t^{Fit}.

During and after this procedure each jet combination needs to pass some additional selection cuts which are aimed at removing incorrect jet combinations where either the kinematic fit did not converge or where the imposed mass constraints were badly fulfilled. The number of points used in the parabolic fit needs to be larger than 5, the a parameter of the fitted parabola needs to be positive and finally χ^2_{min}/ndf needs to be smaller than 10. The number of degrees of freedom within the fit ndf is equal to one, since there is only one real constraint: the W-boson mass constraint. The efficiency of these criteria on correct jet combinations is estimated from simulation to be 99.7%, indicating that only wrong jet combinations are removed by these additional requirements.

The effect of the kinematic fitting procedure on the estimated top-quark mass is shown in Figure 4.17, where m_t^{Fit} is compared with the invariant mass of the three jets matched to the three quarks from top-quark decay after the L5L7 jet energy corrections. Both distributions are fitted with a Gaussian function, which shows that the relative top-quark mass resolution calculated as σ/μ improves significantly from 9.33% to 6.57%. This clearly shows the power of a kinematic fit.

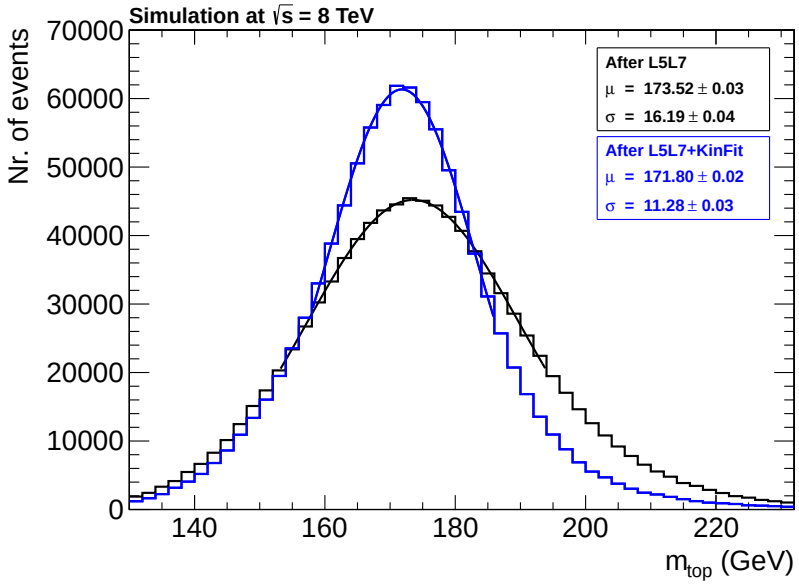

Figure 4.17: Estimated top-quark mass before and after the kinematic fit for correct jet combinations as determined from simulated $t\bar{t}$ ℓ+jets events.

4.2.3 Results after the kinematic fit

The three main variables estimated for each jet combinations by the kinematic fitting procedure (m_t^{Fit}, $\sigma(m_t^{Fit})$ and χ^2_{min}) are used by the Ideogram method to estimate the top-quark mass from the entire event sample as explained in Chapter 5. Therefore a comparison between data and simulation of these quantities is conducted. The additional selection requirements applied on each jet combination during the kinematic fit, including the $\chi^2_{min}/ndf < 10$ requirement, are applied to the results shown here. If an event has no remaining jet combinations it is rejected from further analysis. These requirements were also applied to the results shown in Section 4.1.4.

A comparison between data and simulation of some of these quantities is shown in Figures 4.18 and 4.19 for the 7 TeV analysis and in Figures 4.20 and 4.21 for the 8 TeV analysis. In general a reasonable agreement between data and simulation is observed for both centre-of-mass energies.

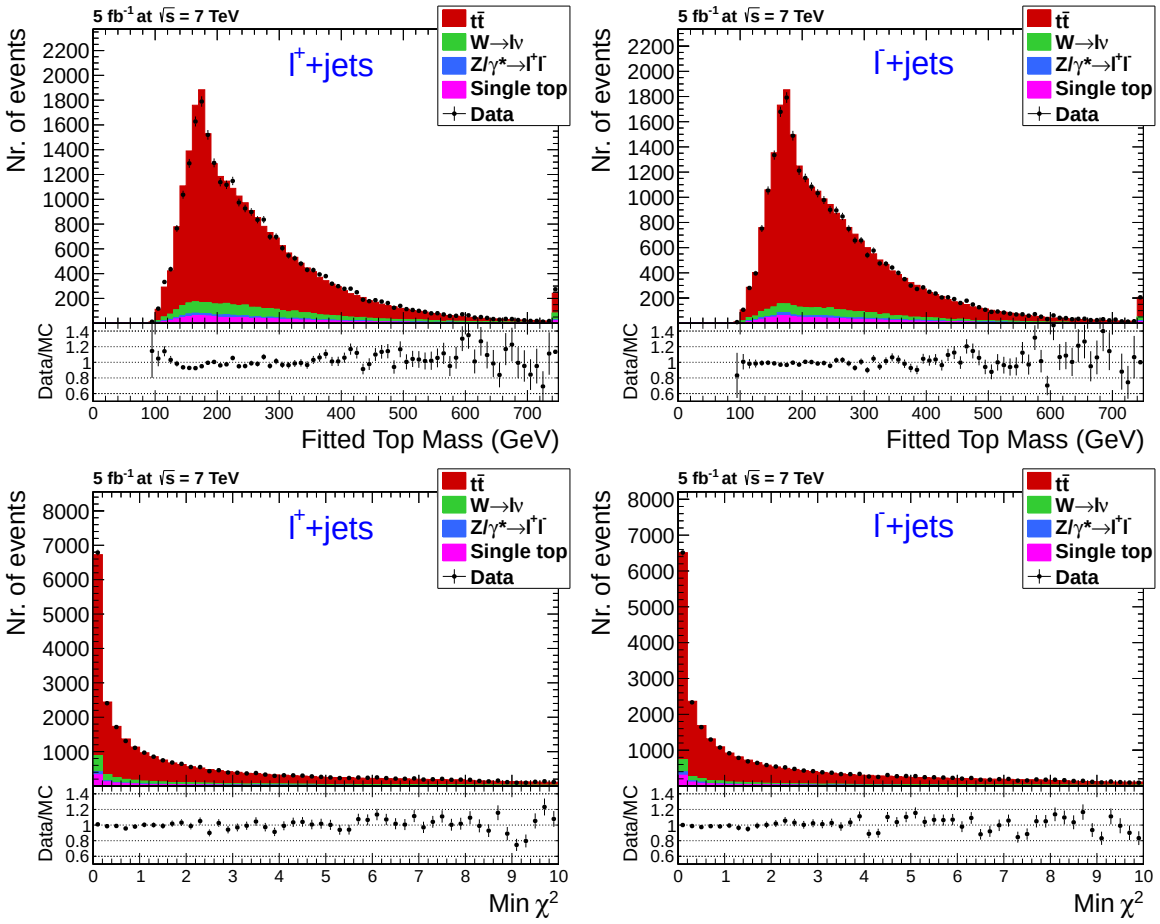

Figure 4.18: Comparison between data and simulation of m_t^{Fit} (top) and χ^2_{min} (bottom) from the jet combination with the smallest χ^2 in each event, shown separately for ℓ^++jets (left) and ℓ^-+jets (right) events.

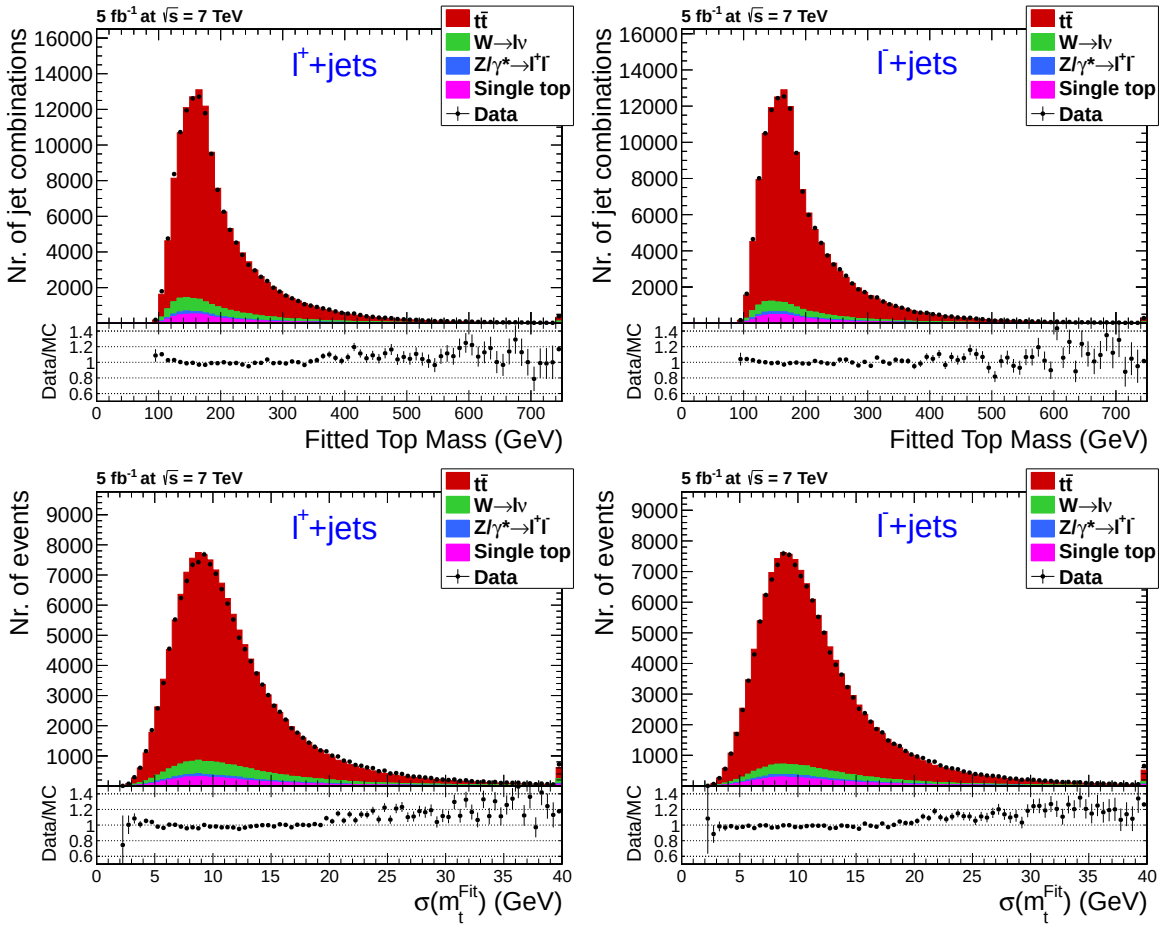

Figure 4.19: Comparison between data and simulation of m_t^{Fit} (top) and $\sigma(m_t^{Fit})$ (bottom) from all selected jet combinations, shown separately for ℓ^++jets (left) and ℓ^-+jets (right) events.

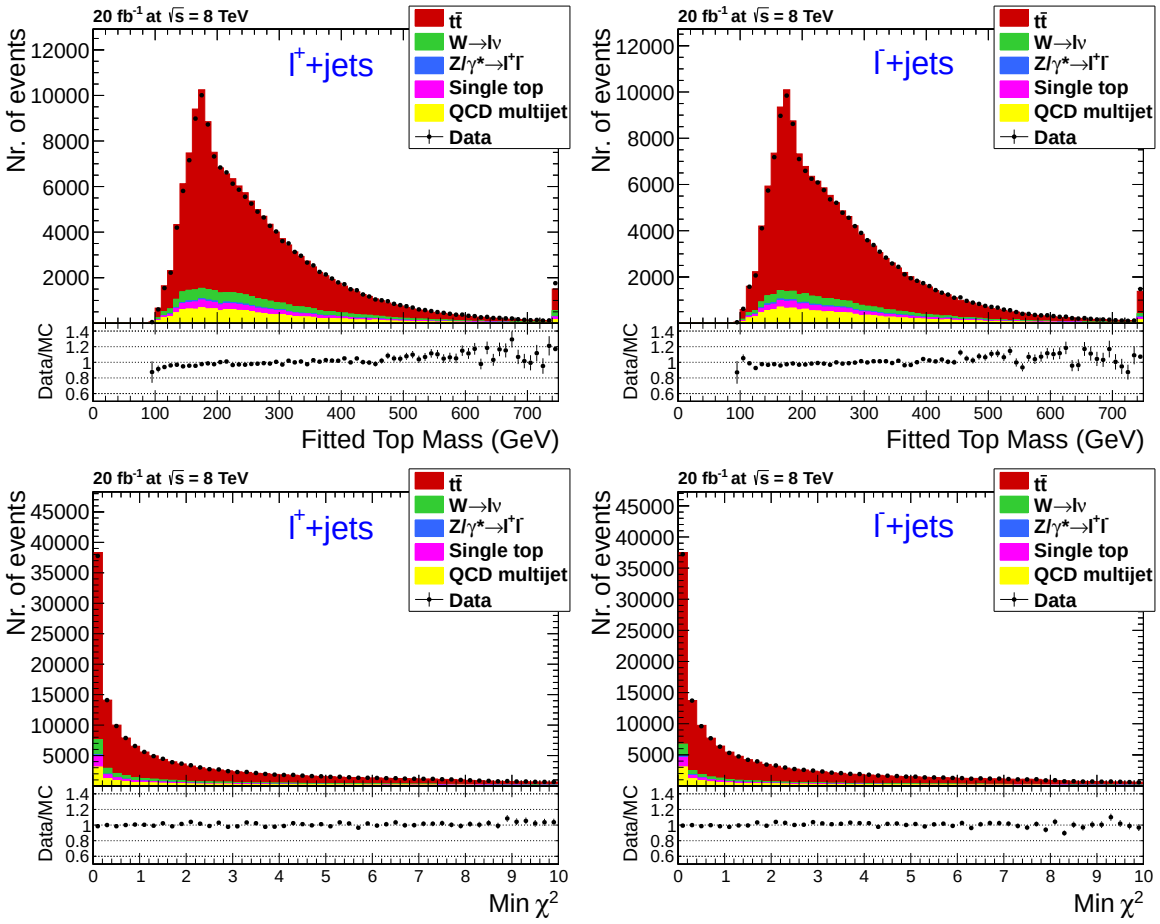

Figure 4.20: Comparison between data and simulation of m_t^{Fit} (top) and χ^2_{min} (bottom) from the jet combination with the smallest χ^2 in each event, shown separately for ℓ^++jets (left) and ℓ^-+jets (right) events.

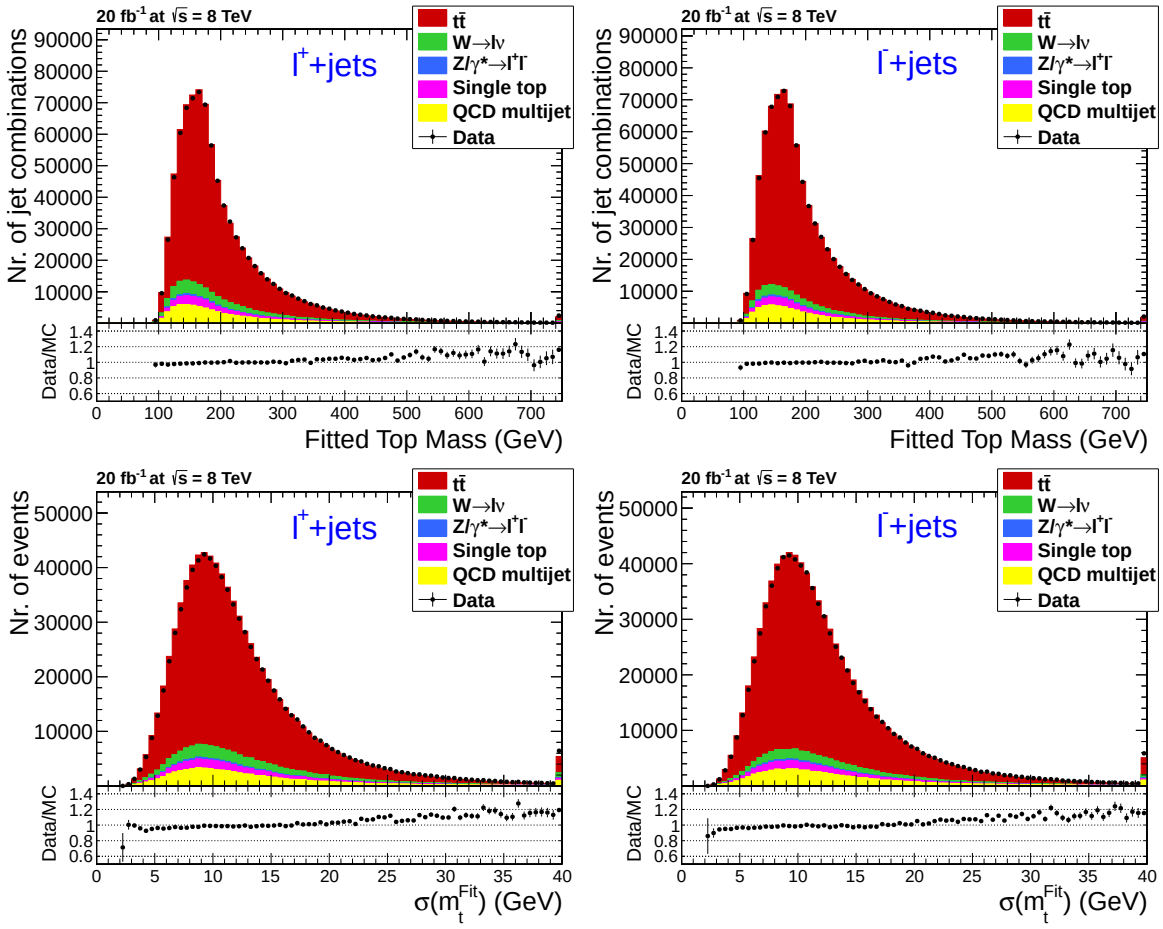

Figure 4.21: Comparison between data and simulation of m_t^{Fit} (top) and $\sigma(m_t^{Fit})$ (bottom) from all selected jet combinations, shown separately for ℓ^++jets (left) and ℓ^-+jets (right) events.

Chapter 5

Top-quark mass measurement with the Ideogram method

The mass of the top quark is one of the fundamental parameters of the Standard Model, therefore a precise measurement of this quantity is of large importance as discussed in Section 1.2.3. The most precise measurements have been performed in the ℓ+jets decay channel and are in general based on the hadronically decaying top quark, since in this case the kinematics of all final-state objects are measured and can be easily combined to estimate the top-quark mass. One of the main limitations of these methods is their high sensitivity to the jet energy scale uncertainty, so to overcome this issue combined measurements of the top-quark mass and the jet energy scale have been performed, where the known W-boson mass was used to constrain the jet energy scale.

In this chapter a more traditional approach will be used, since no combined jet energy scale measurement is performed. The mass measurement itself is conducted with the Ideogram method [128], which has been used before to measure the mass of the W boson and the top quark and is explained in Section 5.1. Due to some assumptions and simplifications made within the method presented here, a calibration procedure needs to be applied as discussed in Section 5.2. The results obtained with the 7 and 8 TeV data and the corresponding systematic uncertainties are presented in Section 5.3. Afterwards the method is optimised in order to reduce the total uncertainty, as explained in Section 5.4. Finally in Section 5.5 the top-quark mass measurement is performed in bins of several kinematic variables.

5.1 The Ideogram method

Within the Ideogram method a likelihood $\mathcal{L}_{\text{event}}(x|m_t)$ is calculated for every event as a function of the assumed top-quark mass m_t based on the output of the kinematic fit for each jet combination as explained in Section 4.2.2. The event likelihood is defined as

$$\mathcal{L}_{\text{event}}(x|m_t) = f_{t\bar{t}}\ P_{t\bar{t}}(x|m_t) + (1 - f_{t\bar{t}})\ P_{\text{bkg}}(x), \tag{5.1}$$

where $P_{t\bar{t}}(x|m_t)$ and $P_{\text{bkg}}(x)$ are the signal and background probabilities. In this equation x represents the measured observables which include the number of b-tagged jets n_b, the lepton charge q_ℓ and the output of the kinematic fit for each jet combination, namely

the fitted top-quark mass m_i, the uncertainty on the fitted top-quark mass σ_i and the minimal χ^2 of the fit χ^2_i. The fraction of signal events $f_{t\bar{t}}$ is calculated from the numbers in Tables 4.2 and 4.3 for the 7 and the 8 TeV analyses, respectively.

In the definition of the signal and background probabilities it is assumed that the number of b-tagged jets and the lepton charge are uncorrelated with the mass information in a given event. This means that the probabilities to observe a certain number of b-tagged jets and a certain lepton charge can be factorised out via

$$P_{t\bar{t}}(x|m_t) = P_{t\bar{t}}(n_b) \cdot P_{t\bar{t}}(q^\ell) \cdot P_{t\bar{t}}(x_{mass}|m_t) \tag{5.2}$$

$$P_{\text{bkg}}(x) = P_{\text{bkg}}(n_b) \cdot P_{\text{bkg}}(q^\ell) \cdot P_{\text{bkg}}(x_{mass}). \tag{5.3}$$

These probability densities for the number of b-tagged jets ($P_{t\bar{t}}(n_b)$ and $P_{\text{bkg}}(n_b)$) and for the lepton charge ($P_{t\bar{t}}(q^\ell)$ and $P_{\text{bkg}}(q^\ell)$) are taken from the signal and background samples in Tables 4.2 and 4.3 for the 7 and the 8 TeV analyses. The inclusion of b-tagging in this way reduces the impact of the backgrounds, while the inclusion of the lepton charge is to account for the charge asymmetry in the W+jets and single top backgrounds. The x_{mass} observable in Equations 5.2 and 5.3 represents the output from the kinematic fit for each jet combination (m_i, σ_i and χ^2_i). The remaining signal and background probabilities, $P_{t\bar{t}}(x_{mass}|m_t)$ and $P_{\text{bkg}}(x_{mass})$, are discussed in Sections 5.1.1 and 5.1.2, respectively, and the final top-mass extraction is explained in Section 5.1.3.

5.1.1 The signal probability

The $t\bar{t}$ signal probability is taken as a weighted sum over all jet combinations passing the selection requirements of Section 4.2.2. It consists of two terms; the first one representing the probability that a jet combination has the correct jet-to-quark assignment and the second one expressing the probability that a jet combination has a wrong jet-to-quark assignment.

$$P_{t\bar{t}}(x_{mass}|m_t) = \sum_{i=1}^{n_{combi}} w_i \left(f_{gc} \int_{m_{min}}^{m_{max}} \mathrm{d}m' \, G(m_i|m', \sigma_i) \, B(m'|m_t, \Gamma_t) + (1 - f_{gc}) \, W(m_i|m_t) \right) \tag{5.4}$$

The fraction of correct jet combinations f_{gc} in this equation is taken from simulated $t\bar{t}$ events, separately for events with $n_b = 1$ or $n_b > 1$.

The weights w_i in the signal probability equation are taken to be equal to

$$w_i = \exp\left(-\frac{1}{2}\chi^2_i\right) w_{b\text{-tag}}, \tag{5.5}$$

where the first term in this product represents the probability of the kinematic fit for the considered jet combination, and the second term reflects the degree of compatibility of the jet-to-quark assignment with the observed b-tagging assignments and is defined as

$$w_{b\text{-tag}} = \prod_j p^j. \tag{5.6}$$

Here the index j runs over all jets considered in the fit and the probabilities p^j are calculated from the b-tag efficiency ε_b and the mistag rate ε_l from Section 3.3.5. These probabilities

are equal to ε_l, $(1-\varepsilon_l)$, ε_b or $(1-\varepsilon_b)$ depending on the flavour assigned to each jet in the jet combinations under study and whether the jet is b-tagged or not. These individual weights w_i are normalised to sum to unity for each event.

The first term in the weighted sum of Equation 5.4, the probability that a jet combination has the correct jet-to-quark assignment, is calculated by taking the convolution of a Gaussian resolution function $G(m_i|m', \sigma_i)$ and a relativistic Breit-Wigner distribution $B(m'|m_t, \Gamma_t)$. This relativistic Breit-Wigner distribution is defined by

$$B(m'|m_t, \Gamma_t) \propto \frac{m'^2}{\left(m'^2 - m_t^2\right)^2 + m'^4\Gamma_t^2/m_t^2}, \tag{5.7}$$

where the width of the top quark Γ_t is fixed to 2 GeV [9]. The Gaussian function within this convolution describes the mass resolution for each jet combination. It is centred at the Breit-Wigner distributed value of the top-quark mass m' and has a standard deviation equal to the uncertainty on the fitted top-quark mass σ_i. If the smallest χ_i^2 in an event (χ^2_{min}) is larger than the number of degrees of freedom $ndf = 1$, then all the σ_i of the event are scaled up by a factor $\sqrt{\chi^2_{min}/ndf}$. This reduces the impact of events where no single jet combination properly fulfils the constraints imposed in the kinematic fit.

The second term of the weighted sum in Equation 5.4, $W(m_i|m_t)$, represents the probability that a jet combination has a wrong jet-to-quark assignment. Its shape is estimated by using jet combinations from the simulated $t\bar{t}$ event samples which are known to have a wrong jet-to-quark assignment. The distribution of the fitted top-quark mass m_i of these wrong jet combinations, taking the weights w_i from Equation 5.5 into account, is fitted with a Landau function in each of the samples with different generated top-quark masses m_t^{gen}. Some examples of these fits are shown in Figures 5.1 and 5.2 for the 7 and the 8 TeV analyses. In general these fitted top-quark mass distributions are reasonable well described by the fitted Landau functions. The mean μ and width σ of these Landau functions depend on the generated top-quark mass. Therefore μ and σ are parametrised as having a linear dependence on the generated top-quark mass, as shown in Figures 5.3 and 5.4 for 7 and 8 TeV, where these dependencies are fitted with a linear function of the form $p_0 + p_1 \cdot m_t^{gen}$. These linear parametrisations define the dependence of $W(m_i|m_t)$ on the assumed top-quark mass m_t.

5.1.2 The background probability

The background probability $P_{bkg}(x_{mass})$ is assumed to be independent on the top-quark mass m_t and its shape is obtained in a similar way as the wrong jet-to-quark assignment probability $W(m_i|m_t)$. The distribution of the fitted top-quark mass m_i of background events, taking the weights w_i from Equation 5.5 into account, is fitted again with a Landau function as shown in Figure 5.5. The backgrounds considered for these shapes are the dominant background sources from Tables 4.2 and 4.3. For the 7 TeV analysis this is the W+jets background, while for the 8 TeV analysis both the W+jets and the QCD multijet background are considered.

Finally, the background probability $P_{bkg}(x_{mass})$ is obtained from this fitted Landau function via a weighted sum over all jet combinations of the individual Landau-distributed

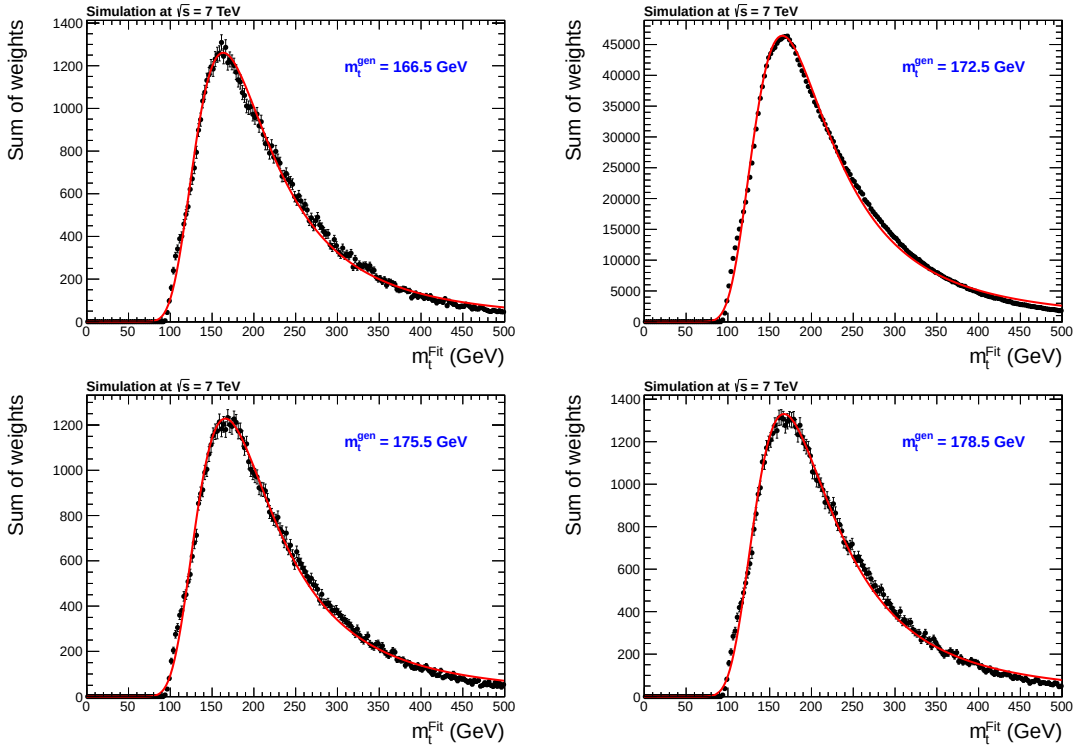

Figure 5.1: Fitted top-quark mass distributions of wrong jet combinations for different generated top-quark mass samples at 7 TeV. All are fitted with a Landau function.

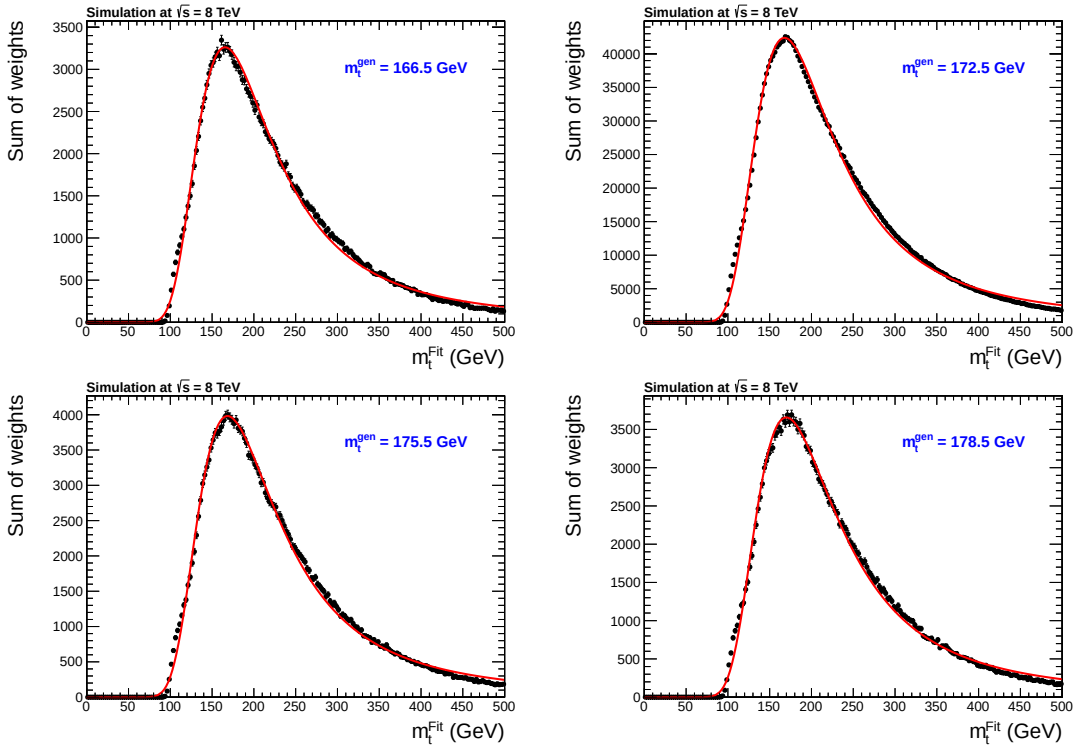

Figure 5.2: Fitted top-quark mass distributions of wrong jet combinations for different generated top-quark mass samples at 8 TeV. All are fitted with a Landau function.

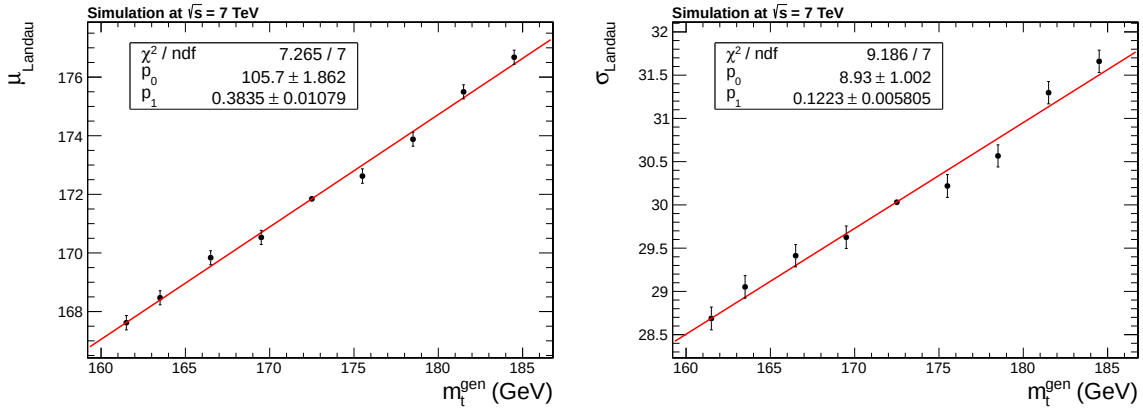

Figure 5.3: Dependence of the mean μ (left) and the width σ (right) of the fitted Landau function on the generated top-quark mass m_t^{gen} for the 7 TeV analysis.

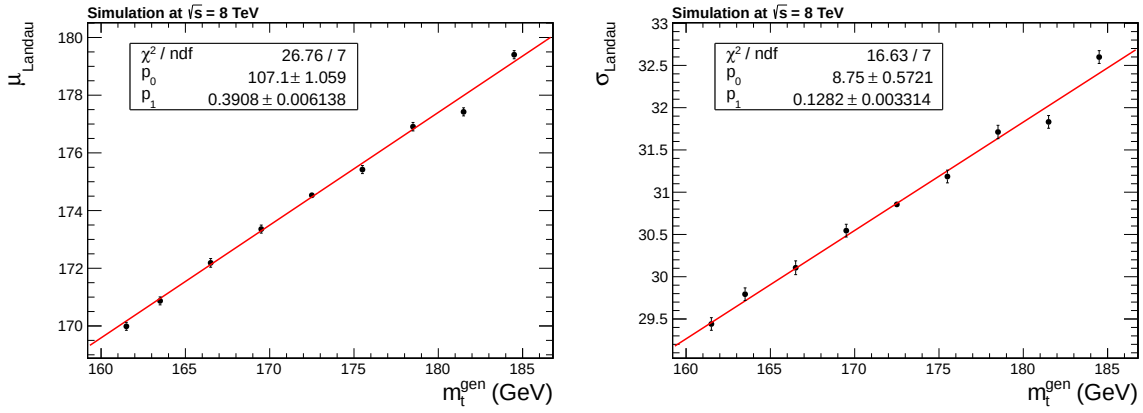

Figure 5.4: Dependence of the mean μ (left) and the width σ (right) of the fitted Landau function on the generated top-quark mass m_t^{gen} for the 8 TeV analysis.

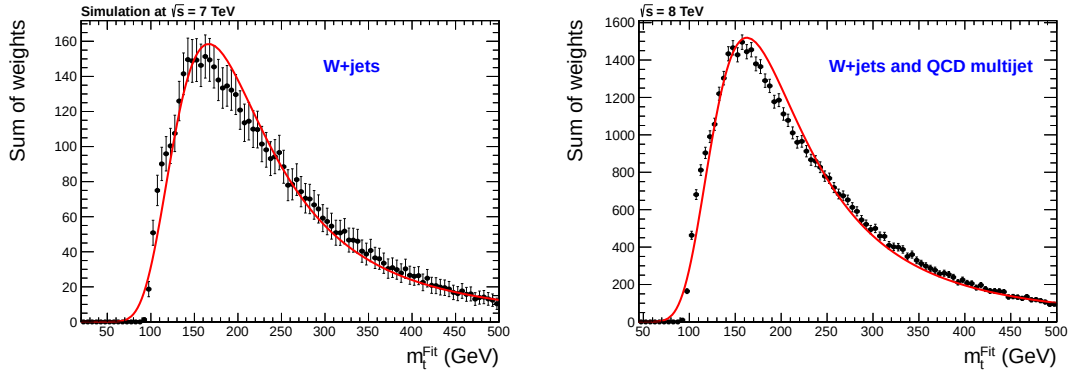

Figure 5.5: Fitted top-quark mass distributions of background events for 7 (left) and 8 TeV (right). Both are fitted with a Landau function.

probabilities $L(m_i)$ of each jet combination:

$$P_{bkg}(x_{mass}) = \sum_{i=1}^{n_{combi}} w_i \; L(m_i). \tag{5.8}$$

From the plots in Figure 5.5 it is clear that the Landau function is not perfectly describing the shapes of the fitted top-quark mass distributions for both the 7 and the 8 TeV analyses. This is one of the simplifications made within the analyses, but the method will be corrected afterwards for any residual bias arising from this and other simplifications as discussed in Section 5.2.

5.1.3 Extraction of the top-quark mass

For the final extraction of the top-quark mass the individual likelihoods for every event are calculated as explained in the previous section. Some example event likelihoods are shown in Figure 5.6 for simulated $t\bar{t}$ events with either 1 b-tagged jet or 2 b-tagged jets. When

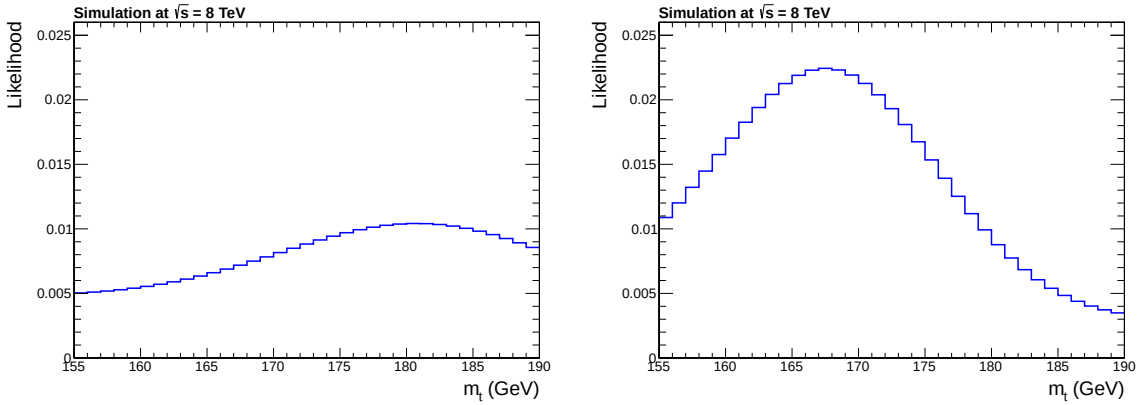

Figure 5.6: Example likelihoods for two simulated $t\bar{t}$ ℓ+jets events with 1 (left) and 2 (right) b-tagged jets.

comparing these likelihoods the effect of the weights w_i from Equation 5.5 is clear. In the event with 2 b-tags the jet combination with the correct b-tagging assignment gets a much higher weight compared to the other jet combinations, resulting in the clear peak around $m_t \sim 168.5$ GeV. In the 1 b-tag event this difference in weight between the jet combinations with correct and wrong b-tagging assignments is smaller and this results in a less pronounced peak compared to the event with 2 b-tags.

The sample likelihood as a function of the assumed top-quark mass m_t is defined as the product of all the individual event likelihoods

$$\mathcal{L}_{\text{sample}}(x|m_t) = \prod_j \mathcal{L}_{\text{event},j}\,(x|m_t). \tag{5.9}$$

In practice it is easier to use the known relationship $\chi^2 = -2\log\mathcal{L}$ and sum the log-likelihood curves of all the events:

$$\chi^2_{\text{sample}}(x|m_t) = -2\log\mathcal{L}_{\text{sample}}(x|m_t) = -2\sum_j \log\mathcal{L}_{\text{event},j}\,(x|m_t). \tag{5.10}$$

An example of such a $\chi^2_{\text{sample}}(x|m_t)$ curve is shown in Figure 5.7 as illustration, where $\Delta\chi^2_{\text{sample}} = \chi^2_{\text{sample}} - \chi^2_{\text{sample, min}}$ is plotted as calculated from the simulated 8 TeV event samples with $m_t^{gen} = 172.5$ GeV. The estimated top-quark mass and its statistical uncer-

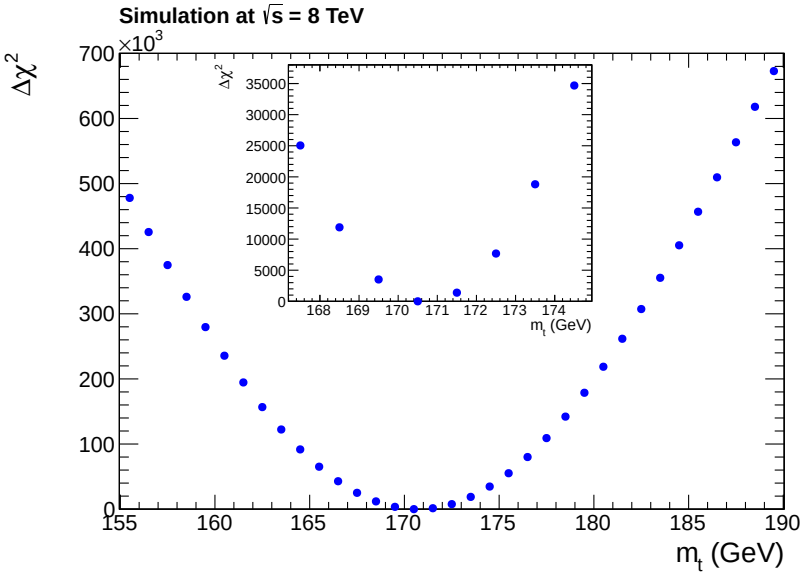

Figure 5.7: Example $\Delta\chi^2_{\text{sample}}(x|m_t)$ curve as calculated from the simulated 8 TeV event samples with $m_t^{gen} = 172.5$ GeV.

tainty are calculated from this $\chi^2_{\text{sample}}(x|m_t)$ curve by using a parabolic interpolation with the 3 χ^2 points closest to the minimum.

5.2 Calibration of the Ideogram method

The likelihood $\mathcal{L}_{\text{event}}$ as calculated for every event contains a number of assumptions, so it is only a simplified representation of the underlying probability processes. This means that the resulting combined likelihood $\mathcal{L}_{\text{sample}}$ is an approximation and needs to be calibrated with simulated events. The procedure which is applied to perform this calibration is explained in Section 5.2.1, while the results after calibration are discussed in Section 5.2.2.

5.2.1 Calibration procedure

The calibration of the Ideogram method is performed with pseudo-experiments using simulated events. In these pseudo-experiments events are picked from the different event samples listed in Tables 4.2 and 4.3 for the 7 and 8 TeV analyses, including Poisson fluctuations around the expected sample composition. In each pseudo-experiment the top-quark mass m_i and its statistical uncertainty σ_i are estimated. Finally, the results of all pseudo-experiments are used to calculate the bias and the pull distribution via

$$\text{bias} = \langle m \rangle - m_t^{gen} \quad \text{and} \quad \text{pull}_i = \frac{m_i - \langle m \rangle}{\sigma_i}, \tag{5.11}$$

where $\langle m \rangle$ is the mean of the estimated top-quark masses over all pseudo-experiments. The width of the pull is defined as the width σ of a Gaussian function which is fitted to the pull distribution.

For each generated top-quark mass value ranging from 161.5 GeV to 184.5 GeV a total of 1000 pseudo-experiments are performed. The resulting biases and pull widths are shown in Figures 5.8 and 5.9 for the 7 and 8 TeV analyses, respectively. These plots clearly show

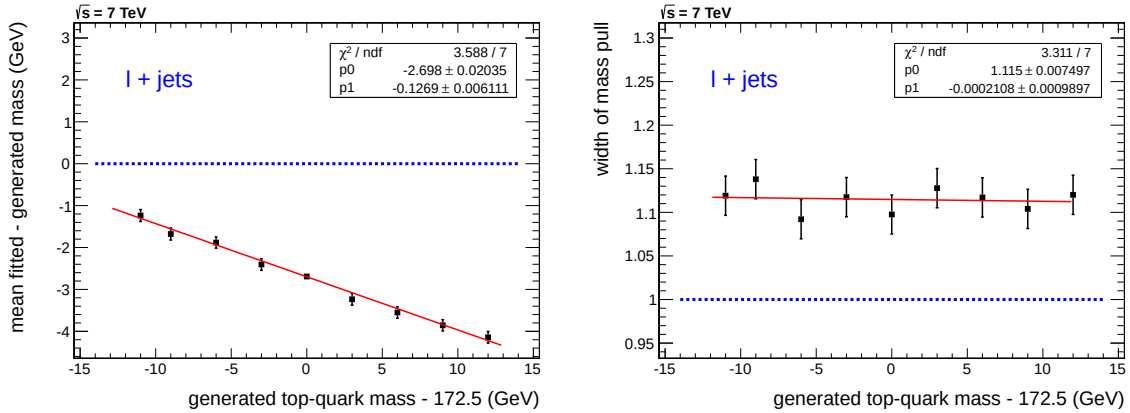

Figure 5.8: Bias on the estimated top-quark mass (left) and width of the pull distribution (right) as a function of the generated top-quark mass for the 7 TeV analysis.

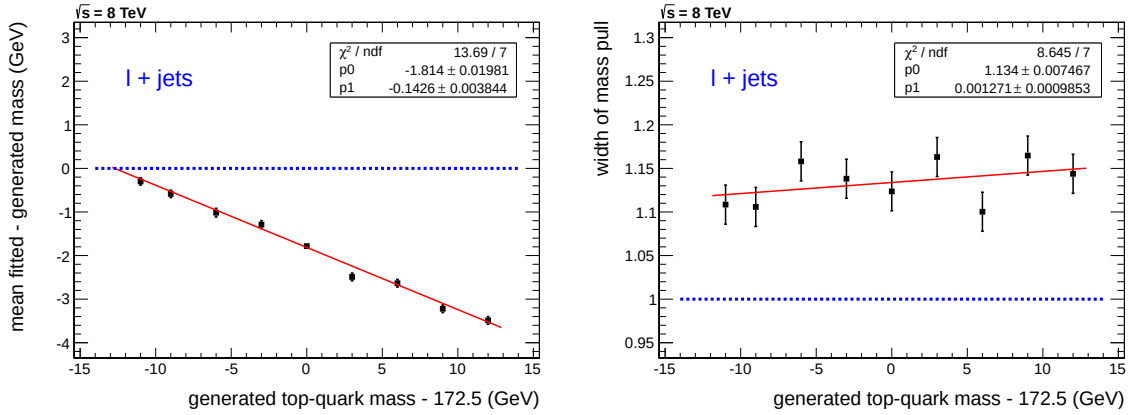

Figure 5.9: Bias on the estimated top-quark mass (left) and width of the pull distribution (right) as a function of the generated top-quark mass for the 8 TeV analysis.

the need of the calibration procedure, since residual biases are present in the method due to the assumptions and simplifications made in the definition of the event likelihood. For both the 7 and the 8 TeV analyses significant biases are observed, which also depend on the generated top-quark mass. The fitted linear calibration curves in both plots are used to correct the final estimated top-quark masses. Since the pull widths in both analyses are slightly larger than one, the statistical uncertainties on the final measurements need to be scaled up by about 11.5% and 13.4% for the 7 and the 8 TeV analyses, respectively.

5.2.2 Results of the calibration

The performance of the calibration procedure is checked by repeating all the pseudo-experiments but this time with the calibration applied to the estimated top-quark mass m_i and its statistical uncertainty σ_i. The resulting biases and pull widths are shown in Figure 5.10 for the 7 TeV analysis and in Figure 5.11 for the 8 TeV analysis. A very good

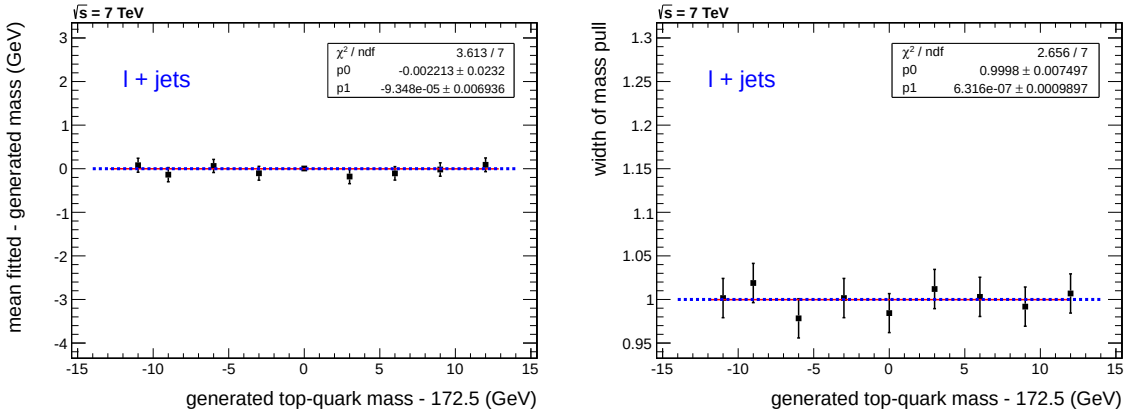

Figure 5.10: Bias on the estimated top-quark mass (left) and width of the pull distribution (right) after calibration as a function of the generated top-quark mass for the 7 TeV analysis.

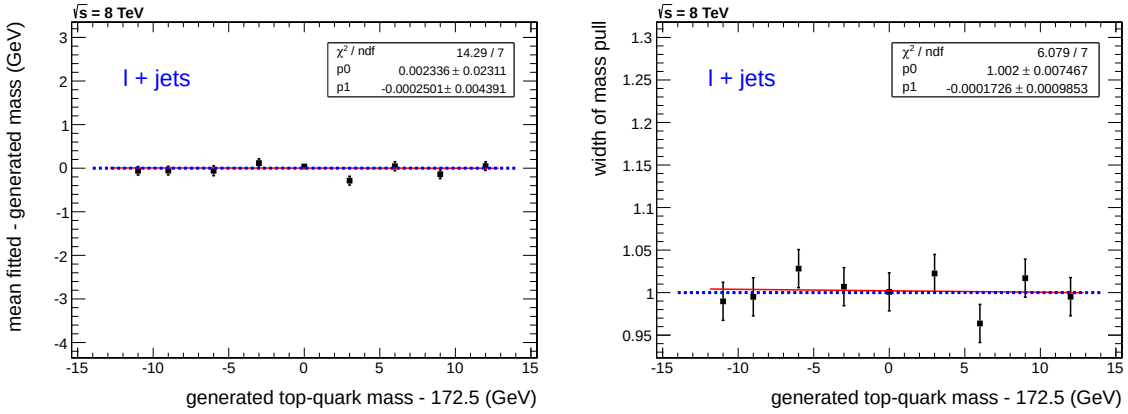

Figure 5.11: Bias on the estimated top-quark mass (left) and width of the pull distribution (right) after calibration as a function of the generated top-quark mass for the 8 TeV analysis.

closure is observed in all cases, meaning that all biases have been removed and all pull widths are compatible with one.

In Figure 5.12 the biases and pull widths observed in the samples after the e+jets and μ+jets event selection are shown. In both cases the inclusive ℓ+jets calibration curves from Figure 5.9 are applied. When comparing both plots, a small residual bias between both channels is observed which can be explained by the fact that the e+jets and μ+jets channels have a different signal fraction and background composition. Their kinematics are also slightly different due to the different event selection criteria applied in both channels.

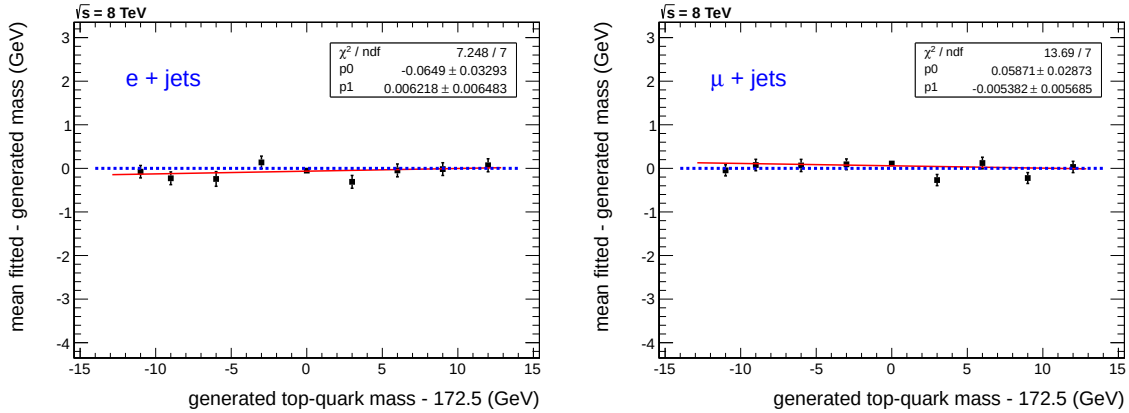

Figure 5.12: Bias on the estimated top-quark mass for events in the e+jets (left) and μ+jets (right) decay channels after the inclusive ℓ+jets calibration as a function of the generated top-quark mass for the 8 TeV analysis.

5.3 Measurement of the top-quark mass

After the calibration procedure the method is bias-free and the statistical uncertainties are known to be properly estimated. Thus the top-quark mass estimation method is applied to the real data collected by the CMS experiment, as explained in Section 5.3.1. The different systematic uncertainties which need to be considered in the context of this measurement are discussed in Section 5.3.2.

5.3.1 Results on data

The $\Delta\chi^2_{\text{sample}}(x|m_t)$ curves obtained by applying the Ideogram method separately to the 7 and the 8 TeV datasets are shown in Figures 5.13 and 5.14, which are calculated before the calibration is applied. From these curves the top-quark masses and the corresponding uncertainties are extracted, and the calibration procedure is applied. The results obtained on both datasets are listed in Table 5.1.

	e+jets	μ+jets	ℓ+jets
7 TeV (4.9 fb^{-1})	173.10 $\pm$ 0.32 GeV	172.77 $\pm$ 0.28 GeV	172.89 $\pm$ 0.21 GeV
8 TeV (19.7 fb^{-1})	173.11 $\pm$ 0.14 GeV	172.63 $\pm$ 0.13 GeV	172.84 $\pm$ 0.10 GeV

Table 5.1: Measured top-quark masses and their statistical uncertainties in the different decay channels and the combined ℓ+jets channel, using the 7 and 8 TeV datasets.

These results show an almost perfect compatibility between the 7 and the 8 TeV measurements. When the results of the e+jets and μ+jets decay channels are compared some small differences are observed, with the μ+jets channels giving slightly lower top-quark mass values. This is expected since the calibration curves were estimated inclusively for ℓ+jets events and not separately for e+jets and μ+jets events, as explained in Section 5.2. With the 7 TeV dataset this difference is measured to be $m_t^e - m_t^\mu =$

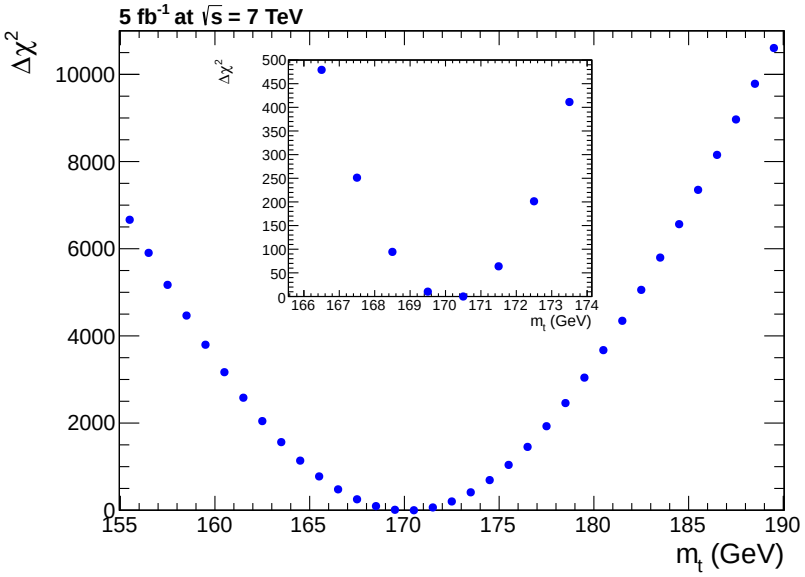

Figure 5.13: $\Delta\chi^2_{\text{sample}}(x|m_t)$ curve obtained with the 7 TeV ℓ+jets dataset.

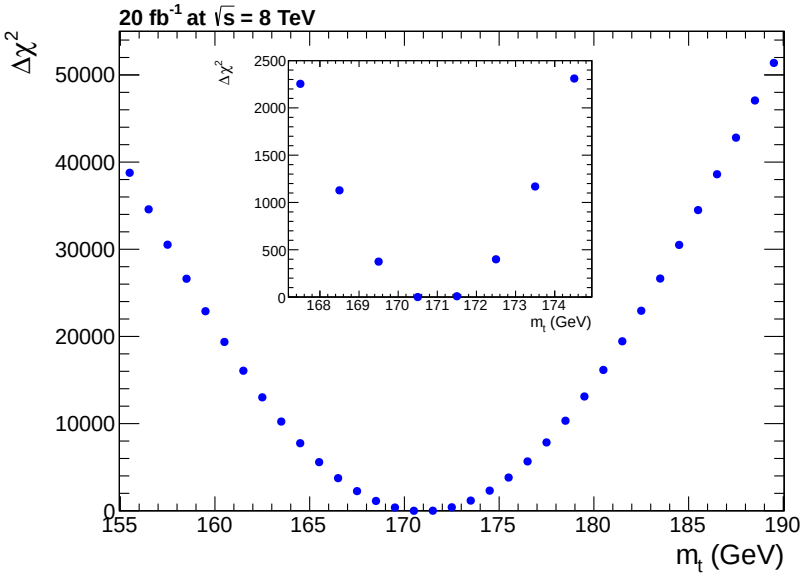

Figure 5.14: $\Delta\chi^2_{\text{sample}}(x|m_t)$ curve obtained with the 8 TeV ℓ+jets dataset.

0.33 ± 0.43 GeV, which is statistically compatible with zero. With the 8 TeV dataset a difference of $m_t^e - m_t^\mu = 0.48 \pm 0.19$ GeV is observed, which is statistically compatible with a difference of 0.17 ± 0.05 GeV as measured in the simulation and with the results shown in Figure 5.12. All individual values shown in the table are also compatible with previous measurements of the top-quark mass, as presented in Section 1.2.2.

5.3.2 Systematic uncertainties on the measured top-quark mass

Several systematic effects can induce a shift on the measured top-quark mass. These can originate from our understanding of the detector performance, the robustness of the reconstruction algorithms, assumptions made within the mass-extraction method, and the description of $t\bar{t}$ events in the simulation.

All systematic effects are evaluated using simulated event samples, by comparing the nominal sample to a sample where the systematic effect is varied by ± 1 standard deviation. Since some of these systematic effects introduce very small shifts on the estimated top-quark mass, the statistical uncertainty of these induced shifts is calculated for every systematic effect. This is done with the Jackknife method as explained below.

The Jackknife method

The Jackknife resampling technique [129] is used to estimate the statistical uncertainty on a complicated quantity from which the underlying probability distribution and/or internal correlations are unknown. When the Jackknife method is applied to an observable X estimated from a sample of n independent measurements, the estimation of X is repeated n times. Every time the estimation is repeated with exactly one measurement i removed from the full sample, resulting in a set of n estimated quantities X_i^{Jack}. From this set the statistical uncertainty on X can be calculated via

$$\sigma_X^{Jack} = \sqrt{\frac{n-1}{n} \sum_{i=1}^{n} \left(X_i^{Jack} - X\right)^2}. \tag{5.12}$$

In the case of the top-quark mass measurement the observable X is taken as the shift introduced due to a certain systematic variation, $X = m_t^{Nominal} - m_t^{Systematic}$. Equation 5.12 is then used to estimate the statistical uncertainty on the observed systematic shift in m_t.

Systematic uncertainties on the 7 and 8 TeV analyses

An overview of all the systematic uncertainties considered for the measurement of the top-quark mass with the 7 and 8 TeV datasets can be found in Table 5.2. The numbers shown here are the maximal observed shifts when the systematic effect is varied by ± 1 standard deviation in simulation, and the corresponding statistical uncertainties on these shifts as calculated with the Jackknife method.

For some systematic uncertainties the statistical significance of the observed shift in m_t is small. Therefore the observed shift is quoted as systematic uncertainty when it is larger than the statistical uncertainty and otherwise just the statistical uncertainty is quoted, as indicated by the bold script in the table. The total systematic uncertainty is

taken to be the quadratic sum of the values quoted for each source. The evaluation of all these systematic uncertainties is explained below. There some potential additional sources which were included in previous top-quark mass measurements are also discussed, like the uncertainty on the modelling of the trigger and on the modelling of hadronisation.

	Estimated effect on m_t (MeV)	
Source	7 TeV	8 TeV
Jet energy scale	**1160** $\pm$ 6	**832** $\pm$ 6
Jet energy resolution	**93** $\pm$ 5	**86** $\pm$ 6
b-tagging efficiency	**42** $\pm$ 4	**22** $\pm$ 3
Pileup	**51.3** $\pm$ 0.1	**29.9** $\pm$ 0.2
Signal fraction	**57** $\pm$ 1	**161** $\pm$ 1
Background composition	**12.7** $\pm$ 0.3	**75.0** $\pm$ 0.3
Method calibration	17 $\pm$ **24**	**39** $\pm$ 27
Parton distribution functions	**142** $\pm$ 1	**172** $\pm$ 2
Q^2-scale; initial- and final-state radiation	**525** $\pm$ 103	**603** $\pm$ 96
ME-PS matching threshold	**226** $\pm$ 97	**110** $\pm$ 94
Colour reconnection	**205** $\pm$ 86*	16 $\pm$ **55**
Underlying event	**227** $\pm$ 90*	64 $\pm$ **89**
Total	1335	1073

Table 5.2: Overview of the systematic uncertainties on m_t. For each contribution the larger among the estimated shift and its statistical uncertainty is quoted, as indicated by the bold script. The asterisk * denotes that the systematic was evaluated with FASTSIM samples.

Jet energy scale

This uncertainty is evaluated by propagating the p_T- and η-dependent jet energy scale uncertainties from Figures 3.11 and 3.12 to the estimated top-quark mass. Therefore the energy of each jet is scaled up/down within these uncertainties. Since the top-quark mass is estimated from its decay into three jets, the measurement is very sensitive to any uncertainty on the jet energy scale, making this the most dominant systematic uncertainty.

Jet energy resolution

For the evaluation of this uncertainty the energy of each jet was smeared up/down within the $|\eta|$-dependent jet energy resolution uncertainties shown in Figure 3.13. Although the top-quark mass as estimated from its three-jet decay is very sensitive to any jet-energy related uncertainty, the effect of the jet energy resolution is

reasonably small since the up/down scaling results only in small shifts of the energy of the jet.

b-tagging efficiency

A mismodelling of the *b*-tag efficiency can bias the measurement via altering the effect of wrong jet combinations and background events. To quantify the impact of the *b*-tagging efficiency, the threshold defining the working point is altered which results in a change in efficiency. An absolute change in efficiency of $\pm 1.9\%$ was introduced for the 7 TeV analysis, while for the 8 TeV analysis an efficiency change of $\pm 1.2\%$ was applied. These numbers correspond to the uncertainties quoted on the *b*-tag efficiencies in Section 3.3.5.

Pileup

The effect of additional pileup collisions is included in all simulated event samples, which are reweighted to match the observed pileup profile in data. To estimate the systematic uncertainty related to pileup the mean number of interactions is changed in simulation by $\pm 5\%$ for the 7 TeV analysis and by $\pm 6\%$ for the 8 TeV analysis. The larger 8 TeV variation is motivated by a larger uncertainty on the measured luminosity at 8 TeV, which is used to estimate the observed pileup profile in data as explained in Section 4.1.4.

The pileup dependence of the top-quark mass measurement is small. This can be seen from the dependence of the measured top-quark mass on the primary vertex multiplicity as shown in Figure 5.15. The data results of both analyses are statistically compatible with no dependence on the number of reconstructed primary vertices.

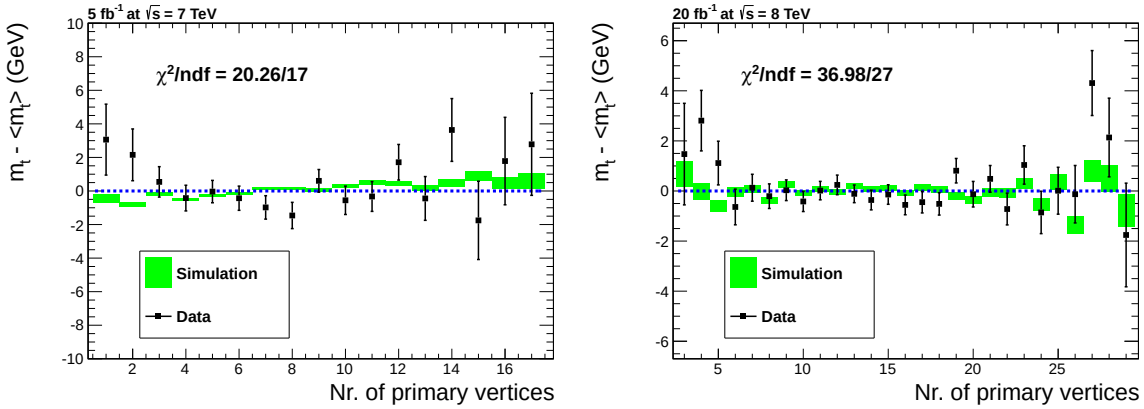

Figure 5.15: Dependence of the top-quark mass on the number of primary vertices in data and simulation for the 7 TeV (left) and 8 TeV (right) analyses. The χ^2/ndf value is calculated for the data with respect to the blue curve, which corresponds to $m_t = \langle m_t \rangle$.

Signal fraction

The signal fraction is one of the input parameters of the Ideogram method. A change in signal fraction will also bias the measurement since signal and background events have different fitted top-quark mass distributions. Therefore the signal fraction is changed by a relative $\pm 10\%$ in both analyses.

Background composition

Residual effects due to the composition of the background can be present, since not all backgrounds have the same fitted top-quark mass distributions. The uncertainty is evaluated by scaling each background source up and down, but keeping the signal fraction fixed. The W+jets, Z/γ^*+jets, single top t-channel and single top tW-channel backgrounds are each independently scaled up/down by a relative 30%, while the QCD multijet background in the 8 TeV analysis is scaled up/down by a relative 50%. The individual top-quark mass shifts induced by these variations are listed in Table 5.3.

	Effect on m_t (MeV)	
Background	7 TeV	8 TeV
W+jets ($\pm$30%)	3.6 $\pm$ 0.2	3.5 $\pm$ 0.2
Z/γ^*+jets ($\pm$30%)	7.78 $\pm$ 0.02	0.10 $\pm$ 0.02
Single top t-channel ($\pm$30%)	7.04 $\pm$ 0.02	11.38 $\pm$ 0.03
Single top tW-channel ($\pm$30%)	6.1 $\pm$ 0.2	12.5 $\pm$ 0.2
QCD multijet ($\pm$50%)	/	73.0 $\pm$ 0.2
Total	12.7 $\pm$ 0.3	75.0 $\pm$ 0.3

Table 5.3: Overview of the subcomponents of the background composition systematic uncertainty. The total uncertainty is calculated as the quadratic sum of all individual components.

Method calibration

The calibration procedure is limited in precision by the statistics of the simulated event samples. Therefore the residual bias after calibration is calculated with the nominal event samples (with $m_t = 172.5$ GeV) and quoted as systematic uncertainty. Both biases are statistically compatible with zero.

Parton distribution functions

As explained in Section 3.1.1, the simulated event samples were generated with the CTEQ parton distribution functions (PDFs) which are described by 22 independent parameters. The up and down variation of each of these parameters results in 22 accompanying PDF up/down possibilities. The nominal simulated event sample is reweighted according to the deviation of each PDF from its original form. The sum of the larger shift (up or down) for each change in PDF is taken in quadrature to define the combined PDF uncertainty.

Q^2-scale; initial- and final-state radiation

The uncertainty on the amount of initial- and final-state radiation and on the choice of Q^2-scale during the event generation is estimated with dedicated simulated $t\bar{t}$ samples. In these samples the Q^2-scale is varied up and down by a factor 2 and the amount of initial- and final-state radiation is also increased and decreased simultaneously.

The significant effect of these variations on the measured top-quark mass is caused by two effects. Firstly, the increase and decrease of the amount of additional radiation jets influence the choice of the four leading jets from which the fitted top-quark masses are estimated. Secondly, final-state radiation originating from one of the quarks of $t \to bq\bar{q}'$ decay will result in a significant mismeasurement of the quarks energy and bias the fitted top-quark masses.

ME-PS matching threshold

The uncertainty on the threshold chosen in the matrix-element parton-shower matching is evaluated with dedicated $t\bar{t}$ samples. The threshold is scaled up and down by a factor 2 in these samples.

Colour reconnection

To estimate the effect of colour reconnection on the measured mass, two $t\bar{t}$ samples with different PYTHIA tunes are compared as explained in Section 3.1.4. The P11 tune, which includes colour reconnection, and the P11noCR tune, where colour reconnection was disabled, are compared. In the 7 TeV analysis this systematic uncertainty is evaluated with FASTSIM samples.

Underlying event

The uncertainty on the modelling of the underlying event is propagated onto the measured top-quark mass with $t\bar{t}$ samples produced with different PYTHIA tunes as discussed in Section 3.1.4. The P11 tune is compared to the P11mpiHi and P11TeV, which have more and less underlying event activity, respectively. The individual mass shifts observed when comparing these samples are listed in Table 5.4. In the 7 TeV analysis this systematic uncertainty is evaluated with FASTSIM samples.

	Effect on m_t (MeV)	
PYTHIA tune	7 TeV	8 TeV
Tune P11 vs. tune P11TeV	197 $\pm$ 64	63 $\pm$ 63
Tune P11 vs. tune P11mpiHi	113 $\pm$ 63	11 $\pm$ 63
Total	227 $\pm$ 90	64 $\pm$ 89

Table 5.4: Overview of the subcomponents of the underlying event systematic uncertainty. The total is calculated as the quadratic sum of the individual components.

Trigger

The trigger requires either the presence of an isolated lepton and at least three jets with $p_T > 30$ GeV, or only the presence of an isolated lepton. As the lepton is not used in the mass reconstruction, no systematic effect is expected from any mismodelling of the lepton trigger efficiency or p_T threshold. The requirement of three jets with $p_T > 30$ GeV in the 7 TeV trigger is highly efficient for events with 4 jets with $p_T > 30$ GeV. Any effect on kinematic distributions of the jets in selected events is thus expected to be small. Therefore no uncertainty is quoted for this source.

Hadronisation

In other m_t measurements the uncertainty on the modelling of the hadronisation is in general evaluated by comparing the two standard hadronisation algorithms, namely PYTHIA and HERWIG. In this analysis the flavour-dependent part of the jet energy scale uncertainty considered in both the 7 and 8 TeV analyses already includes the difference in jet response observed between PYTHIA and HERWIG, as explained in Section 3.3.4. Therefore no additional hadronisation uncertainties need to be taken into account since this will only lead to a double-counting of this uncertainty.

When these systematic uncertainties are combined with the measured top-quark masses from Section 5.3.1 the final results are obtained. With 4.9 fb^{-1} of 7 TeV data a value of

$$m_t = 172.89 \pm 0.21\,(\text{stat.}) \pm 1.34\,(\text{syst.})\ \text{GeV}$$

is obtained, while 19.7 fb^{-1} of 8 TeV data results in a measured top-quark mass of

$$m_t = 172.84 \pm 0.10\,(\text{stat.}) \pm 1.07\,(\text{syst.})\ \text{GeV}.$$

Both results are in excellent agreement with each other and with previous m_t measurements as discussed in Section 1.2.2. The precision of the 8 TeV results is better than the 7 TeV results, which is mainly caused by the reduced jet energy scale uncertainties at 8 TeV.

5.4 Optimisation of the top-quark mass measurement

To further reduce the total uncertainty on the top-quark mass an optimisation procedure is performed which can potentially reduce the main systematic uncertainties. Therefore the entire 8 TeV analysis is repeated multiple times as explained in Section 5.4.1. Here the most optimal analysis is chosen. Additional details and the complete results of the most optimal analysis are shown in Sections 5.4.2 and 5.4.3, respectively.

5.4.1 Optimisation procedure

Since the total uncertainty on the measured top-quark mass is completely dominated by the jet energy scale uncertainty, the optimisation focuses mainly on the reduction of this uncertainty. It was shown in Section 3.3.4 that this jet energy scale uncertainty depends on the reconstructed PF jet p_T and η. The most straightforward option to reduce this uncertainty is thus to tighten the kinematic criteria applied to the jets during the event selection.

To assess which combination of criteria on the p_T and η of the jets is optimal, the analysis as explained before is repeated for every set of criteria. This includes recalculating the shapes for backgrounds and wrong jet combinations that enter the Ideogram method, and redoing the calibration procedure. Each time the total uncertainty is calculated from the statistical uncertainty, the jet energy scale systematic uncertainty, and the Q^2-scale systematic uncertainty; since these are the leading uncertainties. Finally the combination

of criteria with the smallest total uncertainty is chosen and the remaining systematic uncertainties are evaluated for this set of jet criteria.

The results for the different combination of jet selection criteria are shown in Table 5.5. As expected a clear reduction of the jet energy scale uncertainty is seen when going to

			Estimated uncertainty on m_t (GeV)		
Jet selection	$N_{t\bar{t}}$	Data result	JES	Q^2-scale	Total
$p_T > 30$ GeV, $\|\eta\| < 2.4$	256860	172.84 $\pm$ 0.10	**0.83** $\pm$ 0.01	**0.60** $\pm$ 0.10	1.03
$p_T > 30$ GeV, $\|\eta\| < 1.9$	212357	172.62 $\pm$ 0.10	**0.79** $\pm$ 0.01	**0.46** $\pm$ 0.10	0.92
$p_T > 40$ GeV, $\|\eta\| < 1.9$	110490	172.72 $\pm$ 0.14	**0.74** $\pm$ 0.01	**0.69** $\pm$ 0.14	1.00
$p_T > 50$ GeV, $\|\eta\| < 1.9$	51743	172.79 $\pm$ 0.20	**0.78** $\pm$ 0.02	**0.94** $\pm$ 0.21	1.24
$p_T > 30$ GeV, $\|\eta\| < 1.3$	119033	172.51 $\pm$ 0.13	**0.77** $\pm$ 0.01	**0.42** $\pm$ 0.13	0.89

Table 5.5: Results of the application of different kinematic selection criteria on the jets. For each set of criteria the number of $t\bar{t}$ events $N_{t\bar{t}}$ expected from simulation is shown, together with the result as obtained from the 8 TeV data and the statistical, the jet energy scale, and the Q^2-scale systematic uncertainty. The total uncertainty is calculated by taking the quadratic sum of these uncertainties.

tighter jet selection criteria. When tightening the $|\eta|$ criterion, a reduction of the Q^2-scale systematic uncertainty is also observed. However when tightening the p_T criterion the Q^2-scale systematic increases significantly, therefore a tighter jet p_T criterion is ruled out. The smallest total uncertainty is obtained when requiring at least four jets with $p_T > 30$ GeV within $|\eta| < 1.3$. More details about this most optimal analysis are given below.

5.4.2 Ideogram shapes and calibration curves of the most optimal analysis

The fitted top-quark mass distributions of background events and wrong jet combinations depend strongly on the kinematic selection applied to the jets. These shapes need therefore to be re-estimated for every set of kinematic requirements. This means that also the calibration procedure needs to be reapplied because the residual biases will be different.

Like in Section 5.1.1, the fitted top-quark mass distributions of wrong jet-to-quark assignments are fitted with Landau functions for each of the generated top-quark masses. Some examples of these fits are shown in Figure 5.16. In general a similar level of agreement was obtained for the optimised analysis, as can be seen from these plots. The dependence of the mean μ and width σ of these Landau functions is again assumed to be linear and parametrised via the fitted line in Figure 5.17.

The shape of the fitted top-quark mass distribution of background events was also estimated in the same way as explained before (Section 5.1.2). Only the dominant background sources, W+jets and QCD multijet, were used and fitted with Landau functions. Here the

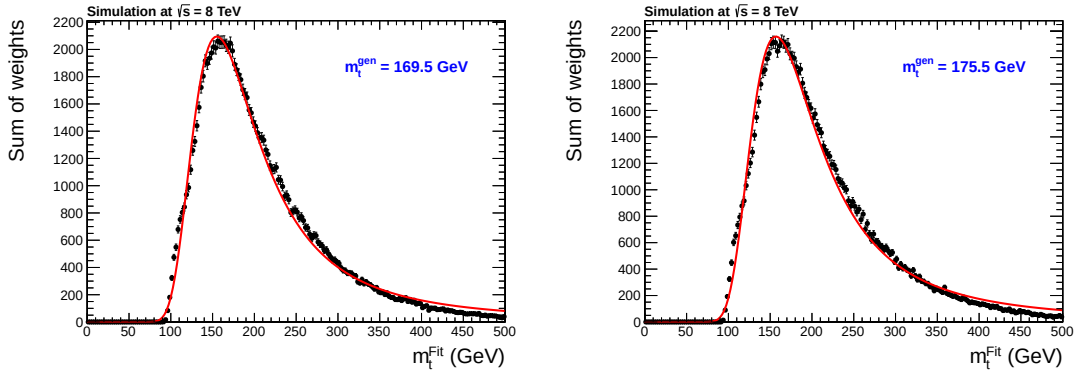

Figure 5.16: Fitted top-quark mass distributions of wrong jet combinations for different generated top-quark masses for the 8 TeV analysis with $|\eta^{PFjet}| < 1.3$. Both distributions are fitted with a Landau function.

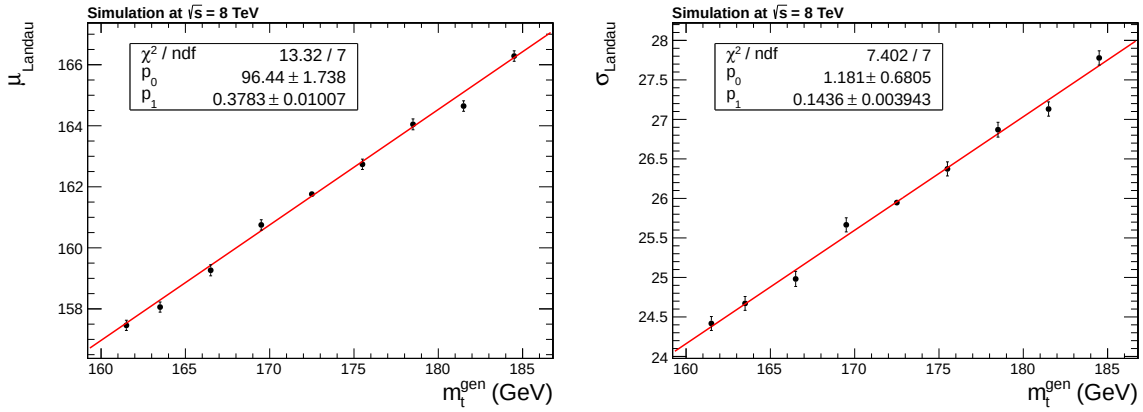

Figure 5.17: Dependence of the mean μ (left) and the width σ (right) of the fitted Landau function on the generated top-quark mass m_t^{gen} for the 8 TeV analysis with $|\eta^{PFjet}| < 1.3$.

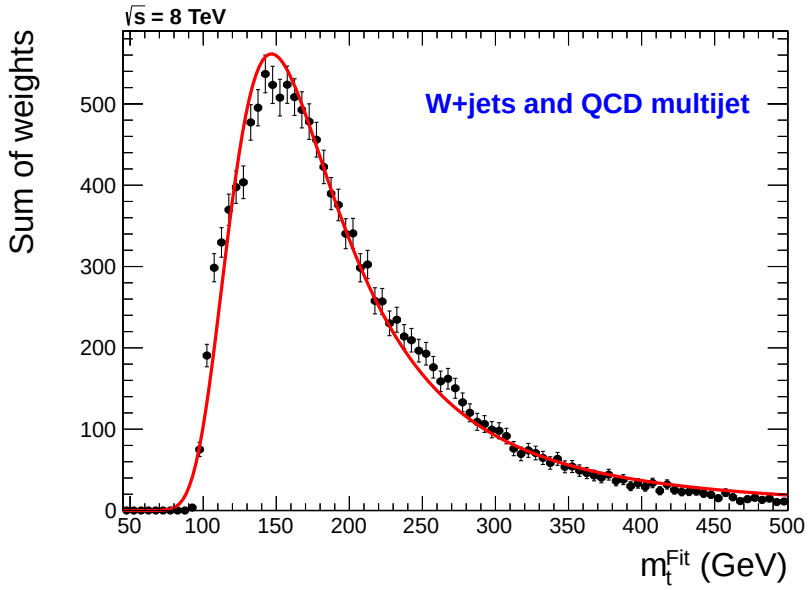

Figure 5.18: Fitted top-quark mass distribution of background events for the 8 TeV analysis with $|\eta^{PFjet}| < 1.3$. The distribution is fitted with a Landau function.

fitted Landau function provides a slightly better description of the shape compared to the standard $|\eta^{PFjet}| < 2.4$ analysis.

The calibration is performed in the same way as before, by performing pseudo-experiments with simulated events using the expected sample composition. The resulting biases and pull widths are shown in Figure 5.19. Significant residual biases are again observed, since

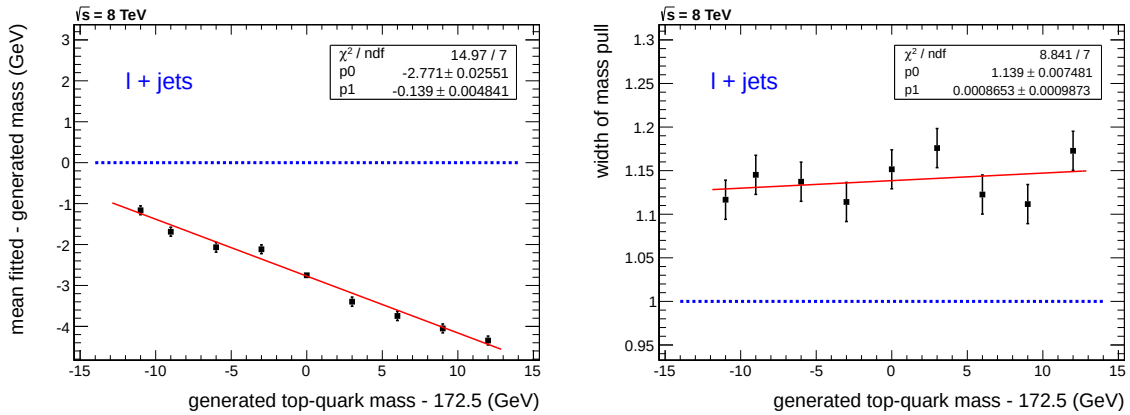

Figure 5.19: Bias on the estimated top-quark mass (left) and width of the pull distribution (right) as a function of the generated top-quark mass for the 8 TeV analysis with $|\eta^{PFjet}| < 1.3$.

the method is still relying on some assumptions and simplifications. When these linear calibration curves are applied to the estimated top-quark masses the residual biases disappear and the pull widths are also compatible with one, as shown in Figure 5.20. This bias-free method is then applied to the data to measure the top-quark mass.

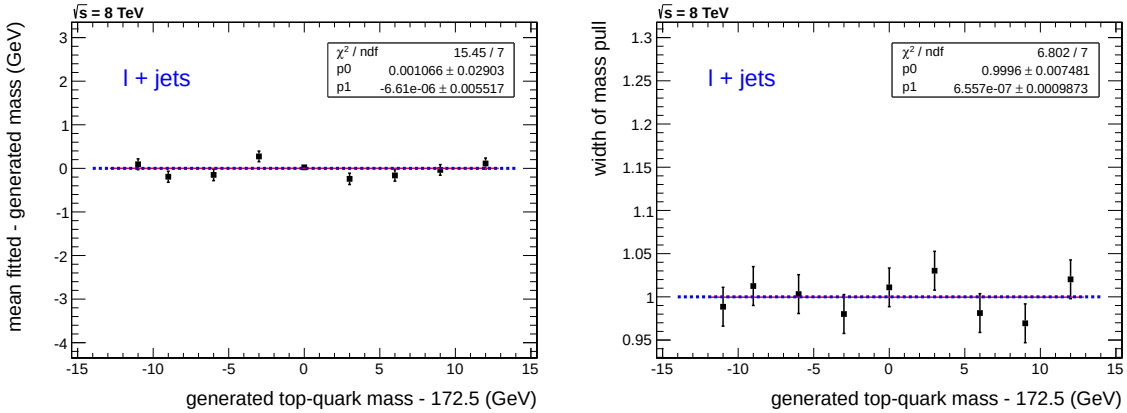

Figure 5.20: Bias on the estimated top-quark mass (left) and width of the pull distribution (right) after calibration as a function of the generated top-quark mass for the 8 TeV analysis with $|\eta^{PFjet}| < 1.3$.

5.4.3 Results and systematic uncertainties of the most optimal analysis

The $\Delta\chi^2_{\text{sample}}(x|m_t)$ curve which is obtained by applying the optimised analysis to the 8 TeV data is plotted in Figure 5.21, which is calculated before applying the calibration procedure. The top-quark masses are extracted from the minimum of this curve and the calibration

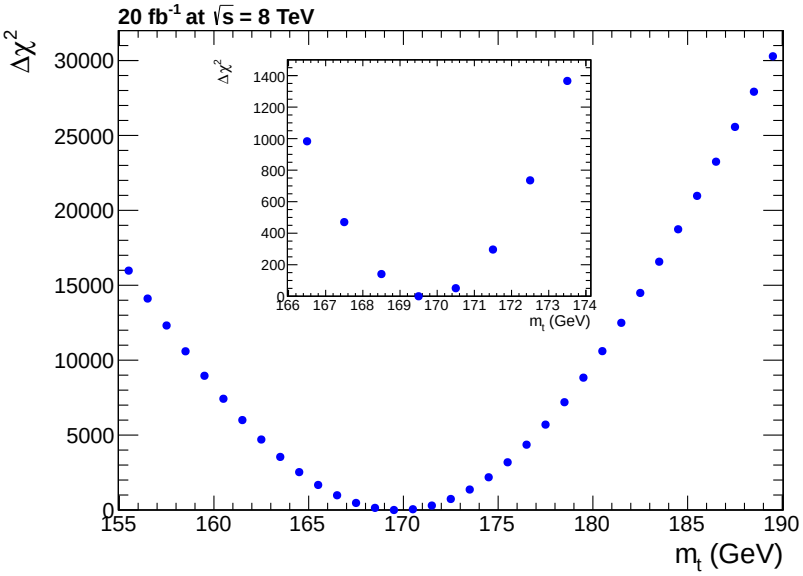

Figure 5.21: $\Delta\chi^2_{\text{sample}}(x|m_t)$ curve obtained with the 8 TeV dataset for the analysis with $|\eta^{PFjet}| < 1.3$.

procedure is applied. The results are shown in Table 5.6 for the different decay channels and for the combined ℓ+jets channel. These results are in general slightly smaller than the results obtained with the standard analysis with $|\eta^{PFjet}| < 2.4$ as shown in Table 5.1. However, these shifts are well within the jet energy scale systematic uncertainties, as can

	e+jets	μ+jets	ℓ+jets
8 TeV (19.7 fb^{-1})	172.65 $\pm$ 0.20 GeV	172.39 $\pm$ 0.18 GeV	172.51 $\pm$ 0.13 GeV

Table 5.6: Measured top-quark masses and their statistical uncertainties in the different decay channels and the combined ℓ+jets channel, using the 8 TeV datasets for the analysis with $|\eta^{PFjet}| < 1.3$.

be seen from the results presented below.

The evaluation of the systematic uncertainty is done in exactly the same way as explained in Section 5.3.2. An overview of all the systematic uncertainties is shown in Table 5.7. These

Source	Estimated effect on m_t (MeV)
Jet energy scale	**771** $\pm$ 7
Jet energy resolution	**31** $\pm$ 7
b-tagging efficiency	**18** $\pm$ 4
Pileup	**18.5** $\pm$ 0.2
Signal fraction	**266** $\pm$ 2
Background composition	**38.1** $\pm$ 0.3
Method calibration	23 $\pm$ **36**
Parton distribution functions	**159** $\pm$ 2
Q^2-scale; initial- and final-state radiation	**420** $\pm$ 125
ME-PS matching threshold	**205** $\pm$ 124
Colour reconnection	34 $\pm$ **77**
Underlying event	60 $\pm$ **124**
Total	967

Table 5.7: Overview of the systematic uncertainties on m_t as measured by the 8 TeV analysis with $|\eta^{PFjet}| < 1.3$. For each contribution the larger among the estimated shift and its statistical uncertainty is quoted, as indicated by the bold script.

results have a significant reduction of the total uncertainty compared to the 8 TeV results with $|\eta^{PFjet}| < 2.4$ from Table 5.2. This is mainly caused by the smaller jet energy scale and Q^2-scale systematic uncertainties, which are the result of the tighter $|\eta^{PFjet}|$ requirement. The breakdown of the systematic uncertainties on the background composition and the underlying event modelling is shown in Tables 5.8 and 5.9, respectively. In general a slight reduction is observed on the systematic uncertainties sensitive to the different backgrounds, since the background fraction is decreased when going to $|\eta^{PFjet}| < 1.3$. No large changes

Background	Effect on m_t (MeV)
W+jets ($\pm$30%)	7.2 $\pm$ 0.2
Z/γ^*+jets ($\pm$30%)	2.35 $\pm$ 0.02
Single top t-channel ($\pm$30%)	0.88 $\pm$ 0.01
Single top tW-channel ($\pm$30%)	21.3 $\pm$ 0.2
QCD multijet ($\pm$50%)	30.6 $\pm$ 0.1
Total	38.1 $\pm$ 0.3

Table 5.8: Overview of the subcomponents of the background composition systematic uncertainty for the 8 TeV analysis with $|\eta^{PFjet}| < 1.3$. The total uncertainty is calculated as the quadratic sum of all individual components.

PYTHIA tune	Effect on m_t (MeV)
Tune P11 vs. tune P11TeV	59 $\pm$ 88
Tune P11 vs. tune P11mpiHi	5 $\pm$ 88
Total	60 $\pm$ 124

Table 5.9: Overview of the subcomponents of the underlying event systematic uncertainty for the 8 TeV analysis with $|\eta^{PFjet}| < 1.3$. The total is calculated as the quadratic sum of the individual components.

in the theoretical systematic uncertainties are observed, except the decrease of the Q^2-scale uncertainty.

The combination of these systematic uncertainties with the results obtained in 19.7 fb^{-1} of 8 TeV data results in a measured top-quark mass of

$$m_t = 172.51 \pm 0.13\,(\text{stat.}) \pm 0.97\,(\text{syst.})\ \text{GeV}.$$

This result is in agreement with previous m_t measurements as discussed in Section 1.2.2. The total uncertainty is reduced by about 9% by going from $|\eta^{PFjet}| < 2.4$ to $|\eta^{PFjet}| < 1.3$.

5.5 Binned measurement of the top-quark mass

To study the description of the top-quark mass observable as measured with the Ideogram method in various regions of phase-space, a measurement of this quantity in bins of different kinematic variables is performed. These results can then be compared to the various simulated $t\bar{t}$ samples with modified underlying-event tunes or different generator-level settings like the amount of initial- and final-state radiation or the matching thresholds.

This measurement is performed with the same events as used for the most optimal measurement discussed in Section 5.4, which has kinematic requirements on the jets of $p_T > 30$ GeV and $|\eta| < 1.3$. Instead of using all jet combinations in an event and assigning them the weights from Equation 5.5, only the jet combination with the highest weight is used and all kinematic quantities are calculated from this particular jet combination. This also means that new Ideogram shapes and new calibration curves need to be calculated, as discussed in Section 5.5.1. The final results in bins of different kinematic quantities are presented in Section 5.5.2.

5.5.1 Ideogram shapes and calibration curves

The wrong-jet-combination and background shapes are determined as before. The fitted top-quark mass distributions of wrong jet combinations, together with the Landau functions which are fitted to these distributions, are shown in Figure 5.22 for two different generated top-quark masses. The dependence of the mean μ and the width σ on the generated top-quark mass is plotted in Figure 5.23. A linear behaviour is again observed and the fitted lines are used as a parametrisation of the linear dependence of μ and σ on the generated top-quark mass. The background shape used in the Ideogram method is also estimated in the same way, by fitting a Landau function to the fitted top-quark mass distribution of the main backgrounds: W+jets and QCD multijet events. This is shown in Figure 5.24. In general a reasonable description of all distributions by the fitted Landau functions is observed.

The biases and the widths of the pull distributions are estimated as before with pseudo-experiments and are plotted in Figure 5.25. A reduction of the biases and an increase in the width of the pull distributions is observed when compared to the calibration curves in Figure 5.19 where all jet combinations are used. Figure 5.26 shows that the calibration procedure is still able to correct the statistical uncertainties for these large observed pull widths. This figure also shows that no residual biases are present.

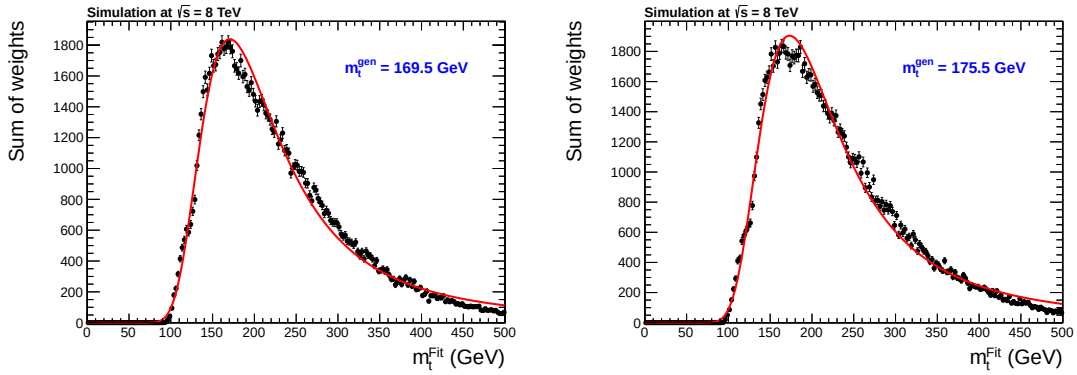

Figure 5.22: Fitted top-quark mass distributions of wrong jet combinations for different generated top-quark masses for the 8 TeV analysis with $|\eta^{PFjet}| < 1.3$, using only the highest-weight jet combination per event. Both distributions are fitted with a Landau function.

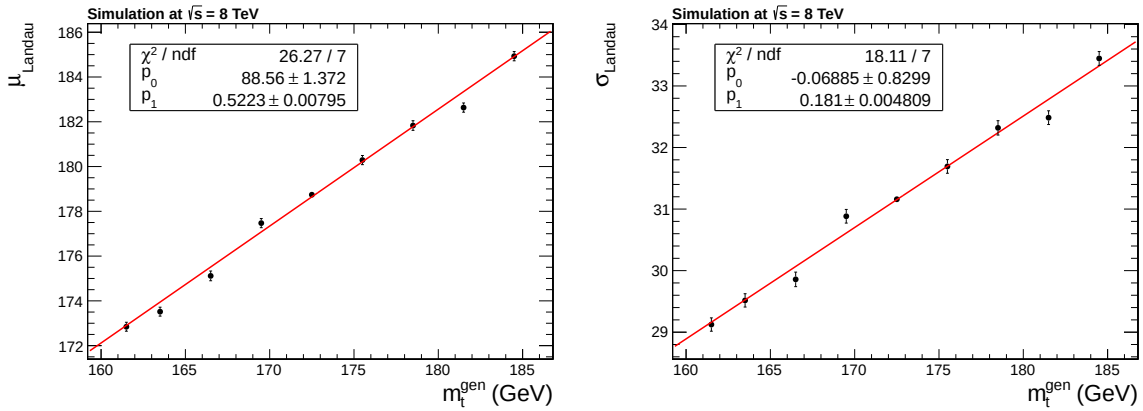

Figure 5.23: Dependence of the mean μ (left) and the width σ (right) of the fitted Landau function on the generated top-quark mass m_t^{gen} for the 8 TeV analysis with $|\eta^{PFjet}| < 1.3$, using only the highest-weight jet combination per event.

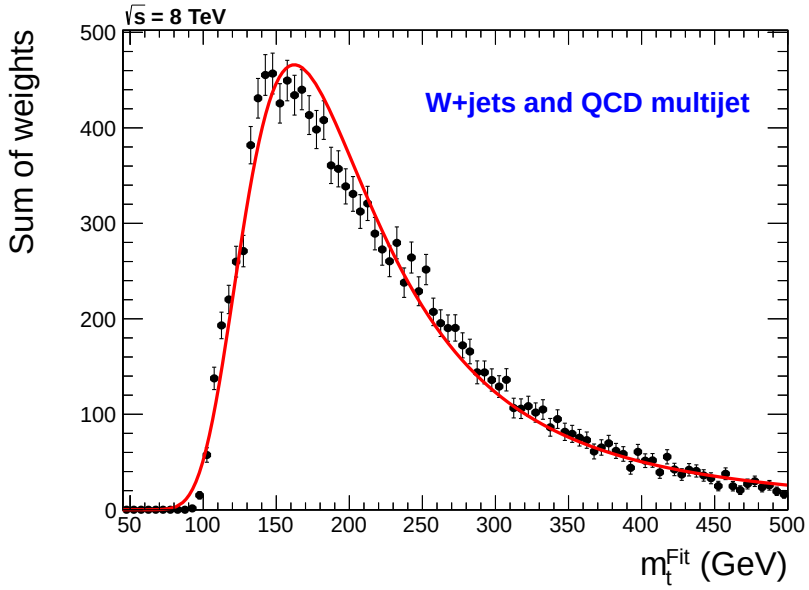

Figure 5.24: Fitted top-quark mass distribution of background events for the 8 TeV analysis with $|\eta^{PFjet}| < 1.3$, using only the highest-weight jet combination per event. The distribution is fitted with a Landau function.

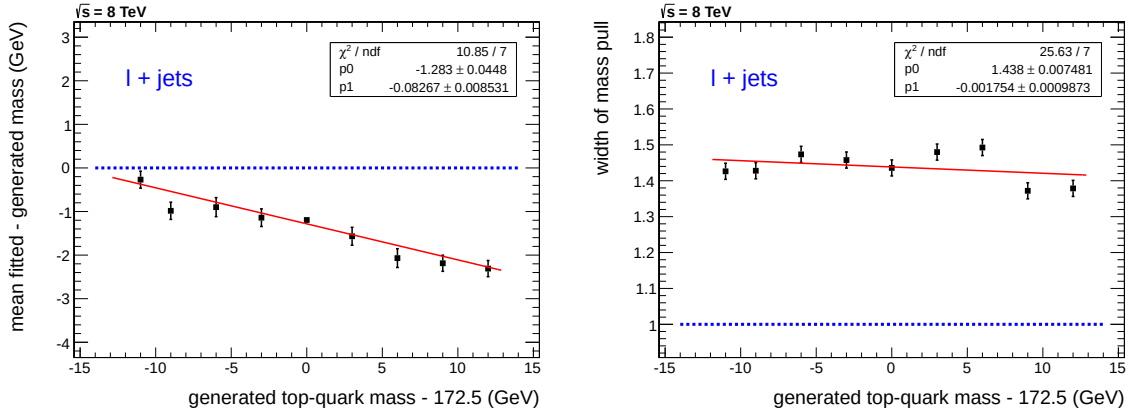

Figure 5.25: Bias on the estimated top-quark mass (left) and width of the pull distribution (right) as a function of the generated top-quark mass for the 8 TeV analysis with $|\eta^{PFjet}| < 1.3$, using only the highest-weight jet combination per event.

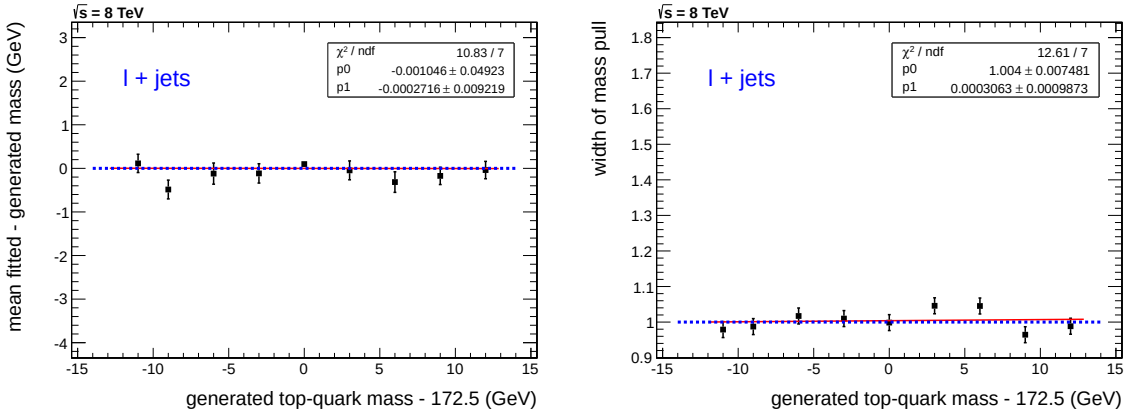

Figure 5.26: Bias on the estimated top-quark mass (left) and width of the pull distribution (right) after calibration as a function of the generated top-quark mass for the 8 TeV analysis with $|\eta^{PFjet}| < 1.3$, using only the highest-weight jet combination per event.

5.5.2 Results in bins of several kinematic variables

To obtain the final results the calibrated method explained in the previous section is applied to the data in bins of several kinematic variables. The choice of the binning was based on the requirement to have as many bins as possible, but with a statistical uncertainty on the measured top-quark mass in each bin below the total systematic uncertainty of ~ 1 GeV. In every bin the quantity $m_t - \langle m_t \rangle$ is plotted instead of the measured top-quark mass m_t, where $\langle m_t \rangle$ is the result of the inclusive measurement on the entire sample. This removes the constant shifts in m_t as expected due to the different systematic uncertainties and allows an easier comparison between the data and the different theoretical predictions.

Together with the statistical uncertainty, the systematic uncertainties on $m_t - \langle m_t \rangle$ related to the detector modelling and the backgrounds are also evaluated in each bin and are added in quadrature to the statistical uncertainties. The systematic uncertainties taken into account are the uncertainties related to the jet energy scale, the jet energy resolution, the b-tagging efficiency, additional pileup interactions, the signal fraction, and the background composition. These are evaluated in each bin in exactly the same way as explained before for the inclusive top-quark mass measurement.

The measured top-quark masses and their corresponding uncertainties are compared to the nominal tune Z2* different and to theoretical predictions, which correspond to variations in the Q^2-scale and the matching threshold, and variations in the underlying event and colour reconnection via the different P11 tunes. For each of these theoretical curves the same quantity $m_t - \langle m_t \rangle$ is estimated in each bin with the Ideogram method taking the relevant background into account, together with its statistical uncertainty which reflects the limited size of the simulated event samples.

The top-quark mass measured in bins of two global event variables, the missing transverse energy $\not{E}_T$ and the scalar sum of the p_T of all selected jets H_T, is shown in Figures 5.27 and 5.28. The measurement in bins of $\not{E}_T$ shows a good agreement between the data and all theoretical predictions. In the measurement in bins of H_T a good agreement is also observed, except for the Q^2-scale down variation which is slightly disfavoured by the data in the H_T region between 330 and 410 GeV.

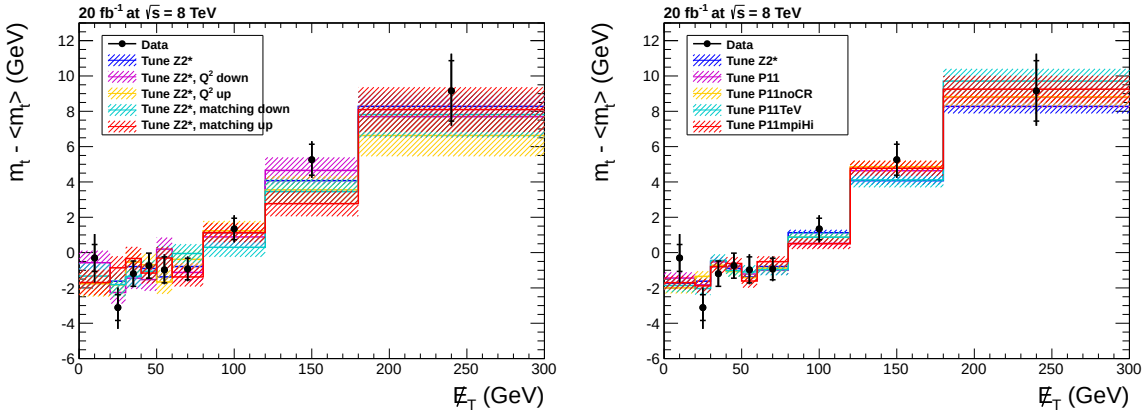

Figure 5.27: Measured top-quark mass as a function of the missing transverse energy, comparing the data with the Q^2-scale and matching variations (left), and with different PYTHIA tune variations (right).

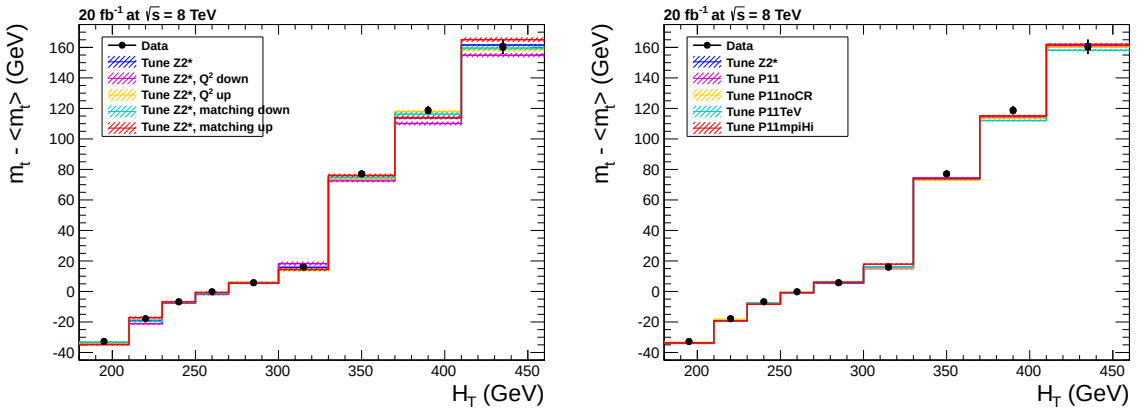

Figure 5.28: Measured top-quark mass as a function of the scalar sum of the p_T of all selected jets, comparing the data with the Q^2-scale and matching variations (left), and with different PYTHIA tune variations (right).

Figures 5.29 and 5.30 show the measured top-quark mass in bins of the number of selected jets and the number of b-tagged jets, respectively. In general a good agreement

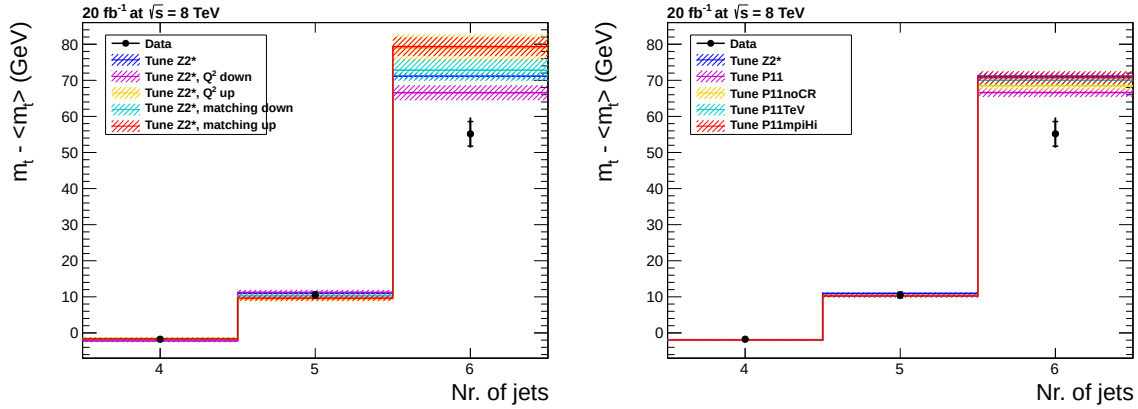

Figure 5.29: Measured top-quark mass as a function of the the number of selected jets, comparing the data with the Q^2-scale and matching variations (left), and with different PYTHIA tune variations (right).

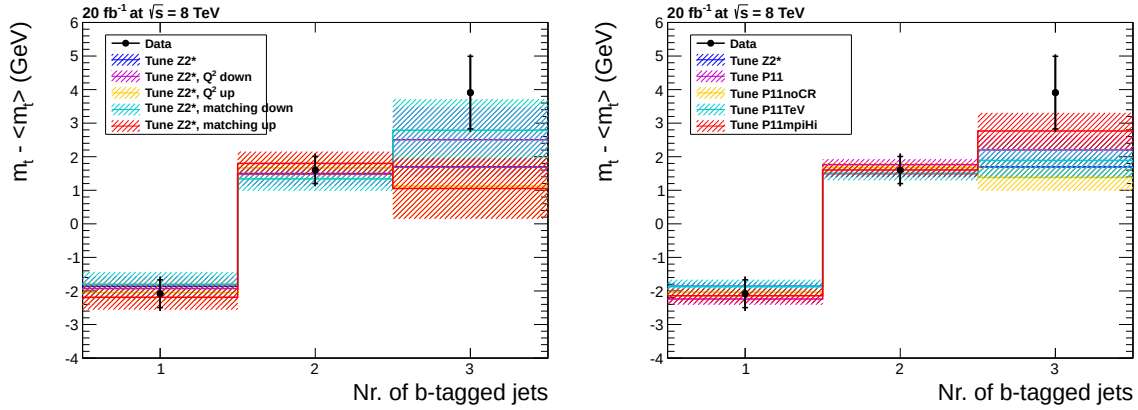

Figure 5.30: Measured top-quark mass as a function of the number of b-tagged jets, comparing the data with the Q^2-scale and matching variations (left), and with different PYTHIA tune variations (right).

between the data and the different theoretical curves is observed for both variables, except for event with 6 selected jets where the data is in disagreement with all theoretical predictions.

The measurement is also performed in bins of the kinematic properties of the $t\bar{t}$ system. The $t\bar{t}$ system is defined in this case by combining the four-momenta of the 4 leading jets in every event with that of the isolated lepton and with the energy and momentum of the neutrino. The p_x and p_y of the neutrino are taken to be equal to the x and y components of the missing transverse energy $\not\!\!E_T$, while the p_z component is calculated by imposing the mass of the W-boson as a constraint via $m_W^2 = (E_\ell + E_\nu)^2 - (\vec{p}_\ell + \vec{p}_\nu)^2$. The top-quark mass measurements in bins of the mass and transverse momentum of the $t\bar{t}$ system are shown in Figures 5.31 and 5.32. A good description of the data by most of the theoretical

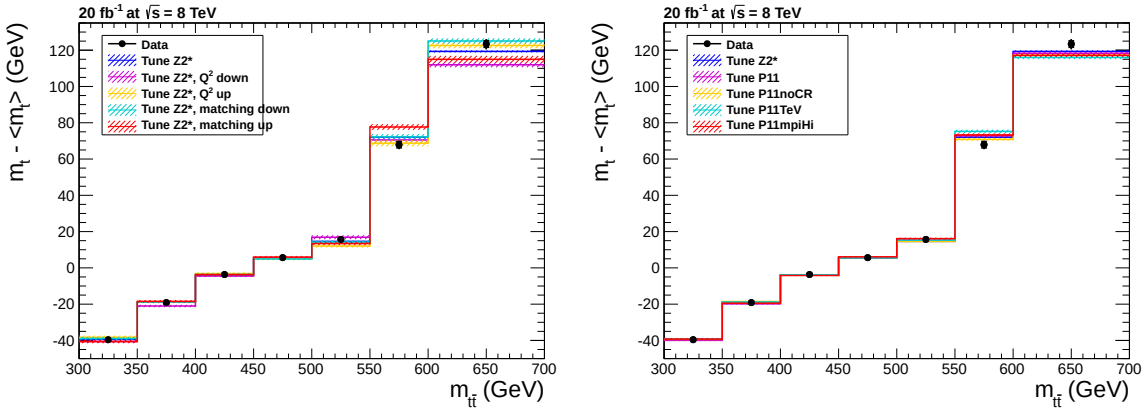

Figure 5.31: Measured top-quark mass as a function of the invariant mass of the $t\bar{t}$ system, comparing the data with the Q^2-scale and matching variations (left), and with different PYTHIA tune variations (right).

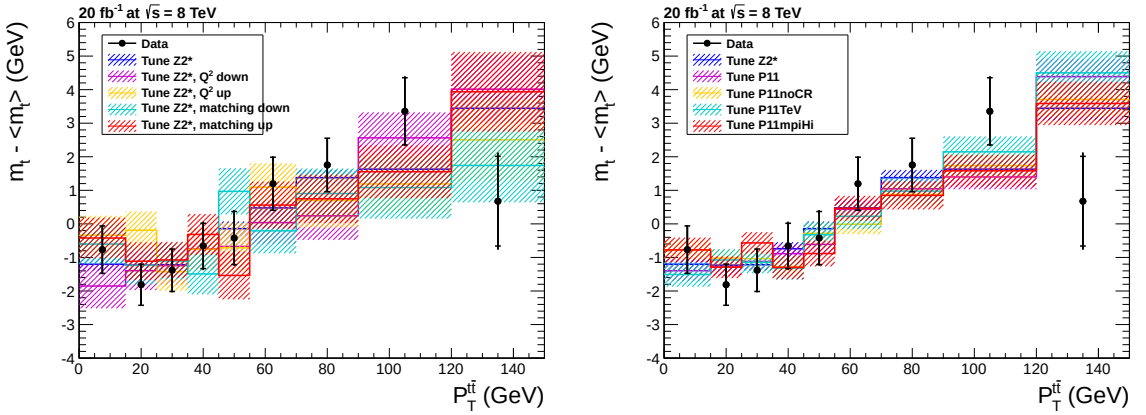

Figure 5.32: Measured top-quark mass as a function of the p_T of the $t\bar{t}$ system, comparing the data with the Q^2-scale and matching variations (left), and with different PYTHIA tune variations (right).

predictions is found for both variables, except at high $m_{t\bar{t}}$. Here some differences between the theoretical predictions are visible and the data is not in agreement with the Q^2-scale down and matching threshold up variations. In this region smaller disagreements are also observed between the data and the P11TeV and P11mpiHi PYTHIA tunes.

In Figures 5.33 and 5.34 the measured top-quark mass is shown in bins of the p_T and $|\eta|$ of the hadronic top quark, which is defined via the combination of three jets with the highest weight as used in the Ideogram method for these binned measurements. A good

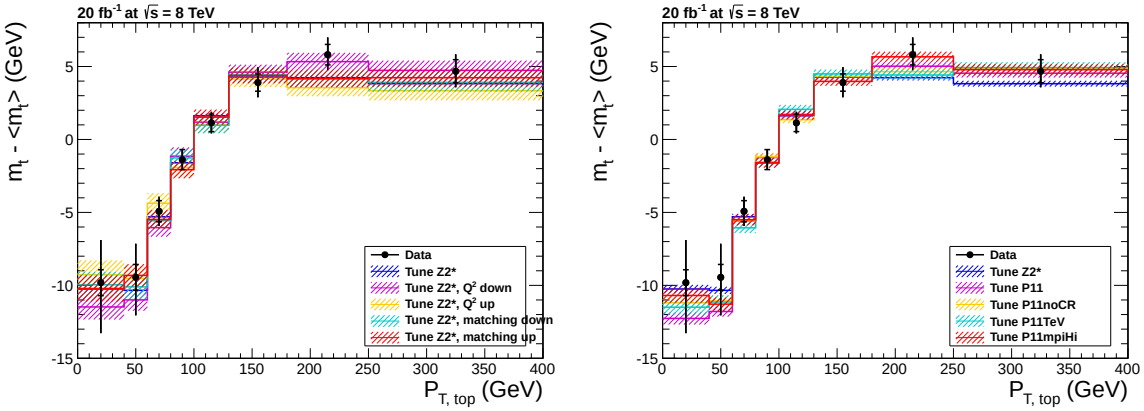

Figure 5.33: Measured top-quark mass as a function of the p_T of the hadronic top quark, comparing the data with the Q^2-scale and matching variations (left), and with different PYTHIA tune variations (right).

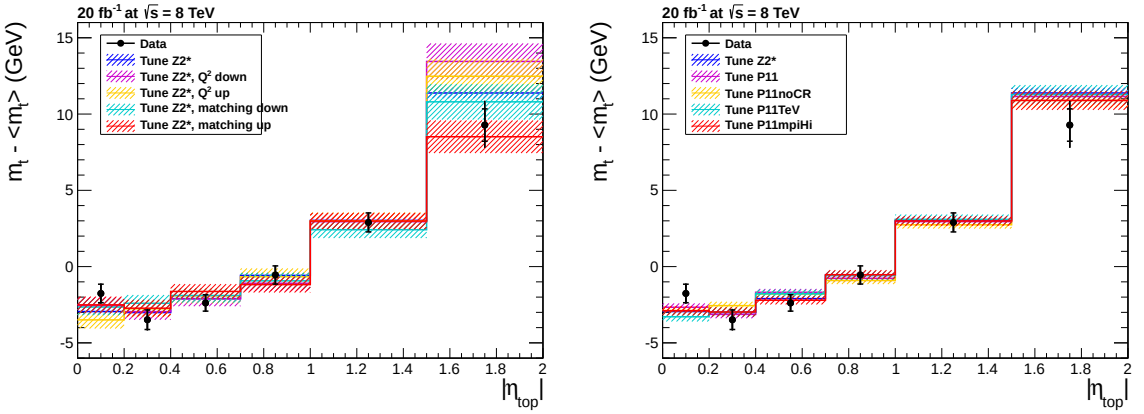

Figure 5.34: Measured top-quark mass as a function of the pseudorapidity of the hadronic top quark, comparing the data with the Q^2-scale and matching variations (left), and with different PYTHIA tune variations (right).

agreement between the measured top-quark masses in data and the theoretical predictions is found.

The dependence of the top-quark mass on the transverse momentum and pseudorapidity of the hadronic b jet of the jet combination with the highest weight is shown in Figures 5.35 and 5.36. Good agreement is again observed between the data and the different theoretical curves. Except at very high transverse momentum of the hadronic b-jet, where the data disfavours the Q^2-scale down variation.

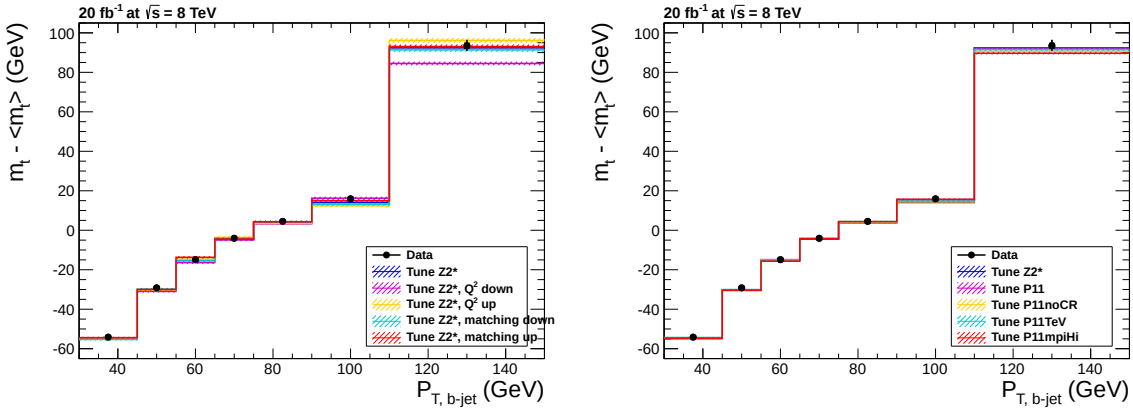

Figure 5.35: Measured top-quark mass as a function of the p_T of the hadronic b jet, comparing the data with the Q^2-scale and matching variations (left), and with different PYTHIA tune variations (right).

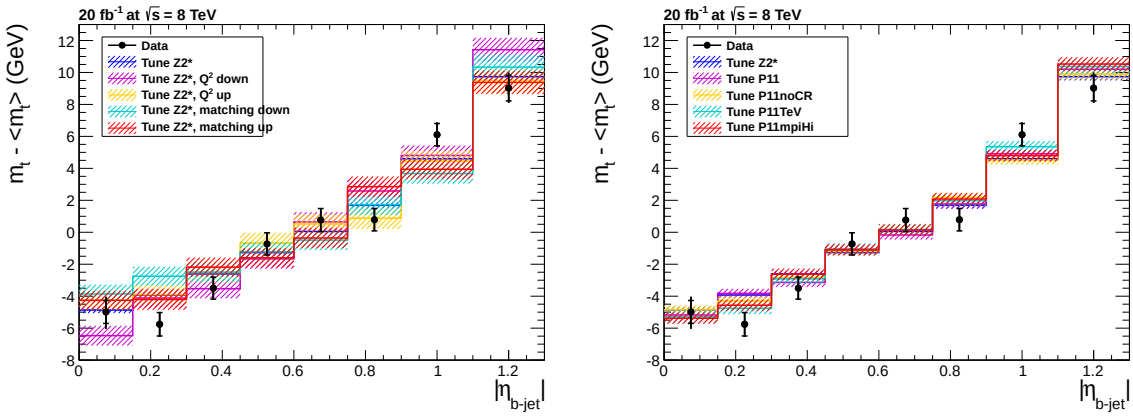

Figure 5.36: Measured top-quark mass as a function of the pseudorapidity of the hadronic b jet, comparing the data with the Q^2-scale and matching variations (left), and with different PYTHIA tune variations (right).

Finally the top-quark mass is measured in bins of the ΔR distances between the decay products of the hadronic top quark. Figure 5.37 shows the top-quark mass in bins of the smallest ΔR distance between any of the two light jets and the hadronic b jet, while in Figure 5.38 the measurement in bins of the ΔR distance between the two light jets is plotted. In most bins a good agreement between the data and the theoretical curves is observed. At

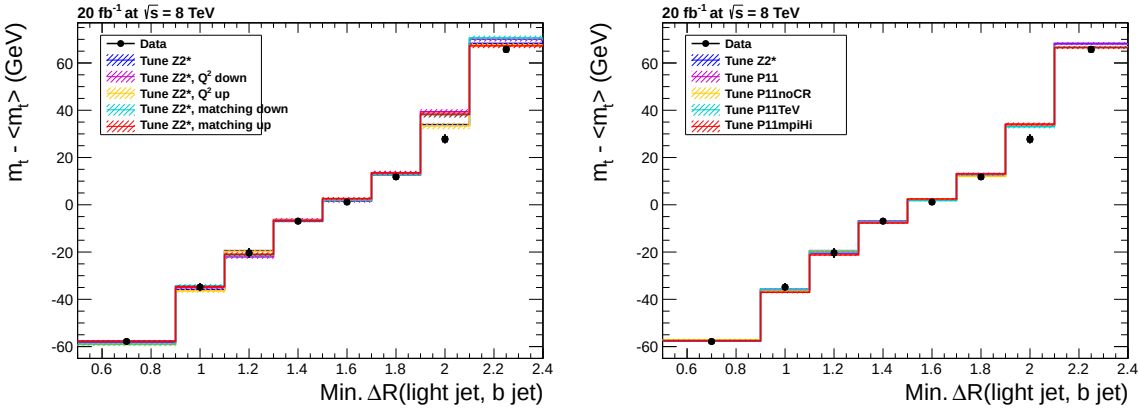

Figure 5.37: Measured top-quark mass as a function of the smallest ΔR distance between the hadronic b jet and one of the two light jets, comparing the data with the Q^2-scale and matching variations (left), and with different PYTHIA tune variations (right).

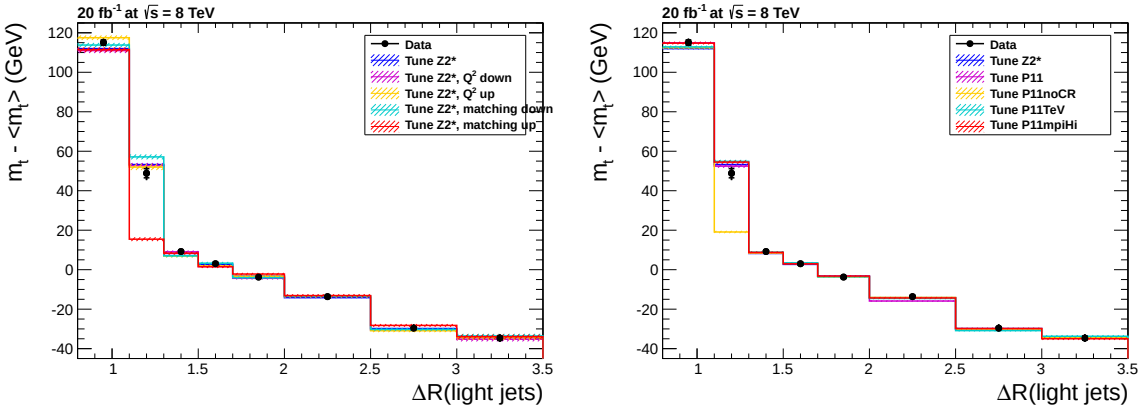

Figure 5.38: Measured top-quark mass as a function of the ΔR distance between the two light jets, comparing the data with the Q^2-scale and matching variations (left), and with different PYTHIA tune variations (right).

high values of the smallest ΔR distance between any of the two light jets and the hadronic b jet some disagreement between the data and the predictions is observed. Here the largest disagreement occurs between the data and both matching threshold variations and the Q^2-scale down variation in the 1.9 – 2.1 bin. Significant disagreement is also observed at low values of the ΔR distance between the two light jets, mainly in the 1.1 – 1.3 bin. Here a clear disagreement with both matching variations and with the P11noCR PYTHIA tune is observed.

Chapter 6

Measurement of the mass difference between the top and the antitop quark

The CPT symmetry of a quantum field theory like the Standard Model can be experimentally verified by measuring the difference in mass between a particle and its corresponding antiparticle, as discussed in Section 1.1.4. Most quarks cannot be observed as free quarks, since they carry colour charge and hadronise into colourless particles before decaying, meaning that a direct measurement of this mass difference is not possible. The lone exception is the top quark, which due to its short lifetime decays before hadronisation can take place. Therefore top quarks are used to measure this mass difference ($\Delta m_t = m_t - m_{\bar{t}}$), which serves a test of CPT symmetry.

The mass difference between the top quark and its antiquark was measured previously by other experiments, as presented in Section 1.2.2, and showed no significant deviation from zero. In this thesis both the 7 and 8 TeV datasets collected by the CMS experiment are used to measured this quantity, by using a technique based on the Ideogram method as explained in Section 6.1. The results obtained with this method when applied to the data are presented in Section 6.2 and the corresponding systematic uncertainties are discussed in Section 6.3.

6.1 Procedure to measure the top-antitop mass difference

The Ideogram method, which was used in Chapter 5 to measure the top-quark mass, is also used here to measure the mass difference between the top and the antitop quark using the measurement technique discussed in Section 6.1.1. The closure and residual biases of this method were also checked and will be discussed in Section 6.1.2.

6.1.1 Measuring Δm_t with the Ideogram method

This measurement starts from exactly the same event selection as presented in Chapter 4. However, the entire dataset is split via the lepton charge into ℓ^- and ℓ^+ samples. The

ℓ^-+jets sample will contain mainly hadronically decaying top quarks ($t\bar{t} \to bW^+\bar{b}W^- \to bq\bar{q}'\bar{b}\ell^-\nu_\ell$), while the ℓ^++jets sample will consist mainly of hadronically decaying antitop quarks ($t\bar{t} \to bW^+\bar{b}W^- \to b\ell^+\bar{\nu}_\ell\bar{b}\bar{q}q'$). For each event category the mass of the hadronically decaying top or antitop quark is measured, from which the mass difference is obtained via

$$\Delta m_t = m^{\ell^-} - m^{\ell^+}. \tag{6.1}$$

Here m^{ℓ^-} and m^{ℓ^+} are the top-quark masses as measured in the ℓ^-+jets and ℓ^++jets samples, respectively. The statistical uncertainty on Δm_t is calculated by taking the quadratic sum of the statistical uncertainties on the individual m^{ℓ^-} and m^{ℓ^+} measurements.

The individual top-quark masses in the ℓ^-+jets and ℓ^++jets samples, m^{ℓ^-} and m^{ℓ^+}, are measured with the same method as used for the top-quark mass measurement in Chapter 5. Exactly the same kinematic fit is applied with the W-boson mass fixed to 80.4 GeV, implying that the W^+ and W^- bosons are assumed to have the same mass. Their mass difference has been measured to be compatible with zero within an accuracy of 0.6 GeV [9]. The output of the kinematic fit of all jet combinations passing the standard selection requirements from Section 4.2.2 is again used as input to the Ideogram method.

This Ideogram method is applied separately to both the ℓ^-+jets and ℓ^++jets data samples to obtain the mass of the hadronically decaying top quarks in each sample, m^{ℓ^-} and m^{ℓ^+}. Exactly the same method as for the top-quark mass measurement is used, hence also the same parametrisations and inclusive ℓ+jets calibration curves as discussed in Sections 5.1 and 5.2. The performance of this approach in measuring a possible mass difference between the top and the antitop quark will be discussed below.

6.1.2 Performance of the method

Since identical parametrisations and calibration curves as for the top-quark mass measurement are used to estimate both the m^{ℓ^-} and m^{ℓ^+} masses, the residual biases on both masses need to be studied. Therefore the standard calibration procedure with pseudo-experiments is used to estimate the biases and the widths of the pull distribution after the application of the inclusive ℓ+jets calibration.

The residual biases and pull widths after calibration are shown separately for ℓ^++jets and ℓ^-+jets events in Figures 6.1 and 6.2 for the 7 and 8 TeV analyses, respectively. From the upper plots in these figures it is clear that no statistically significant residual biases or residual slopes exist in the separate m^{ℓ^-} and m^{ℓ^+} measurements after the inclusive ℓ+jets calibration, for both the 7 and the 8 TeV analyses. This means that the inclusive $\ell + jets$ calibration can be safely applied to both ℓ^-+jets and ℓ^++jets events. All pull distribution widths are also compatible with one, meaning that the statistical uncertainties on both masses are properly estimated.

An additional cross-check of the method is conducted where the measured difference in mass between the top and the antitop quark is compared to the mass difference at generator level. In principle this can be performed by generating several additional $t\bar{t}$ samples, where each sample has a certain generator-level difference in mass between the top and antitop quark. An easier option which does not involve the generation of additional samples is to use a $t\bar{t}$ sample where the top quarks are generated with a non-zero mass width, as shown in the left plot of Figure 6.3 for the nominal 8 TeV $t\bar{t}$ sample used in this thesis. In such

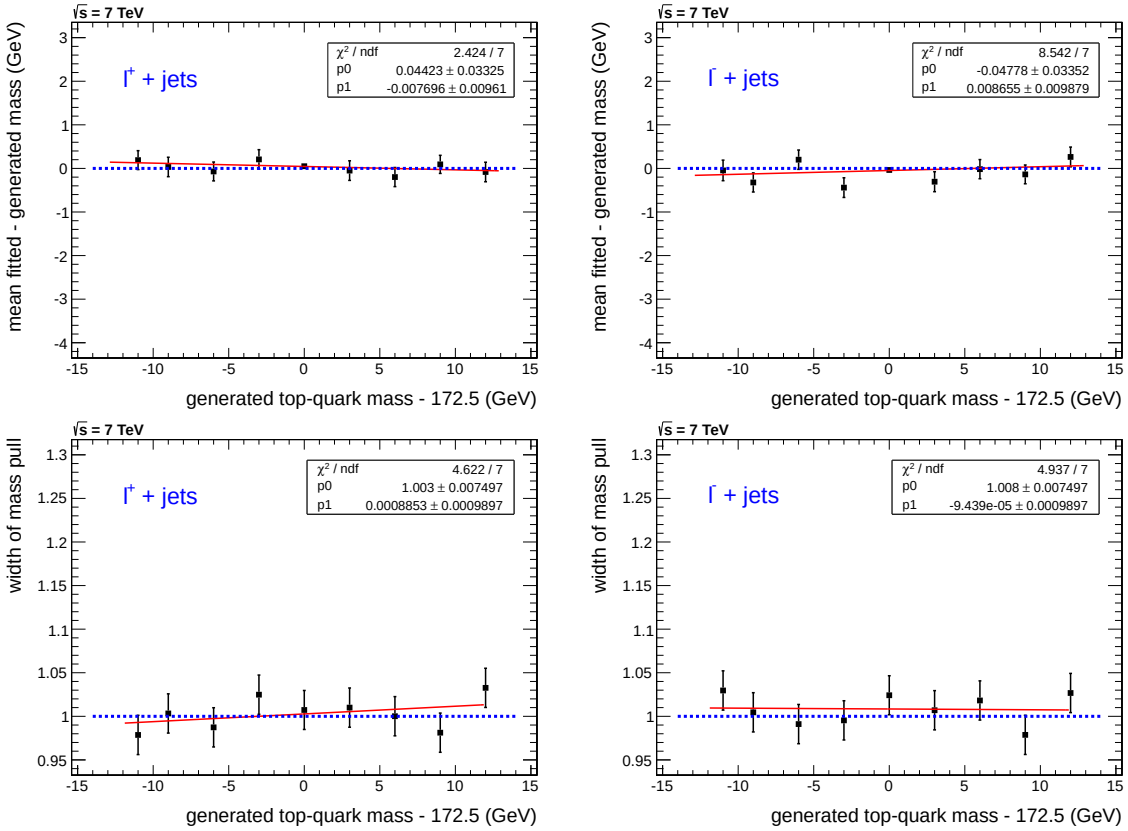

Figure 6.1: Bias on the estimated top-quark mass (top) and width of the pull distribution (bottom) as a function of the generated top quark mass with the 7 TeV analysis, using ℓ^++jets events (left) and ℓ^-+jets events (right) after the inclusive ℓ+jets calibration from Figure 5.8.

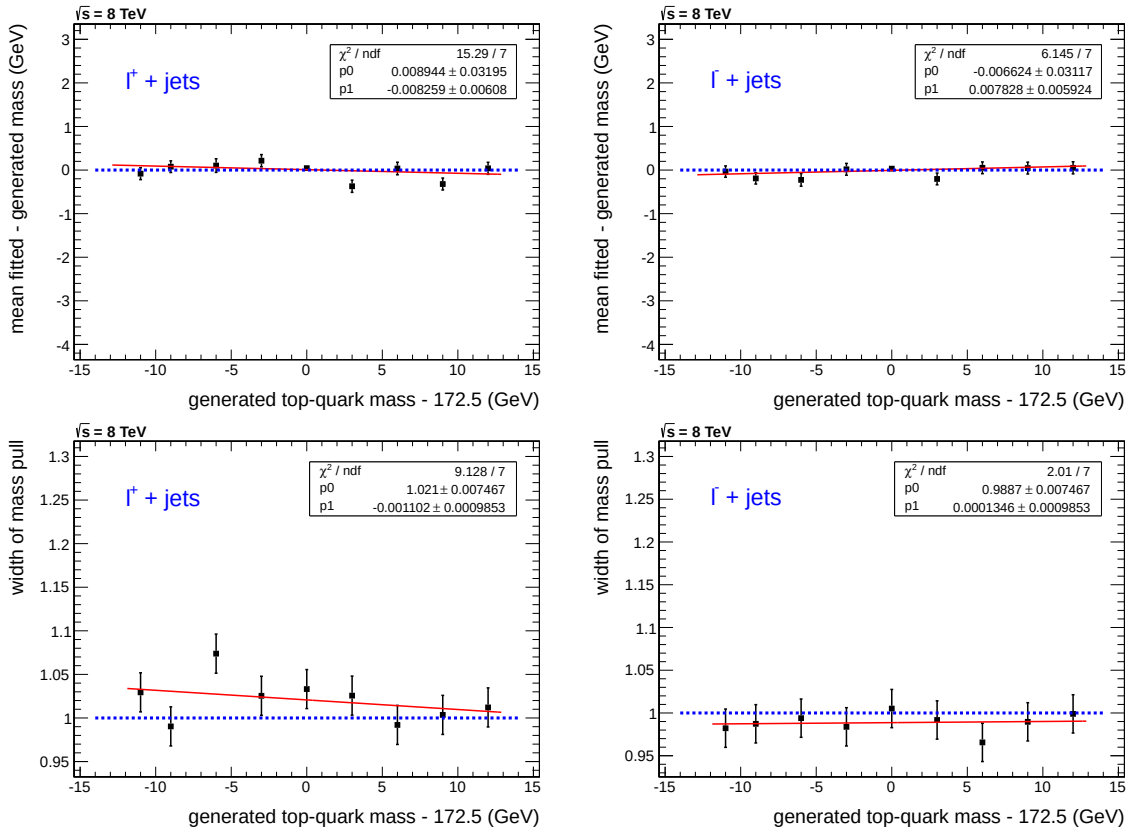

Figure 6.2: Bias on the estimated top-quark mass (top) and width of the pull distribution (bottom) as a function of the generated top quark mass with the 8 TeV analysis, using ℓ^++jets events (left) and ℓ^-+jets events (right) after the inclusive ℓ+jets calibration from Figure 5.9.

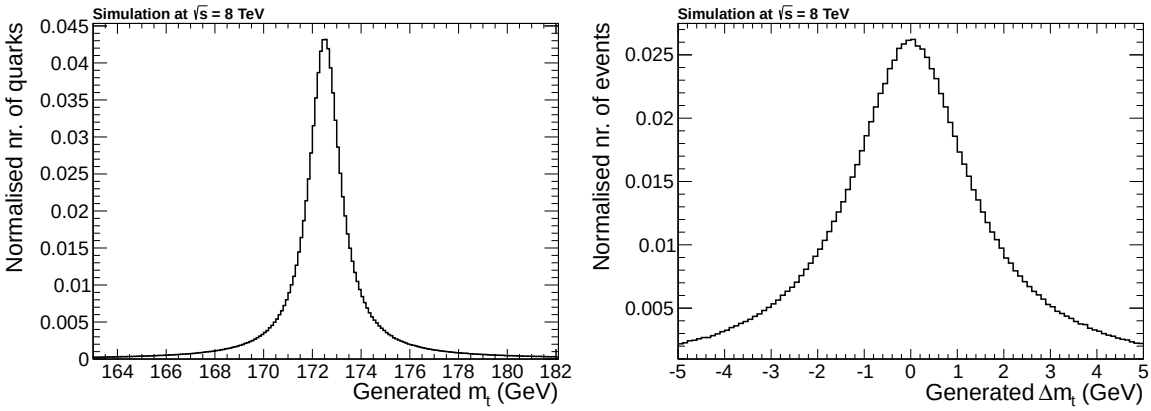

Figure 6.3: Distributions of the generated top-quark mass (left) and the generator-level mass difference between the top and the antitop quark (right) for $t\bar{t}$ ℓ+jets events at 8 TeV.

a sample the top and antitop quarks can already have a generator-level mass difference, as shown for the nominal 8 TeV $t\bar{t}$ sample in the left plot of Figure 6.3. This plot clearly shows that large enough mass differences are present at generator level.

The difference in mass between the top and the antitop quark is then measured in several bins of the generator-level mass difference using the method explained before and taking the standard backgrounds into account. The resulting dependence of the measured Δm_t on the generated Δm_t is plotted in Figure 6.4. This clearly shows that no residual biases

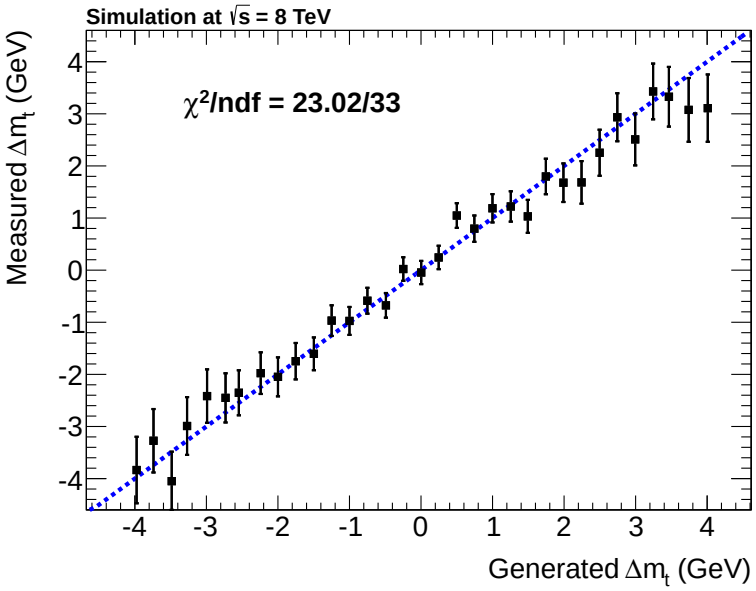

Figure 6.4: Measured Δm_t as a function of the generator-level Δm_t. The χ^2/ndf value is calculated with respect to the blue curve, which corresponds to the case where the measured Δm_t is equal to the generator-level Δm_t.

are present in the method since a reasonable χ^2/ndf of the measured points with respect to the blue curve is obtained, which corresponds to the case where the measured Δm_t is equal to the generator-level Δm_t. The generated mass width in the 7 TeV $t\bar{t}$ samples is equal to zero, therefore this cross-check can only be performed for the 8 TeV analysis.

6.2 Results on data

Since the method is free of any residual biases it can be applied to both the 7 and 8 TeV datasets. The resulting $\Delta\chi^2_{\text{sample}}(x|m_t)$ curves are plotted in Figures 6.5 and 6.6 for the ℓ^-+jets and ℓ^++jets samples. From the minima of these curves the top-quark masses and the corresponding statistical uncertainties in the ℓ^-+jets and ℓ^++jets channels are estimated. The inclusive ℓ+jets calibration procedure is applied and the resulting difference in mass between the top and the antitop quark is calculated, as shown in Table 6.1.

Like in the top-quark mass measurement a good agreement between the 7 and the 8 TeV results is found within statistical uncertainties. The separate Δm_t results in the e+jets and μ+jets channels are also compatible. The μ+jets channel tends to give slightly higher Δm_t values, but this is not significant when the statistical uncertainties are taken into account.

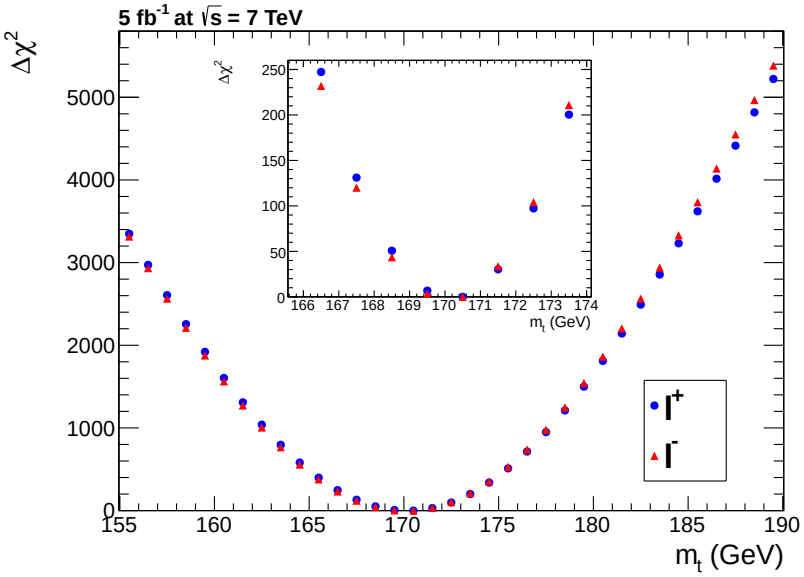

Figure 6.5: $\Delta\chi^2_{\text{sample}}(x|m_t)$ curve obtained with the 7 TeV dataset in the ℓ^-+jets and ℓ^++jets channels.

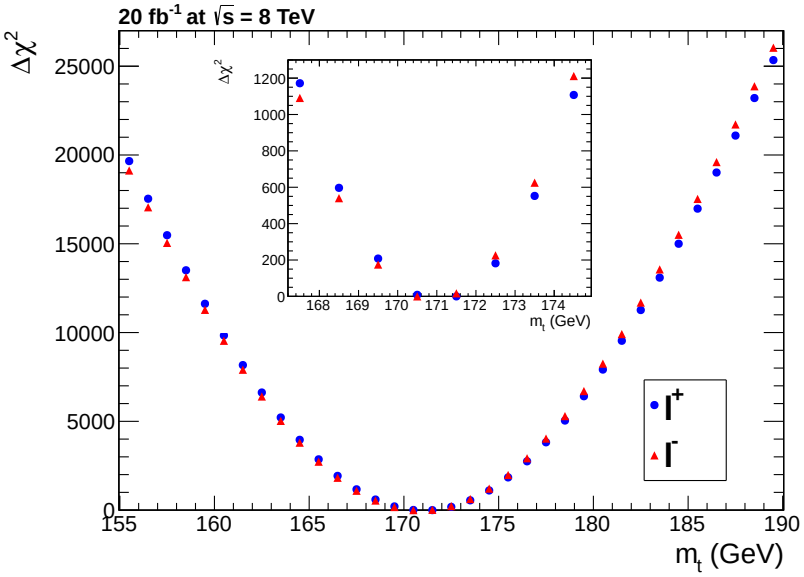

Figure 6.6: $\Delta\chi^2_{\text{sample}}(x|m_t)$ curve obtained with the 8 TeV dataset in the ℓ^-+jets and ℓ^++jets channels.

	e+jets	μ+jets	ℓ+jets
7 TeV (4.9 fb^{-1})	-763 $\pm$ 635 MeV	376 $\pm$ 560 MeV	-106 $\pm$ 420 MeV
8 TeV (19.7 fb^{-1})	-189 $\pm$ 285 MeV	-133 $\pm$ 258 MeV	-155 $\pm$ 191 MeV

Table 6.1: Measured mass differences between the top and the antitop quark and their statistical uncertainties in the different decay channels and the combined ℓ+jets channel, using the 7 and 8 TeV datasets.

All measured values are also compatible with the CPT hypothesis of $\Delta m_t = 0$, even without taking systematic uncertainties into account.

6.3 Systematic uncertainties

Many of the systematic uncertainties that affect the top-quark mass measurement as discussed in Section 5.3.2 have a significantly reduced impact in the measurement of the mass difference since they alter the top and antitop quarks in the same way. Some of these systematic uncertainties related to the modelling of the physical processes are not expected to affect the Δm_t measurement, thus they are not considered in this analysis. These include the uncertainties on the modelling of hadronisation, the underlying event, colour reconnection, Q^2-scale and initial- and final-state radiation, and the matching between the matrix element and the parton showers.

The remaining effects which were considered in the top-quark mass measurement are included, together with some additional sources which are potentially relevant for the Δm_t measurement such as lepton-charge identification and a possible difference in jet energy response between b and $\bar{b}$ jets. An overview of all the systematic effects considered in this analysis is given in Table 6.3. These numbers are the maximal observed shifts when the systematic effect is varied by ± 1 standard deviation in simulation. The corresponding statistical uncertainties on these shifts are calculated using the Jackknife method, as explained in Section 5.3.2.

The statistical significance of the observed shift in Δm_t is sometimes small. Therefore the observed shift is quoted as systematic uncertainty when it is larger than the statistical uncertainty and otherwise just the statistical uncertainty is quoted, as indicated by the bold script in the table. The total systematic uncertainty is calculated via the quadratic sum of the values quoted for each source. The evaluation of all these systematic effects is explained below, together with some potential additional sources, such as the uncertainty on the modelling of the trigger and on the assignment of the lepton charge.

Jet energy scale

Top and antitop quarks are produced at the LHC with slightly different rapidity distributions, therefore the η-dependence of the jet energy scale uncertainty can lead to an effect on Δm_t. This is evaluated by scaling the energy of all jets in simulation up/down within the p_T- and η-dependent uncertainties from Figures 3.11 and 3.12.

	Estimated effect on Δm_t (MeV)	
Source	7 TeV	8 TeV
Jet energy scale	$11 \pm \mathbf{12}$	$\mathbf{14} \pm 11$
b vs. $\bar{b}$ jet response	$\mathbf{50} \pm 1$	$\mathbf{51} \pm 1$
Jet energy resolution	$\mathbf{30} \pm 10$	$10 \pm \mathbf{11}$
b-tagging efficiency	$\mathbf{42} \pm 8$	$\mathbf{24} \pm 7$
b vs. $\bar{b}$ tagging efficiency	$\mathbf{49} \pm 6$	$\mathbf{11} \pm 7$
Pileup	$\mathbf{5.3} \pm 0.2$	$\mathbf{9.1} \pm 0.3$
Signal fraction	$\mathbf{49} \pm 1$	$\mathbf{27} \pm 2$
Background composition	$\mathbf{8} \pm 1$	$\mathbf{28} \pm 1$
Background charge asymmetry	$\mathbf{9.21} \pm 0.02$	$\mathbf{11.86} \pm 0.03$
Method calibration	$\mathbf{82} \pm 48$	$3 \pm \mathbf{53}$
Parton distribution functions	$\mathbf{7} \pm 3$	$\mathbf{9} \pm 3$
Total	131	91

Table 6.2: Overview of the systematic uncertainties on Δm_{t}. For each contribution the larger among the estimated shift and its statistical uncertainty is quoted, as indicated by the bold script.

b vs. $\bar{b}$ jet response

Since a difference in jet response between b and $\bar{b}$ jets can introduce a bias in this measurement, a dedicated study of this possible response difference was performed. The E_T of the reconstructed PF jets was compared with the original parton E_T in two simulated $t\bar{t}$ samples: the nominal sample which was generated with MadGraph and showered with pythia, and a $t\bar{t}$ sample which was generated with mc@nlo and showered with herwig. After the application of the L5L7 jet energy corrections from Section 4.2.1 the ratio of b to $\bar{b}$ jet response was measured in both samples in several bins of the transverse momentum of the PF jet. Finally, the difference between both response ratios was calculated as shown in Figure 6.7. The observed differences are in general compatible with or close to zero. Also the average difference, which is estimated by fitting a horizontal line to all the points, is close to zero.

In the 8 TeV analysis both the p_T-dependent difference and the average difference of $0.0779 \pm 0.0396\%$ are propagated, resulting in Δm_t shifts of 38 ± 1 MeV and 51 ± 1 MeV, respectively. The propagation of the average difference results in the largest shift, which is quoted a systematic uncertainty. In the 7 TeV analysis the same average difference of $0.0779 \pm 0.0396\%$ is propagated to Δm_t, since this

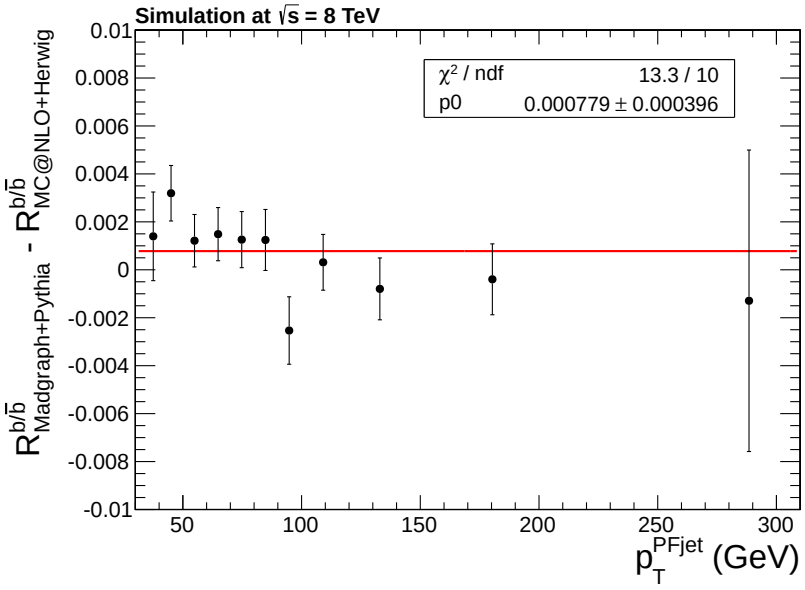

Figure 6.7: Difference between MADGRAPH+PYTHIA and MC@NLO+HERWIG of the b to $\bar{b}$ jet E_T response in simulated $t\bar{t}$ events at 8 TeV. The red horizontal line is fitted to all the points.

difference in jet response is not expected to depend on the centre-of-mass energy.

Jet energy resolution

To evaluate this uncertainty, the energy of each jet in simulation was smeared up/down within the jet energy resolution uncertainties shown in Figure 3.13. These uncertainties are $|\eta|$-dependent, so a residual effect on Δm_t is possible since signal and background have different pseudorapidity distributions.

***b*-tagging efficiency and *b* vs. $\bar{b}$ tagging efficiency**

The Ideogram method as used for this analysis utilises the b-tagging assignments in each event, so a possible mismodelling in simulation of the b-tagging efficiency can bias the measurement. The impact of this mismodelling is quantified by changing the b-tagging efficiency in simulation via the variation of the threshold on the b-tag discriminator.

The *b-tagging efficiency* systematic uncertainty is estimated by introducing absolute b-tagging efficiency changes of $\pm 1.9\%$ and $\pm 1.2\%$ in the 7 and the 8 TeV analyses, respectively. These numbers correspond to the uncertainties quoted on the b-tag efficiencies in Section 3.3.5. For the *b vs. $\bar{b}$ tagging efficiency* different working points for the ℓ^-+jets and ℓ^++jets samples are used, yielding an absolute difference in b-tagging efficiency between both samples of 1.9% and 1.2% in the 7 and the 8 TeV analyses.

Pileup

The effect of pileup on the Δm_t measurement is estimated in the same way as for the top-quark mass measurement. The mean number of interactions is changed in simulation by $\pm 5\%$ for the 7 TeV analysis and by $\pm 6\%$ for the 8 TeV analysis. The dependence of the measurement on the primary vertex multiplicity is again

small, as shown in Figure 6.8. The χ^2/ndf values show that the results are largely independent from the primary vertex multiplicity.

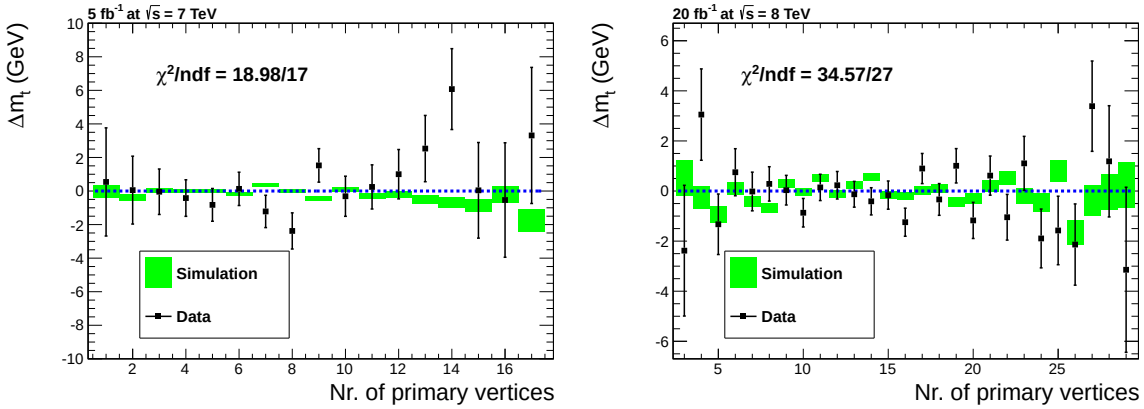

Figure 6.8: Dependence of the mass difference between the top and the antitop quark on the number of primary vertices in data and simulation for the 7 TeV (left) and 8 TeV (right) analyses. The χ^2/ndf value is calculated for the data with respect to the blue curve, which corresponds to $\Delta m_t = 0$.

Signal fraction

The ℓ^-+jets and ℓ^++jets sample have a different signal fraction, which means that a change in the global signal fraction can influence each sample in a different way. The systematic uncertainty is estimated by varying the signal fraction in simulation by a relative $\pm 10\%$ in both analyses.

Background composition

Possible residual effects due to the background composition are evaluated by scaling each background up and down, but keeping the signal fraction fixed. Like in the top-quark mass measurement, the W+jets, Z/γ^*+jets, single top t-channel and single top tW-channel backgrounds are each scaled independently by a relative $\pm 30\%$ while the QCD multijet background in the 8 TeV analysis is scaled by a relative $\pm 50\%$. The Δm_t shifts observed when separately scaling each background source are listed in Table 6.3.

Background charge asymmetry

A difference in the charge asymmetry of the backgrounds leads to different background levels and to a different background composition in the ℓ^-+jets and ℓ^++jets samples, which can bias the Δm_t measurement. The inclusive W^+/W^- ratio at 7 and 8 TeV is in agreement with theoretical predictions within precisions of 3.5% and 2%, respectively [130, 131], but since this ratio depends on the number of jets conservative uncertainties of 7% and 4% are used. The systematic uncertainty is estimated by varying W^+ and W^- events in opposite directions, thereby affecting the W^+/W^- ratio.

The W+jets background contains non-negligible contributions from $W+c\bar{c}$ and $W+b\bar{b}$ events, whose relative W^+ to W^- ratio is affected by larger uncertainties. Therefore a relative $\pm 20\%$ change in this ratio is applied separately for $W+c\bar{c}$ and $W+b\bar{b}$ events.

	Effect on Δm_t (MeV)	
Background	7 TeV	8 TeV
W+jets ($\pm$30%)	3.6 $\pm$ 0.5	1.3 $\pm$ 0.3
Z/γ^*+jets ($\pm$30%)	0.87 $\pm$ 0.04	1.99 $\pm$ 0.03
Single top t-channel ($\pm$30%)	1.60 $\pm$ 0.03	6.9 $\pm$ 0.1
Single top tW-channel ($\pm$30%)	7.3 $\pm$ 0.3	1.4 $\pm$ 0.3
QCD multijet ($\pm$50%)	/	26.8 $\pm$ 0.3
Total	8 $\pm$ 1	28 $\pm$ 1

Table 6.3: Overview of the subcomponents of the background composition systematic uncertainty. The total uncertainty is calculated as the quadratic sum of all individual components.

Single top quarks produced via the t-channel also posses a charge asymmetry, which was measured at 7 and 8 TeV with relative precisions of 13% and 15%, respectively [52, 53]. This charge asymmetry was changed in simulation by a relative $\pm$15%.

An overview of all the individual Δm_t shifts is shown in Table 6.4. In general the systematic uncertainties estimated at 7 TeV are slightly smaller than at 8 TeV. This is caused by the slightly smaller charge asymmetry in the 7 TeV dataset, as observed when comparing Tables 4.2 and 4.3.

	Effect on Δm_t (MeV)	
Background	7 TeV	8 TeV
W^+/W^- ($\pm$7% or $\pm$4%)	3.22 $\pm$ 0.01	3.72 $\pm$ 0.01
W^+cc/W^-cc ($\pm$20%)	8.12 $\pm$ 0.01	9.05 $\pm$ 0.02
W^+bb/W^-bb ($\pm$20%)	0.06 $\pm$ 0.01	5.83 $\pm$ 0.02
Single top t-channel $t/\bar{t}$ ($\pm$15%)	2.94 $\pm$ 0.01	3.298 $\pm$ 0.005
Total	9.21 $\pm$ 0.02	11.86 $\pm$ 0.03

Table 6.4: Overview of the subcomponents of the background charge-asymmetry systematic uncertainty. The total is calculated as the quadratic sum of all individual components.

Method calibration

The calibration procedure is limited in precision by the statistics of the simulated event samples. Therefore, like in the m_t analysis, the residual bias after calibration is calculated with the nominal event samples (with m_t = 172.5 GeV) and quoted as systematic uncertainty.

Parton distribution functions

The uncertainty on the parton distribution functions (PDF) can affect the Δm_{t} measurement in multiple ways. The PDFs determine, for example, the difference in production of W^+ and W^- events. The CTEQ PDFs which were used to generate the simulated event samples are described by 22 independent parameters, as explained in Section 3.1.1. The up and down variation of each of these parameters results in 44 accompanying PDF up/down possibilities. The nominal simulated event sample is reweighted according to the deviation of each PDF from its original form. The sum of the larger shift (up or down) for each change in PDF is taken in quadrature to define the combined PDF uncertainty.

Lepton charge misassignment

A misassignment of the lepton charge can affect the calibration and can also lead to a dilution of the measurement. The charge misassignment probabilities for muons and electrons were measured in data and are smaller than 0.1% and 0.4%, respectively, as discussed in Sections 3.3.1 and 3.3.2. This means that the systematic uncertainty from charge misassignment is below 1% of the measured Δm_t value, which is negligible and therefore ignored.

Trigger

The trigger requires either the presence of an isolated lepton and at least three jets with $p_T > 30$ GeV, or only the presence of an isolated lepton. The lepton is not used in the mass reconstruction, so no systematic effect is expected from any mismodelling of the lepton trigger efficiency or p_T threshold. The requirement of three jets with $p_T > 30$ GeV in the 7 TeV trigger is highly efficient for events with 4 jets with $p_T > 30$ GeV. Any effect on kinematic distributions of the jets in selected events is thus expected to be small and should affect the top and antitop quarks in a similar way. Therefore no uncertainty needs to be quoted.

The combination of these systematic uncertainties with the Δm_t values measured in data from Section 6.2 gives a Δm_t result of

$$\Delta m_t = -106 \pm 420\,(\text{stat.}) \pm 131\,(\text{syst.})\ \text{MeV}$$

at 7 TeV and of

$$\Delta m_t = -155 \pm 191\,(\text{stat.}) \pm 91\,(\text{syst.})\ \text{MeV}$$

at 8 TeV. These are both compatible with the hypothesis of CPT symmetry, which predicts no difference in mass between the top and the antitop quark.

Chapter 7

Conclusions

The discovery of the top quark in 1995 by the CDF and DØ experiments opened the quest for the precise measurement of its mass and several other properties. At the LHC the high luminosity and the large $t\bar{t}$ production cross-section result in a gigantic number of top-quark pairs. This allows a more detailed study of the different aspects of the top quark to further test the Standard Model. One of the most important properties of the top quark is its mass, which is one of the main input parameters to the global electroweak fits which test the global consistency of the Standard Model, as explained in Section 1.2.3.

In this thesis a detailed study of the mass of the top quark was conducted based on ℓ+jets $t\bar{t}$ events produced during the 2011 and 2012 LHC run and recorded by the CMS experiment at centre-of-mass energies of 7 and 8 TeV. This entire study was performed with the Ideogram method and resulted in a very precise measurement of the top-quark mass, of the kinematic dependence of the top-quark mass estimator, and of the mass difference between the top and the antitop quark.

In Section 7.1 an overview of the main results of the top-quark mass measurements obtained in this thesis is given, together with a discussion of the interpretation. Some perspectives on further improvements of the precision are also presented. The final results of the measurement of the mass difference between the top and the antitop is reviewed in Section 7.2, together with the precision which can be expected once the LHC starts delivering collisions at a centre-of-mass energy of 13 TeV.

7.1 Measurement of the top-quark mass and its kinematic dependence

7.1.1 Inclusive top-quark mass measurement

The hadronic top-quark decay $t \to bq\bar{q}'$ in ℓ+jets $t\bar{t}$ events is used within this thesis to estimate the mass of the top quark. Via the combination of a kinematic fit and the Ideogram method as explained in Section 5.1 a robust top-quark mass estimator was constructed and applied to both the 7 and 8 TeV datasets. Together with the low jet energy scale uncertainty, which was achieved for the 8 TeV dataset as discussed in Section 3.3.4, a precise measurement of the top-quark mass is made possible. This measurement was further improved by tightening the p_T- and $|\eta|$-constraints applied on the jets in the event

selection. The most optimal kinematic constraints on the jets, $p_T > 30$ GeV and $|\eta| < 1.3$, results in a measured top-quark mass of

$$m_t = 172.51 \pm 0.13\,(\text{stat.}) \pm 0.97\,(\text{syst.})\ \text{GeV} = 172.51 \pm 0.98\ \text{GeV},$$

which is the first single measurement of the top-quark mass with a total uncertainty below 1 GeV.

In Figure 7.1 this measurement is compared to the most precise measurements from the ATLAS, CMS and Tevatron experiments. The top-quark mass measured in this thesis

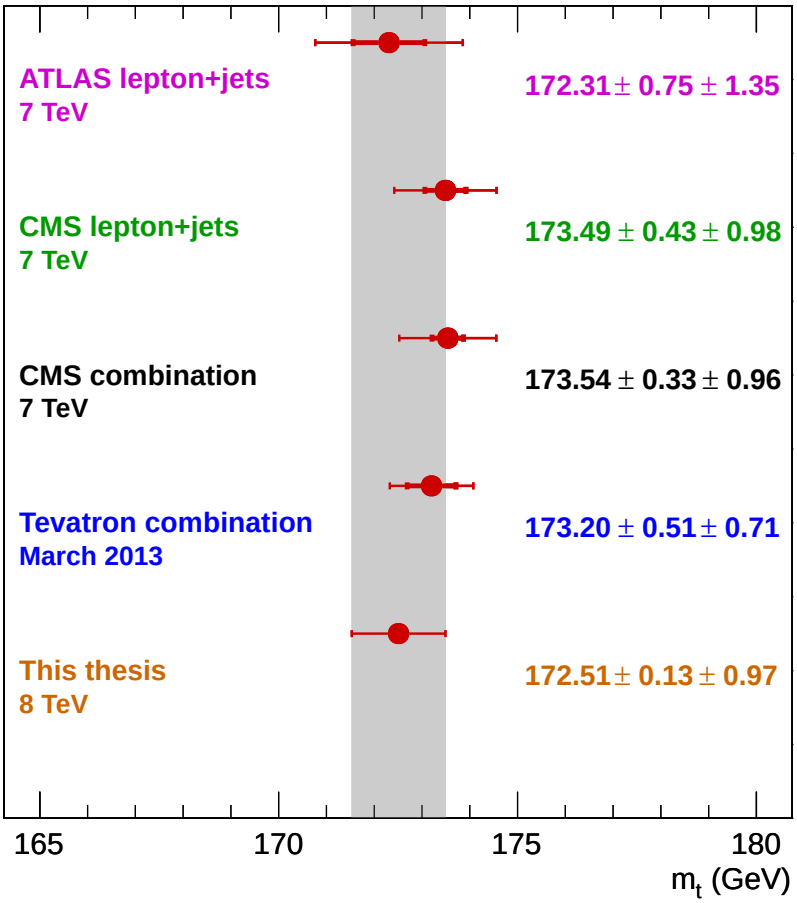

Figure 7.1: Comparison between the top-quark mass measured in this thesis and the most precise measurements performed by the ATLAS, CMS and Tevatron experiments.

is in good agreement with previous measurements when the uncertainties are taken into account. It has a total uncertainty of 0.98 GeV, which corresponds to a relative precision of 0.57%. The most precise measurements from ATLAS and CMS have relative precisions of 0.90% and 0.58%, respectively, and the most precise Tevatron measurement has a relative precision of 0.50%. This clearly shows the very good precision of the top-quark mass results obtained here.

An important aspect to take into account in this comparison is the fact that these other results were estimated via a combined measurement of the top-quark mass and the jet energy scale (using the known W-boson mass). This means that they have a significantly

reduced jet energy scale systematic uncertainty, but this combined measurements results in higher systematic uncertainties due to colour reconnection and other theoretical modelling uncertainties. The different sensitivity to the various systematic uncertainties of the result obtained in this thesis with respect to the other measurements means that a future combination of all these measurements can result in a significantly reduced total uncertainty on the top-quark mass.

Perspectives

The total uncertainty on the measured top-quark mass is completely dominated by two systematic uncertainties: the jet energy scale uncertainty, and the uncertainty on the Q^2-scale and the amount of initial- and final-state radiation. This means that for a sizeable improvement in the total uncertainty on the top-quark mass both systematic effects need to be reduced.

To reduce the effect of the jet energy scale on the top-quark mass only a reduction of the jet energy scale uncertainty in the barrel part of the CMS detector needs to be considered, since the measurement can easily be performed using only jets within the barrel as shown in this thesis. Figure 3.12 shows that the jet energy scale uncertainty at $|\eta^{PFjet}| = 0$ is dominated by the flavour-dependent and the absolute components. A reduction of the flavour-dependent jet energy scale uncertainty can be achieved by measuring the jet energy scale separately for jets originating from gluons, *uds* quarks, *c* quarks and *b* quarks. The absolute jet energy scale uncertainty, on the other hand, consists of several components as explained in Section 3.3.4. The uncertainty is dominated by two components, namely the uncertainty on the combined photon and lepton reference scales and the uncertainty on the single-particle response of the HCAL calorimeter. A reduction of the former is difficult but could be achieved via a better calibration of the reconstructed photons and leptons, while the latter could be reduced with a more precise measurement of the HCAL response to isolated charged hadrons in data.

A reduction of the uncertainty on the Q^2-scale and on the amount of initial- and final-state radiation is also necessary for future improved top-quark mass measurements. Therefore the current systematic variations of these parameters need to be compared to measurements in data of quantities which are sensitive to additional radiation. A promising example of such a physical quantity is the so-called gap fraction [132]. This is the fraction of events without an additional jet above a certain p_T threshold. The measurement of this gap fraction as a function of the p_T of this additional jet is shown in Figure 7.2. This clearly shows that the current Q^2-scale and ISR/FSR variations are significantly over-estimating the effect of additional radiation. Therefore future top-quark mass measurements could evaluate the corresponding systematic uncertainties by taking smaller variations of these parameters in simulation, leading to a reduced systematic uncertainty.

Besides the reduction of the total uncertainty on the measured top-quark mass additional progress on the theoretical interpretation of the measured top-quark mass is also necessary. The standard measurement techniques, including the one applied within this thesis, measure a top-quark mass which is defined as the input parameter m_t^{MC} of some event generator, but the theoretical interpretation of this parameter is highly non-trivial. This relates to the fact that the top quark is a colour charged object. Therefore it needs to connect to other partons via its decay products, meaning that the invariant mass of the final-state particles

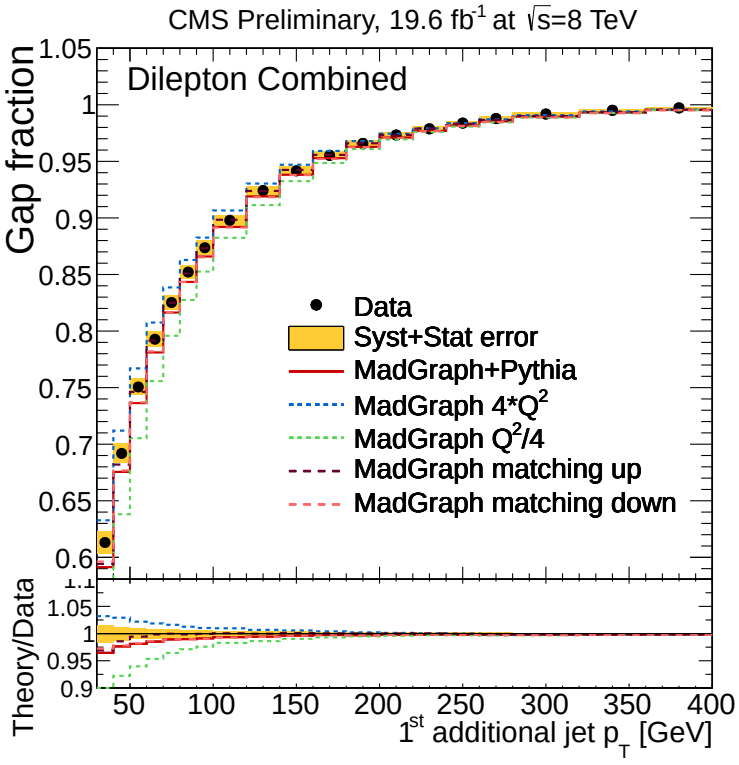

Figure 7.2: Measurement of the gap fraction, which is the fraction of events without an additional jet above a certain p_T threshold, as a function of this p_T threshold [132].

detected by any experiment is not unambiguously related to the initial top quark. So an additional understanding of these theoretical aspects is also of vital importance.

7.1.2 Binned top-quark mass measurement

The same Ideogram method was used to measure the dependence of the estimated top-quark mass on several kinematic quantities like the kinematics of the hadronic top-quark system or global event kinematics, as discussed in Section 5.5. In total the measurement was performed in bins of 12 kinematic variables and a comparison between the dependence measured in data and simulation was performed. In general a good description of the data by the central simulated $t\bar{t}$ sample is observed, which was generated with MadGraph and showered with the Z2* tune of Pythia.

Several theoretical variations of the central MadGraph sample were also compared to the measurements obtained from the data. These include underlying event variations, colour reconnection variations, variations of the Q^2-scale and the initial- and final-state radiation, and variations of the matching thresholds. For most theoretical curves a reasonable agreement with the measurements performed in data is observed, except in some kinematic regions where disagreement between the data and a few theoretical predictions is found.

In general the largest disagreements are observed with respect to the matching threshold variations and the downwards Q^2-scale variation. This observation is compatible with the measurement of the gap fraction shown in Figure 7.2, which shows also a larger disagreement

between the data and the downwards Q^2-scale variation. In the measurement of the top-quark mass dependence on the ΔR distance between the two light jets from the W-boson decay a large discrepancy between the data and the PYTHIA variation without colour reconnection is observed.

Once the LHC starts delivering additional proton-proton collisions to its experiments, which is expected to happen in 2015, additional data can be used to further improve these binned measurements. They can be performed in more bins resulting in more stringent tests of the theoretical modelling of the measured top-quark mass. With more data one should also think about additional theoretical variations which are more realistic than simply turning a certain effect on and off in the event generation as was done for colour reconnection. Curves with an intermediate level of colour reconnection should also be compared to the data.

7.2 Measurement of the mass difference between the top and the antitop quark

A stringent test of the CPT symmetry of relativistic quantum field theories like the Standard Model was also performed in this thesis, as discussed in Chapter 6. Therefore a measurement of the mass difference between the top and the antitop quark was carried out and resulted in a value of

$$\Delta m_t = -155 \pm 191\,(\text{stat.}) \pm 91\,(\text{syst.})\ \text{MeV}.$$

This value has a total uncertainty of 212 MeV and is in agreement with the expectation from CPT symmetry, which requires no mass difference between a particle and its antiparticle.

The precision of this result is significantly better than any of the previous measurements as listed in Table 1.4. The main reason for this significant decrease in total uncertainty is the large dataset which was taken by the CMS experiment in 2012, resulting in a gigantic number of top quarks. With respect to the 7 TeV result from CMS shown in this table the systematic uncertainty was also significantly reduced from 270 to 91 MeV. This reduction in systematic uncertainty was mainly caused by the fact that all systematic uncertainties in this 7 TeV result were statistically compatible with zero but with large uncertainties, therefore the total systematic uncertainty had a large statistical component. For the new result presented in this thesis significantly larger simulated signal and background samples were used, resulting in more accurate estimates of the systematic uncertainties and hence a reduction of these uncertainties. For some of these systematic uncertainties, like the jet energy scale and the b-tagging efficiency, the increased dataset also helped towards a more precise measurement of the corresponding uncertainties.

Perspectives

Since the total uncertainty on the Δm_t measurement is still dominated by the statistical uncertainty, the additional collision data which is expected to be taken from 2015 onwards at $\sqrt{s} = 13$ TeV will result in a significant reduction of the total uncertainty. The theoretical $t\bar{t}$ production cross-section at a centre-of-mass energy of 13 TeV is equal to $\sigma_{t\bar{t}}(13\ \text{TeV}) =$

806 pb [29, 30], which is a factor 3.28 larger than the $t\bar{t}$ production cross-section at 8 TeV: $\sigma_{t\bar{t}}(8\text{ TeV}) = 246$ pb.

This large increase in the $t\bar{t}$ production cross section will thus lead to an even larger number of top quarks available for analysis. Assuming a dataset of 20 fb^{-1} at $\sqrt{s} = 13$ TeV the statistical uncertainty on Δm_t will reduce by a factor $\sqrt{3.28}$ from 191 MeV to 106 MeV, resulting in a total uncertainty of 140 MeV. At this point the systematic uncertainty becomes important so only a limited additional reduction of the total uncertainty can be achieved by adding even more collision data.

For an improvement in the precision on Δm_t the main systematic uncertainty, namely the b vs. $\bar{b}$ jet response, needs to be reduced. This could be done with a $b\bar{b}$ event sample, where one of the two b jets contains a soft electron or muon. The charge of this soft lepton can then be used to obtain the flavour of the jet (b or $\bar{b}$) and a p_T-balancing technique can be applied to measure the difference in jet response between b and $\bar{b}$ jets.

Bibliography

[1] CMS Collaboration, "Measurement of the mass difference between top and antitop quarks", *JHEP* **1206** (2012) 109, doi:10.1007/JHEP06(2012)109, arXiv:1204.2807.

[2] CMS Collaboration, "Measurement of the top - antitop mass difference in pp collisions at sqrt(s) = 8 TeV", CMS Physics Analysis Summary CMS-PAS-TOP-12-031, (2012).

[3] F. Mandl and G. Shaw, "Quantum Field Theory". (1984) 358 p.

[4] M. E. Peskin and D. V. Schroeder, "An Introduction to Quantum Field Theory". (1995) 842 p.

[5] S. Weinberg, "The Quantum Theory of Fields. Vol. 1: Foundations". (1995) 609 p.

[6] S. L. Glashow, "Partial Symmetries of Weak Interactions", *Nucl. Phys.* **22** (1961) 579, doi:10.1016/0029-5582(61)90469-2.

[7] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (1967) 1264, doi:10.1103/PhysRevLett.19.1264.

[8] A. Salam, *Elementary Particle Theory, proceedings 8th Nobel Symposium Aspenas Gärden* (1968) 367.

[9] J. Beringer et al., "The Review of Particle Physics", *Phys. Rev. D* **86** (2012) 010001, doi:10.1103/PhysRevD.86.010001.

[10] CMS Collaboration, "Measurements of the properties of the new boson with a mass near 125 GeV", CMS Physics Analysis Summary CMS-PAS-HIG-13-005, (2013).

[11] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett. B* **716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[12] ATLAS Collaboration, "Observation of a new particle in the search for the standard model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[13] F. Englert and R. Brout, "Broken symmetries and the masses of gauge bosons", *Phys. Rev. Lett.* **13** (1964) 321, doi:10.1103/PhysRevLett.13.321.

[14] P. W. Higgs, "Broken symmetries and the masses of gauge bosons", *Phys. Rev. Lett.* **13** (1964) 508, doi:10.1103/PhysRevLett.13.508.

[15] G. Guralnik, C. Hagen, and T. Kibble, "Global Conservation Laws and Massless Particles", *Phys.Rev.Lett.* **13** (1964) 585, doi:10.1103/PhysRevLett.13.585.

[16] R. G. Sachs, "The Physics of Time Reversal". (1987) 309 p.

[17] KLOE Collaboration, "CPT symmetry and quantum mechanics tests in the neutral kaon system at KLOE", *Found. Phys.* **40** (2010) 852–866, doi:10.1007/s10701-009-9366-x.

[18] KTeV Collaboration, "Precise Measurements of Direct CP Violation, CPT Symmetry, and Other Parameters in the Neutral Kaon System", *Phys. Rev. D* **83** (2011) 092001, doi:10.1103/PhysRevD.83.092001, arXiv:1011.0127.

[19] M. Baak et al., "The Electroweak Fit of the Standard Model after the Discovery of a New Boson at the LHC", *Eur. Phys. J. C* **72** (2012) 2205, doi:10.1140/epjc/s10052-012-2205-9, arXiv:1209.2716.

[20] Planck Collaboration, "Planck 2013 results. XVI. Cosmological parameters", *Submitted to Astronomy & Astrophysics* arXiv:1303.5076.

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

[22] DØ Collaboration, "Observation of the Top Quark", *Phys. Rev. Lett.* **74** (1995) 2632–2637, doi:10.1103/PhysRevLett.74.2632, arXiv:hep-ex/9503003.

[23] "The CDF Experiment". http://www-cdf.fnal.gov/.

[24] "The DØ Experiment". http://www-d0.fnal.gov/.

[25] "The Tevatron Collider". http://www-bdnew.fnal.gov/tevatron/.

[26] ATLAS Collaboration, "The ATLAS experiment at the CERN Large Hadron Collider", *JINST* **03** (2008) S08003, doi:10.1088/1748-0221/3/08/S08003.

[27] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **03** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[28] L. Evans and P. Bryant (editors), "LHC Machine", *JINST* **03** (2008) S08001, doi:10.1088/1748-0221/3/08/S08001.

[29] M. Czakon, P. Fiedler, and A. Mitov, "The total top quark pair production cross-section at hadron colliders through $O(\alpha_s^4)$", arXiv:1303.6254.

[30] M. Czakon et al., "Constraints on the gluon PDF from top quark pair production at hadron colliders", arXiv:1303.7215.

[31] The Tevatron Electroweak Working Group, "Combination of the $t\bar{t}$ production cross section measurements from the Tevatron Collider", *DØ Note 6363 and CDF Note 10926* (2012).

[32] CMS Collaboration, "Measurement of the $t\bar{t}$ production cross section in the dilepton channel in pp collisions at sqrt(s) = 7 TeV", *JHEP* **11** (2012) 067, doi:10.1007/JHEP11(2012)067, arXiv:1208.2671.

[33] CMS Collaboration, "Measurement of the $t\bar{t}$ production cross section in the dilepton channel in *pp* collisions at $\sqrt{s} = 8$ TeV", CMS Physics Analysis Summary CMS-PAS-TOP-12-007, (2012).

[34] A. Quadt, "Top quark physics at hadron colliders", *Eur. Phys. J.* **C48** (2006) 835–1000, doi:10.1140/epjc/s2006-02631-6.

[35] D0 Collaboration, "An improved determination of the width of the top quark", *Phys. Rev. D* **85** (2012) 091104, doi:10.1103/PhysRevD.85.091104, arXiv:1201.4156.

[36] CMS Collaboration, "Measurement of the ratio $B(t \to Wb)/B(t \to Wq)$", CMS Physics Analysis Summary CMS-PAS-TOP-12-035, (2012).

[37] CMS Collaboration, "Search for flavor changing neutral currents in top quark decays in proton-proton collisions at $\sqrt{s} = 8$ TeV", CMS Physics Analysis Summary CMS-PAS-TOP-12-037, (2013).

[38] CMS Collaboration, "Search for baryon number violating top-quark decays in *pp* collisions at $\sqrt{s} = 8$ TeV", CMS Physics Analysis Summary CMS-PAS-B2G-12-023, (2013).

[39] ATLAS Collaboration, "Measurements of top quark pair relative differential cross-sections with ATLAS in *pp* collisions at $\sqrt{s = 7}$ TeV", *Eur. Phys. J. C* **73** (2013) 2261, doi:10.1140/epjc/s10052-012-2261-1, arXiv:1207.5644.

[40] CMS Collaboration, "Measurement of differential top-quark pair production cross sections in *pp* collisions at sqrt(s) = 7 TeV", *Eur. Phys. J. C* **73** (2013) 2339, doi:10.1140/epjc/s10052-013-2339-4, arXiv:1211.2220.

[41] CMS Collaboration, "Measurement of differential top-quark pair production cross sections in the lepton+jets channel in pp collisions at 8 TeV", CMS Physics Analysis Summary CMS-PAS-TOP-12-027, (2012).

[42] CMS Collaboration, "Measurement of the differential ttbar cross section in the dilepton channel at 8 TeV", CMS Physics Analysis Summary CMS-PAS-TOP-12-028, (2012).

[43] CDF Collaboration, "Measurement of Single Top Quark Production in 7.5 fb^{-1} of CDF Data Using Neural Networks", *CDF Note 10793* (2013).

[44] D0 Collaboration, "Evidence for s-channel single top quark production in $p\bar{p}$ collisions at sqrt(s) = 1.96 TeV", *Submitted to Phys. Lett. B.* (2013) arXiv:1307.0731.

[45] ATLAS Collaboration, "Measurement of the t-channel single top-quark production cross section in pp collisions at sqrt(s) = 7 TeV with the ATLAS detector", *Phys. Lett. B* **717** (2012) 330, doi:10.1016/j.physletb.2012.09.031, arXiv:1205.3130.

[46] CMS Collaboration, "Measurement of the single-top-quark t-channel cross section in pp collisions at sqrt(s) = 7 TeV", *JHEP* **12** (2012) 035, doi:10.1007/JHEP12(2012)035, arXiv:1209.4533.

[47] ATLAS Collaboration, "Measurement of t-Channel Single Top-Quark Production in pp Collisions at sqrt(s) = 8 TeV with the ATLAS detector", ATLAS Note ATLAS-CONF-2012-132, (2012).

[48] CMS Collaboration, "Measurement of the single-top t-channel cross section in pp collisions at centre-of-mass energy of 8 TeV", CMS Physics Analysis Summary CMS-PAS-TOP-12-011, (2013).

[49] ATLAS Collaboration, "Evidence for the associated production of a W boson and a top quark in ATLAS at sqrt(s) = 7 TeV", *Phys. Lett. B* **716** (2012) 142, doi:10.1016/j.physletb.2012.08.011, arXiv:1205.5764.

[50] CMS Collaboration, "Evidence for associated production of a single top quark and W boson in pp collisions at sqrt(s) = 7 TeV", *Phys. Rev. Lett.* **110** (2012) 022003, doi:10.1103/PhysRevLett.110.022003, arXiv:1209.3489.

[51] CMS Collaboration, "Single Top associated tW production at 8 TeV in the two lepton final state", CMS Physics Analysis Summary CMS-PAS-TOP-12-040, (2013).

[52] ATLAS Collaboration, "Measurement of the t-channel single top-quark and top-antiquark cross-sections and their ratio in pp collisions at $\sqrt{s} = 7$ TeV", ATLAS Note ATLAS-CONF-2012-056, (2012).

[53] CMS Collaboration, "Measurement of the single-top t-channel charge ratio at 8 TeV", CMS Physics Analysis Summary CMS-PAS-TOP-12-038, (2013).

[54] ATLAS Collaboration, "Top-quark mass measurement in the $e\mu$ channel using the m_{T2} variable at ATLAS", ATLAS Note ATLAS-CONF-2012-082, (2013).

[55] ATLAS Collaboration, "Measurement of the Top Quark Mass from $\sqrt{s} = 7$ TeV ATLAS Data using a 3-dimensional Template Fit", ATLAS Note ATLAS-CONF-2013-046, (2013).

[56] ATLAS Collaboration, "Determination of the Top Quark Mass with a Template Method in the All-Hadronic Decay Channel using 2.04 fb^{-1} of ATLAS Data", ATLAS Note ATLAS-CONF-2012-030, (2012).

[57] CMS Collaboration, "Measurement of the top-quark mass in $t\bar{t}$ events with dilepton final states in pp collisions at $\sqrt{s} = 7$ TeV", *Eur. Phys. J. C* **72** (2012) 2202, doi:10.1140/epjc/s10052-012-2202-z, arXiv:1209.2319.

[58] CMS Collaboration, "Measurement of the top-quark mass in $t\bar{t}$ events with lepton+jets final states in pp collisions at $\sqrt{s} = 7$ TeV", *JHEP* **12** (2012) 105, doi:10.1007/JHEP12(2012)105, arXiv:1209.2319.

[59] CMS Collaboration, "Measurement of the top-quark mass in all-jets t t-bar events in pp collisions at $\sqrt{s} = 7$ TeV", *Submitted to Eur. Phys. J. C* (2013) arXiv:1307.4617.

[60] CMS Collaboration, "Combination of CMS results on the mass of the top quark using up to 5.0 fb^{-1} of data", CMS Physics Analysis Summary CMS-PAS-TOP-11-018, (2012).

[61] The Tevatron Electroweak Working Group, "Combination of CDF and DØ results on the mass of the top quark using up to 8.7 fb^{-1} at the Tevatron", *DØ Note 6381 and CDF Note 10976* (2013) arXiv:1305.3929.

[62] D0 Collaboration, "Direct measurement of the mass difference between top and antitop quarks", *Phys. Rev. D* **84** (2011) 052005, doi:10.1103/PhysRevD.84.052005, arXiv:1106.2063.

[63] CDF Collaboration, "Measurement of the Mass Difference Between Top and Anti-top Quarks at CDF", *Phys. Rev. D* **87** (2013) 052013, doi:10.1103/PhysRevD.87.052013, arXiv:1210.6131.

[64] ATLAS Collaboration, "Observation of spin correlation in $t\bar{t}$ events from pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector", *Phys. Rev. Lett.* **108** (2012) 212001, doi:10.1103/PhysRevLett.108.212001, arXiv:1203.4081.

[65] CMS Collaboration, "Measurement of Spin Correlations in $t\bar{t}$ events in the dilepton channels in pp collisions at $\sqrt{s} = 7$ TeV", CMS Physics Analysis Summary CMS-PAS-TOP-12-004, (2012).

[66] ATLAS Collaboration, "Measurement of top quark polarisation in $t\bar{t}$ events with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV", ATLAS Note ATLAS-CONF-2012-133, (2012).

[67] CMS Collaboration, "Measurement of the top polarization in the dilepton final state", CMS Physics Analysis Summary CMS-PAS-TOP-12-016, (2012).

[68] ATLAS Collaboration, "Measurement of the top quark charge in pp collisions at $\sqrt{s} = 7$ TeV in the ATLAS experiment", ATLAS Note ATLAS-CONF-2011-141, (2011).

[69] CMS Collaboration, "Constraints on the Top-Quark Charge from Top-Pair Events", CMS Physics Analysis Summary CMS-PAS-TOP-11-031, (2012).

[70] CMS and ATLAS Collaborations, "Combination of the ATLAS and CMS measurements of the W-boson polarization in top-quark decays", CMS Physics Analysis Summary CMS-PAS-TOP-12-025 and ATLAS Note ATLAS-CONF-2013-033, (2013).

[71] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) 04013, doi:10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.

[72] ATLAS Collaboration, "Measuring the b-tag efficiency in a $t\bar{t}$ sample with 4.7 fb^{-1} of data from the ATLAS detector", ATLAS Note ATLAS-CONF-2012-097, (2012).

[73] P. Van Mulders, "Calibration of the jet energy scale using top quark events at the LHC". PhD thesis, Vrije Universiteit Brussel, 2010. http://cdsweb.cern.ch/record/1308729.

[74] "CERN". http://home.web.cern.ch/.

[75] "The Large Electron-Positron Collider". http://home.web.cern.ch/about/accelerators/large-electron-positron-collider.

[76] CMS Collaboration, "CMS Physics TDR: Volume I, Detector Performace and Software", CERN-LHCC-2006-001, (2006).

[77] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **03** (2008) S08002, doi:10.1088/1748-0221/3/08/S08002.

[78] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **03** (2008) S08005, doi:10.1088/1748-0221/3/08/S08005.

[79] LHCf Collaboration, "The LHCf detector at the CERN Large Hadron Collider", *JINST* **03** (2008) S08006, doi:10.1088/1748-0221/3/08/S08006.

[80] TOTEM Collaboration, "The TOTEM Experiment at the CERN Large Hadron Collider", *JINST* **03** (2008) S08007, doi:10.1088/1748-0221/3/08/S08007.

[81] CMS Collaboration, "Tracking and Primary Vertex Results in First 7 TeV Collisions", CMS Physics Analysis Summary CMS-PAS-TRK-10-005, (2010).

[82] CMS Collaboration, "Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at sqrt(s) = 7 TeV", *Submitted to JINST,* arXiv:1306.2016.

[83] C. Eck et al., "LHC computing Grid : Technical Design Report", CERN-LHCC-2005-024, (2005).

[84] T. Sjöstrand, "Monte Carlo Tools", (2009). arXiv:0911.5286.

[85] M. Dobbs et al., "Les Houches Guidebook to Monte Carlo Generators for Hadron Collider Physics", (2004). arXiv:hep-ph/0403045.

[86] P. Skands, "QCD for Collider Physics", (2012). arXiv:1104.2863.

[87] P. M. Nadolsky et al., "Implications of CTEQ global analysis for collider observables", *Phys. Rev. D* **78** (2008) doi:10.1103/PhysRevD.78.013004, arXiv:0802.0007.

[88] 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.

[89] J. Alwall et al., "MadGraph 5 : Going Beyond", *JHEP* **06** (2011) 128, doi:10.1007/JHEP06(2011)128, arXiv:1106.0522.

[90] 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.

[91] S. Frixione and B. R. Webber, "The MC@NLO 3.4 Event Generator", arXiv:0812.0770.

[92] Y. L. Dokshitzer, "Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics", *Sov. Phys. JETP* **46** (1977) 641.

[93] V. N. Gribov and L. N. Lipatov, "Deep inelastic e p scattering in perturbation theory", *Sov. J. Nucl. Phys.* **15** (1972) 438.

[94] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language", *Nucl. Phys. B* **126** (1977) 298, doi:10.1016/0550-3213(77)90384-4.

[95] G. Corcella et al., "HERWIG 6.5 Release Note", arXiv:hep-ph/0210213.

[96] J. Alwall et al., "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions", *Eur. Phys. J. C* **53** (2008) 473, doi:10.1140/epjc/s10052-007-0490-5, arXiv:0706.2569.

[97] M. L. Mangano et al., "Matching matrix elements and shower evolution for top-quark production in hadronic collisions", *JHEP* **0701** (2007) 013, doi:10.1088/1126-6708/2007/01/013, arXiv:hep-ph/0611129.

[98] B. Andersson, "The Lund Model". (1998) 471 p.

[99] M. H. Seymour and M. Marx, "Monte Carlo Event Generators", (2013). arXiv:1304.6677.

[100] S. Jadach, J. H. Kuhn, and Z. Was, "TAUOLA - a library of Monte Carlo programs to simulate decays of polarized τ leptons", *Computer Physics Communications* **64** (1990) 275, doi:10.1016/0010-4655(91)90038-M.

[101] D. Wicke and P. Z. Skands, "Non-perturbative QCD Effects and the Top Mass at the Tevatron", *Nuovo Cim. B* **123** (2008) arXiv:0802.0007.

[102] R. Field, "Early LHC Underlying Event Data - Findings and Surprises", (2010). arXiv:1010.3558.

[103] P. Z. Skands, "Tuning Monte Carlo Generators: The Perugia Tunes", *Phys. Rev. D* **82** (2010) 074018, doi:10.1103/PhysRevD.82.074018, arXiv:1005.3457.

[104] J. Allison et al., "Geant4 developments and applications", *IEEE Trans. Nucl. Sci.* **53** (2006) 270, doi:10.1109/TNS.2006.869826.

[105] CMS Collaboration, "Performance of CMS muon reconstruction in pp collision events at sqrt(s) = 7 TeV", *JINST* **07** (2012) 10002, doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071.

[106] CMS Collaboration, "Performance of CMS Muon Reconstruction in Cosmic-Ray Events", *JINST* **05** (2010) 03022, doi:10.1088/1748-0221/5/03/T03022, arXiv:0911.4994.

[107] CMS Collaboration, "Electron reconstruction and identification at sqrt(s) = 7 TeV", CMS Physics Analysis Summary CMS-PAS-EGM-10-004, (2010).

[108] W. Adam et al., "Reconstruction of electrons with the Gaussian-sum filter in theCMS tracker at the LHC", *J. Phys. G: Nucl. Part. Phys.* **31** (2005) N9.

[109] CMS Collaboration, "Electron performance with 19.6 fb^{-1} of data collected at $\sqrt{s} = 8$ TeV with the CMS detector", CMS Detector Performance Summaries CMS-DP-2013-003, (2013).

[110] CMS Collaboration, "Measurement of the lepton charge asymmetry in inclusive W production in pp collisions at $\sqrt{s} = 7$ TeV", *JHEP* **04** (2011) 050, doi:10.1007/JHEP04(2011)050, arXiv:1103.3470.

[111] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", CMS Physics Analysis Summary CMS-PAS-PFT-09-001, (2009).

[112] CMS Collaboration, "Particle-flow commissioning with muons and electrons from J/Psi and W events at 7 TeV", CMS Physics Analysis Summary CMS-PAS-PFT-10-003, (2010).

[113] "Particle Flow in PAT (PF2PAT)". http://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuidePF2PAT.

[114] CMS Collaboration, "Commissioning of the Particle-flow Event Reconstruction in Minimum-Bias and Jet Events from pp collisions at 7 TeV", CMS Physics Analysis Summary CMS-PAS-PFT-10-002, (2010).

[115] C. Buttar et al., "Standard Model Handles and Candles Working Group: Tools and Jets Summary Report", (2008). arXiv:0803.0678.

[116] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[117] CMS Collaboration, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS", *JINST* **6** (2011) 11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.

[118] CMS Collaboration, "Jet Energy Scale performance in 2011", CMS Detector Performance Summaries CMS-DP-2012-006, (2012).

[119] CMS Collaboration, "Status of the 8 TeV Jet Energy Corrections and Uncertainties based on 11 fb^{-1} of data in CMS", CMS Detector Performance Summaries CMS-DP-2013-011, (2013).

[120] A. Perloff (on behalf of the CMS Collaboration), "Pileup measurement and mitigation techniques in CMS", *J. Phys.: Conf. Ser.* **404** (2012) 012045, doi:10.1088/1742-6596/404/1/012045.

[121] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Phys. Lett. B* **659** (2008) 119, doi:10.1016/j.physletb.2007.09.077, arXiv:0707.1378.

[122] M. Cacciari, G. P. Salam, and G. Soyez, "The Catchment Area of Jets", *JHEP* **04** (2008) 005, doi:10.1088/1126-6708/2008/04/005, arXiv:0802.1188.

[123] H. Kirschenmann (on behalf of the CMS Collaboration), "Determination of the Jet Energy Scale in CMS", *J. Phys.: Conf. Ser.* **404** (2012) 012013, doi:10.1088/1742-6596/404/1/012013.

[124] CMS Collaboration, "Missing transverse energy performance of the CMS detector", *JINST* **6** (2011) 09001, doi:10.1088/1748-0221/6/09/P09001, arXiv:1106.5048v1.

[125] CMS Collaboration, "Single Muon efficiencies in 2012 Data", CMS Detector Performance Summaries CMS-DP-2013-009, (2013).

[126] J. Goh et al., "Electron Efficiency Measurement for Top Quark Physics at $\sqrt{s} = 8$ TeV", CMS Analysis Note AN-2012/429, (2013).

[127] J. D'Hondt et al., "Fitting of Event Topologies with External Kinematic Constraints in CMS", CMS Note 2006/023, (2006).

[128] DELPHI Collaboration, "Measurement of the Mass and Width of the W Boson in e^+e^- Collisions at $\sqrt{s} = 161$ - 209 GeV", *Eur. Phys. J. D* **55** (2008) 1, doi:10.1140/epjc/s10052-008-0585-7, arXiv:0803.2534.

[129] R. G. Miller, "The jackknife – a review", *Biometrika* **61** (1974) 1.

[130] CMS Collaboration, "Measurement of inclusive W and Z boson cross sections in pp collisions at $\sqrt{s} = 8$ TeV", CMS Physics Analysis Summary CMS-PAS-SMP-12-011, (2012).

[131] CMS Collaboration, "Measurements of inclusive W and Z cross sections in pp collisions at $\sqrt{s} = 7$ TeV", *JHEP* **01** (2011) 080, doi:10.1007/JHEP01(2011)080, arXiv:1107.4789.

[132] CMS Collaboration, "Measurement of the Jet Multiplicity in dileptonic Top Quark Pair Events at 8 TeV", CMS Physics Analysis Summary CMS-PAS-TOP-12-041, (2013).

Summary

The large $t\bar{t}$ dataset which was delivered in 2011 and 2012 by the Large Hadron Collider to the CMS experiment is used within this thesis to perform very precise measurements of some top-quark properties. Hereby proton-proton collisions are used where a top quark pair was produced which decayed in the ℓ+jets channel ($t\bar{t} \to b\bar{b}W^+W^- \to b\bar{b}q\bar{q}'\ell\nu_\ell$). The invariant mass of the hadronically decaying top quark is estimated via the Ideogram method which is interfaced to a kinematic fit. With this method several top-quark mass related measurements are performed. This allows stringent tests of the Standard Model of elementary particle physics, since the top-quark mass is one of its important input parameters which needs to be measured in data.

With 19.7 fb^{-1} of collision data taken at a centre-of-mass energy of 8 TeV a top-quark mass value of

$$m_t = 172.51 \pm 0.13\,(\text{stat.}) \pm 0.97\,(\text{syst.})\ \text{GeV} = 172.51 \pm 0.98\ \text{GeV}$$

was measured by using only jets with $p_T > 30$ GeV within $|\eta| < 1.3$. This is the first top-quark mass measurement with a total uncertainty below 1 GeV. The systematic uncertainty on this measurement is completely dominated by two components, namely the jet energy scale uncertainty and the uncertainty on the modelling of radiation. Compared to previous measurements of the top-quark mass the systematic uncertainty due to colour reconnection is significantly reduced, since no combined measurement of the top-quark mass and the jet energy scale is performed within this thesis.

Exactly the same analysis method is also applied in bins of several kinematic quantities like for example the invariant mass of the $t\bar{t}$ system, the kinematics of the hadronic top-quark system and the ΔR distances between the decay products. In general a good description of the kinematic dependence of the measured top quark mass on the event kinematics is observed. Except in some kinematic regions where significant disagreements were observed with respect to the theoretical predictions with a modified modelling of the radiation, and with respect to theoretical predictions not including colour reconnection.

Finally the mass difference between the top and the antitop quark was also measured, resulting in a difference of

$$\Delta m_t = -155 \pm 191\,(\text{stat.}) \pm 91\,(\text{syst.})\ \text{MeV} = -155 \pm 212\ \text{MeV}.$$

This measurement is in excellent agreement with the expectation from CPT symmetry, which requires no mass difference between a particle and its antiparticle. The uncertainty on this value is still dominated by the statistical uncertainty, since the systematic effects which influence the absolute top-quark mass measurement have a significantly reduced impact on this mass difference because they alter the individual masses of the top and antitop quarks in a fully correlated way.

Samenvatting

Meting van de top-quark massa en van het verschil in massa tussen top en antitop quarks aan de LHC

De grote $t\bar{t}$ dataset die in 2011 en 2012 geproduceerd werd door de Large Hadron Collider en gedetecteerd werd door het CMS experiment wordt in deze thesis gebruikt om enkele zeer precieze metingen van de eigenschappen van de top quark te doen. Hiervoor worden proton-proton botsingen gebruikt waarin een top-quark paar werd geproduceerd dat vervalt in het ℓ+jets kanaal ($t\bar{t} \to b\bar{b}W^+W^- \to b\bar{b}q\bar{q}'\ell\nu_\ell$). De invariante massa van de hadronisch vervallende top quark wordt geschat met behulp van de Ideogram methode, die gecombineerd wordt met een kinematische fit. Met deze methode worden verschillende top-quark massa gerelateerde metingen gedaan. Deze laten zeer precieze tests van het Standaard Model van de elementaire deeltjesfysica toe, aangezien de top-quark massa een van de belangrijkste parameters is die experimenteel moet worden gemeten.

Met 19.7 fb^{-1} aan botsingsgegevens, die werden genomen bij een energie van 8 TeV, werd een top-quark massa waarde van

$$m_t = 172.51 \pm 0.13\,(\text{stat.}) \pm 0.97\,(\text{syst.})\ \text{GeV} = 172.51 \pm 0.98\ \text{GeV}$$

gemeten. Hierbij werden enkel jets met $p_T > 30$ GeV en $|\eta| < 1.3$ gebruikt. Dit is de eerste meting van de top-quark massa met een totale onzekerheid die kleiner is dan 1 GeV. De systematische onzekerheid op deze meting is volledig gedomineerd door twee componenten, namelijk de jet energie calibratie onzekerheid en de onzekerheid op de emissie van extra straling. Deze meting heeft een sterk gereduceerde systematische onzekerheid door de beschrijving van *colour reconnection* in vergelijking met vorige metingen van de top-quark massa, aangezien er in deze thesis geen gecombineerde meting van de top-quark massa en de jet energie calibratie werd uitgevoerd.

Exact dezelfde methode werd ook toegepast in bins van verschillende kinematische grootheden, zoals bijvoorbeeld de invariante massa van het $t\bar{t}$ systeem, de kinematica van het hadronische top-quark systeem en de ΔR afstand tussen de vervalproducten. In het algemeen werd er een goede beschrijving van de kinematische afhankelijkheid van de gemeten top-quark massa vastgesteld. Behalve in enkele kinematische regio's waar een significant verschil werd waargenomen ten opzichte van de theoretische voorspellingen met een gewijzigde emissie van extra straling en ten opzichte van de theoretische voorspelling zonder *colour reconnection*.

Tenslotte werd ook het massa verschil tussen de top en de antitop quark gemeten, wat resulteerde in een verschil van

$$\Delta m_t = -155 \pm 191\,(\text{stat.}) \pm 91\,(\text{syst.})\ \text{MeV} = -155 \pm 212\ \text{MeV}.$$

Deze meting is in goede overeenkomst met de verwachting van CPT symmetrie, die geen verschil in massa tussen een deeltje en zijn antideeltje toelaat. De onzekerheid op deze waarde is nog steeds gedomineerd door de statistische onzekerheid, aangezien de meeste systematische effecten die de meting van de top-quark massa beinvloeden een kleinere impact hebben op dit verschil in massa. Dit is een rechtstreeks gevolg van het feit dat deze effecten dezelfde invloed hebben op de individuele massa's van de top en antitop quarks.