Vrije Universiteit Brussel

Faculteit Wetenschappen
Departement Natuurkunde

Universiteit Antwerpen

Faculteit Wetenschappen
Departement Fysica

Calibration of the jet energy scale using top quark events at the LHC

Petra Van Mulders

Promotoren
Prof. Dr. Jorgen D'Hondt
Prof. Dr. Nick Van Remortel

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

Juni 2010

Doctoral examination commission:

Prof. Dr. Nick Van Remortel (Universiteit Antwerpen)
Prof. Dr. Eddi De Wolf (Universiteit Antwerpen)
Prof. Dr. Robert Roosen (Vrije Universiteit Brussel, *chair*)
Prof. Dr. Alexandre Sevrin (Vrije Universtiteit Brussel)
Prof. Dr. Stefaan Tavernier (Vrije Universiteit Brussel)
Dr. Tim Christiansen (CERN)
Dr. Wouter Verkerke (NIKHEF)

This thesis is realised with the financial support of IWT-Vlaanderen.

Cover illustration: Different visualisations of the reconstructed particles in the CMS detector originating from the decay of a semi-muonic $t\bar{t}$ event at a center of mass energy of 10 TeV.

Print: Silhouet, Maldegem

2010 Uitgeverij ASP (Academic and Scientific Publishers nv)
Ravensteingalerij 28
B-1000 Brussels
Tel. + 32 (0)2 289 26 50
Fax + 32 (0)2 289 26 59
E-mail: infoaspeditions.be
www.aspeditions.be

ISBN 978 90 5487 761 5
NUR 924
Legal Deposit D/2010/11.161/085

We have to remember that what we observe is not nature in itself but nature exposed to our method of questioning.

Werner Heisenberg

Contents

Introduction

The elementary particles and their fundamental interactions, with the exception of gravity, are described by the Standard Model of particle physics. This quantum field theory endured numerous stringent experimental tests over the last decades. Up to date, the various predictions of the Standard Model are confirmed by diverse high energy experiments. There is one predicted particle however, the Standard Model Higgs boson, which is assumed to be responsible for the generation of the masses of the fundamental particles, escaping detection so far. Despite the successes of the Standard Model in describing the empirical data, the theory is believed to be only an effective low energy field theory. New physics phenomena are expected at energy scales which became currently accessible with the most recent accelerator experiment.

Since March 2010, the Large Hadron Collider (LHC) located at the CERN laboratory near Geneva, is colliding proton beams at a center of mass energy of 7 TeV. One of the main goals of the general-purpose detectors at the LHC is the observation of the Higgs boson as the last missing building block of the Standard Model. In addition, the detectors are designed to study possible new physics phenomena. These phenomena can be studied for instance by searching for effects in the differential cross sections of observables sensitive to new physics. The top quark is the heaviest particle in the Standard Model, therefore, observables which are related to top quark physics provide a probe for the physics expected at higher energy scales.

For detailed studies at the LHC, the commissioning and calibration of the detectors and reconstruction tools are essential. Many of the physics studies at the LHC are based on the reconstruction of jets or particles formed by hadronizing partons, i.e. quarks and gluons. The precise reconstruction of the energy of the orginal partons is one of the key requirements for robust measurements based on jets. Due to the high production rate of top quarks at the LHC, the top quark can be used for the first time since its discovery in 1995 as a calibration tool. The energies of the jets can be calibrated directly from data using top quark pair events. This thesis is devoted to the development of a data-driven method to measure the jet energy corrections.

In the first chapter of this thesis, the Standard Model of particle physics is introduced with emphasis on the top quark sector. A motivation for the construction of the LHC is given in Chapter 2 followed by the description of the LHC and one of the general-purpose detectors, the Compact Muon Solenoid (CMS) experiment. Since the LHC became operational only recently, the physics analyses presented in this thesis are developed with simulated proton collisions. The physics models and assumptions on which the simulation is based are discussed in Chapter 3. Physics objects observed in the final state of each interaction need to be reconstructed from the electronic sig-

nals in the CMS detector as described in Chapter 4. Emphasis is put on those objects needed for the analyses, in particular the reconstruction of jets is studied in detail. The selection and reconstruction of the top quark events on which the method relies are summarized in Chapter 5. The method for an inclusive measurement of the jet energy correction factors is explained in Chapter 6, while differential jet energy corrections are the topic of discussion in Chapter 7. Finally, Chapter 8 reports on the results and its applications. Perspectives are given for the current physics program at the LHC and for possible future analyses based on the developed method.

If I could remember the names of all these particles, I'd be a botanist.

Enrico Fermi

Chapter 1

The top quark sector of the Standard Model

Particle physics aims to explain and describe the nature of the universe at it most fundamental level. The Standard Model of particle physics describes the basic constituents of matter and their interactions. Up to date, this framework is the best formalism of our current knowledge of particle physics and is extensively tested by experimental physicists over the last decades. However, it is believed that the Standard Model of particle physics introduced in the first section of this chapter, is an effective theory up to an energy scale of 1 TeV. New physics phenomena are expected at higher energy scales.

The heaviest elementary particle observed in nature and described by the Standard Model is the top quark, discovered about 15 years ago at the Tevatron collider. The top quark has some interesting properties that trigger a broad range of physics studies such as the study of possible new physics phenomena in the top quark sector or the usage of the observed top quarks as a calibration tool. In a second section some specific top quark properties are discussed. Motivations are formulated to perform even more detailed studies of the top quark sector in and beyond the Standard Model.

1.1 The Standard Model of particle physics

The current experimental knowledge of particle physics is well described by the Standard Model [1, 2]. Apart from the excellent agreement between experimental measurements and theoretical predictions, the Standard Model is able to describe three of the four fundamental forces in nature. The idea of unification of these fundamental forces is inspired by the observations and is the basis of the Standard Model.

The building blocks of the Standard Model, namely the fundamental particles and their interactions, are introduced in Section 1.1.1. Using these building blocks the Standard Model is built as a quantum field theory as discussed in Section 1.1.2. In this section, it is explained how interactions are obtained by requiring invariance under local symmetry transformations. The symmetry groups on which the Standard Model is based are listed together with a mechanism that is believed to be responsible for

the masses of the particles. Furthermore, in Section 1.1.3 some shortcomings of the Standard Model are mentioned together with possible extensions of the effective field theory.

1.1.1 Fundamental particles and interactions

Matter is made up of particles with half-integer spin, so-called fermions. There are twelve of these elementary matter particles and each of these particles f, has its corresponding anti-particle $\bar{f}$, with the same quantum properties but an opposite electric charge. These twelve particles can be subdivided into two groups, namely quarks and leptons. Both the leptons and quarks can be grouped in three generations. Each generation is a copy of the first generation with the exception of an increased mass for the corresponding particles in the higher generations. All stable matter observed in the universe is made from the fermions in the first generation. An atom consists of electrons together with protons and neutrons forming the nucleus. A proton consists of two up quarks and one down quark, while a neutron consists of two down quarks and one up quark. The charged fermions in the second and third generation decay to the fermions in the first generation and are only observed in high energy environments. Table 1.1 gives an overview the three generations of fermions together with their electric charge Q.

	generation			
	1	2	3	Q
leptons	electron neutrino ν_e	muon neutrino ν_μ	tau neutrino ν_τ	0
	electron e^-	muon μ^-	tau τ^-	-1
quarks	up u	charm c	top t	2/3
	down d	strange s	bottom b	-1/3

Table 1.1: The three generations of fermions in the Standard Model with their electric charge Q. Each of them has a corresponding anti-fermion particle.

Forces are carried in the Standard Model by integer spin particles, so-called gauge bosons. These bosons, responsible for the mediation of interactions between the fermions, are listed in Table 1.2. The photon is the carrier of the electromagnetic force between

	interaction	mass (GeV/c^2)
photon γ	electromagnetic	0
W^+ and W^-	charged current weak	80.398 ± 0.025
Z^0	neutral current weak	91.1876 ± 0.0021
gluons g	strong	0

Table 1.2: The gauge bosons of the Standard Model and their measured mass value [3].

particles with a non-zero electric charge, while the massive W and Z bosons mediate the weak force. The strong interaction between the quarks is mediated by eight gluons.

Since the W bosons are electrically charged they also couple to the electromagnetic interactions.

There is one last particle in the Standard Model for which there is no experimental evidence yet. The so-called Higgs boson is a spin-0 particle that provides mass to the elementary particles through a mechanism introduced in the next section.

1.1.2 The Standard Model as a quantum field theory

The Standard Model can be described in the framework of quantum field theory. In quantum field theory the particles are associated to fields that depend on the space time coordinates x. The dynamics of the field is determined by an action S

$$S = \int \mathcal{L}(x) d^4x, \tag{1.1}$$

with $\mathcal{L}(x)$ the Lagrangian density. Therefore, to build the Standard Model, a Lagrangian density needs to be found that describes the experimentally observed particles and their interactions. The strategy to obtain this Lagrangian is summarized in this section.

Interactions from symmetry transformations

Consider the case of a free fermion field ψ with mass m. Using the Dirac matrices γ_μ [1], the Dirac equation of motion can be written as [2]

$$i\gamma^\mu \partial_\mu \psi - m\psi = 0. \tag{1.2}$$

If $\bar{\psi}$ is the field representing the anti-fermion, the Dirac Lagrangian is defined as

$$\mathcal{L}_{\text{Dirac}} = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi. \tag{1.3}$$

Following the principle of gauge invariance [4], it is required that $\mathcal{L}_{\text{Dirac}}$ is invariant under a local phase transformation (i.e. a symmetry or gauge transformation)

$$\psi' = U\psi = \mathrm{e}^{ig\vec{\epsilon}(x)\frac{\vec{\tau}}{2}}\psi, \tag{1.4}$$

with rotation parameters $\vec{\epsilon}(x)$ in an internal phase space represented by the generators $\vec{\tau}$. The quantum-mechanical observables, depending on $|\psi|^2$ remain unchanged under this transformation. In order to make the derivative term in $\mathcal{L}_{\text{Dirac}}$ invariant, as many vector gauge fields $\vec{A}_\mu$ as generators $\vec{\tau}$ need to be introduced. The gauge invariant Lagrangian $\mathcal{L}$ is then written as

$$\mathcal{L} = i\bar{\psi}\gamma^\mu \partial_\mu \psi - g\bar{\psi}\gamma^\mu \frac{\vec{\tau}}{2}\vec{A}_\mu \psi - m\bar{\psi}\psi = i\bar{\psi}\gamma^\mu D_\mu \psi - m\bar{\psi}\psi. \tag{1.5}$$

[1] The Dirac matrices are defined by $\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}$ with the Minkowski metric $g^{\mu\nu}$ on space-time. The notation {a,b} denotes the anti-commutator, i.e. $ab + ba$.

[2] Natural units are used, which means $\hbar = c = 1$.

where the covariant derivative D_μ is defined as

$$D_\mu = \partial_\mu + ig\frac{\vec{\tau}}{2}\vec{A}_\mu. \tag{1.6}$$

In Equation 1.5 the term with $\vec{A}_\mu$ corresponds to the coupling between the interacting gauge field and the fermion field. The parameter g is interpreted as the interaction strength or coupling constant. The transformation relations for the components of the field $\vec{A}_\mu$ are derived to be

$$\frac{\vec{\tau}}{2}\vec{A}'_\mu = -\frac{i}{g}U(\partial_\mu U^{-1}) + U\frac{\vec{\tau}}{2}\vec{A}_\mu U^{-1}\,. \tag{1.7}$$

as obtained by requiring that

$$\mathcal{D}'_\mu\psi' = U(\mathcal{D}_\mu\psi), \tag{1.8}$$

such that the Langrangian is invariant for the transformation given in Equation 1.4.

The gauge transformations can be either based on Abelian or non-Abelian groups [3].

Symmetry groups in the Standard Model

In the previous section it was explained that by requiring gauge invariance under symmetry transformations gauge fields (bosons) are introduced that are responsible for the interactions between the particles. Based on the experimental observations the underlying symmetries are proposed to build the theoretical framework of the Standard Model of particle physics.

The Standard Model is based on the gauge group

$$G_{SM} \equiv SU(3)_c \otimes SU(2)_L \otimes U(1)_Y. \tag{1.9}$$

While $U(1)_Y$ is an Abelian group, which introduces a gauge field B_μ, $SU(2)_L$ and $SU(3)_c$ are non-Abelian groups, introducing respectively three field W^i_μ with $(i = 1,\ldots,3)$ and eight fields G^a_μ with $(a = 1,\ldots,8)$. The covariant derivative ensuring local gauge invariance with respect to G_{SM} is written as

$$D_\mu = \partial_\mu + ig_1\frac{Y}{2}B_\mu + ig_2\frac{\tau^i}{2}W^i_\mu + ig_3\frac{\lambda^a}{2}G^a_\mu, \tag{1.10}$$

where g_1, g_2 and g_3 represent the coupling strengths and the parameters Y, τ^i and λ^a denote respectively the hypercharge, the Pauli matrices and the Gell-Mann matrices [4]. The Lagrangian describing the Standard Model arises by writing down the Dirac terms for the fermions ($\mathcal{L}_{\text{Dirac}}$) and by taking into account the covariant derivative of Equation 1.10. In addition for each gauge field a gauge invariant kinetic term needs to be included to describe the propagation of free gauge bosons. The three different groups that form G_{SM} are interpreted as:

[3] An Abelian group is a group represented by commuting generators $\vec{\tau}$, i.e. $[\tau_i, \tau_j] = 0$, while for a non-Abelian group the commutator of the generators does not vanish but is given by a linear combination of the generators.

[4] The summation over i and a is implicit.

- **Quantum chromodynamics ($SU(3)_c$)**
  The symmetry group $SU(3)_c$ is the symmetry group of the strong force with eight massless gauge boson fields G^a_μ or gluons. The gluons interact with all particles carrying the colour charge (therefore the subscript c), i.e. with the quarks and because $SU(3)_c$ is non-Abelian, also with themselves. The leptons do not carry colour charge and are therefore considered as singlets under $SU(3)_c$ transformations.

- **Electroweak theory ($SU(2)_L \otimes U(1)_Y$)**
  To explain the phenomenon of parity violation in the weak interactions, the gauge bosons W^i_μ of $SU(2)_L$ are only allowed to couple with left-handed fermion doublets [5] (hence the subscript L). The Pauli matrices τ^i are related to the generators of the weak isospin space. The symmetry group $U(1)_Y$ introduces the scalar Y denoting the hypercharge defined by $Y = 2(Q - I_3)$, where Q is the electric charge and I_3 the third component of the weak isospin. The boson fields B_μ and W^i_μ correspond, when arranged into appropriate linear combinations to the photon γ (A_μ) and to the Z^0 (Z_μ) and $W^\pm$ ($W^\pm_\mu$) bosons.

$$A_\mu = \sin\theta_{\mathrm{W}} \mathrm{W}^3_\mu + \cos\theta_{\mathrm{W}} \mathrm{B}_\mu, \tag{1.11}$$

$$Z_\mu = \cos\theta_{\mathrm{W}} \mathrm{W}^3_\mu - \sin\theta_{\mathrm{W}} \mathrm{B}_\mu, \tag{1.12}$$

$$W^\pm_\mu = \sqrt{\frac{1}{2}}(W^1_\mu \mp i W^2_\mu), \tag{1.13}$$

  where θ_W is the Weinberg mixing angle, defined as

$$\tan\theta_{\mathrm{W}} = \frac{\mathrm{g}_1}{\mathrm{g}_2}. \tag{1.14}$$

  The $W^\pm$ and Z^0 bosons have non-zero masses. It must however be stressed that the introduction of explicit mass terms in the Lagrangian would break gauge invariance. A procedure to break the electroweak symmetry is needed to obtain non-zero masses for these bosons.

Electroweak symmetry breaking

The gluon and photon are gauge bosons of the strong and electroweak forces. As a consequence of the gauge symmetry the gluon and photon are massless, which is in contrast with the non-zero masses of the electroweak $W^\pm$ and Z^0 bosons, while these are also governed by a gauge theory. The observation of non-zero masses for the $W^\pm$ and Z^0 bosons implies a spontaneous breaking of the electroweak symmetry. A mechanism is formulated, referred to as the Higgs mechanism [5–7], which preserves the gauge invariance under symmetry transformations but which involves a vacuum

[5] The chirality projections of a Dirac field ψ are $\psi_L = \frac{1}{2}(1 - \gamma_5)\psi$ for left-handed chiralities and $\psi_R = \frac{1}{2}(1 + \gamma_5)\psi$ for right-handed chiralities. Right-handed fermions have a spin lined up to the direction of motion, while for left-handed particles the scalar product of the spin and the direction vectors is negative.

state that does not possess this symmetry. In the Standard Model it is assumed that the electroweak symmetry $SU(2)_L \otimes U(1)_Y$ is broken by the introduction of an extra scalar field that is an electroweak doublet

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \tag{1.15}$$

where ϕ^+ and ϕ^0 are complex scalar fields. The gauge invariant renormalizable terms that can be added to the Lagrangian are

$$\mathcal{L}_\Phi = (D^\mu \Phi)^\dagger (D_\mu \Phi) - \mu^2 \left(\Phi^\dagger \Phi\right) - \lambda \left(\Phi^\dagger \Phi\right)^2, \tag{1.16}$$

where the two last terms represent the potential of the scalar field and μ^2 and $\lambda > 0$ represent respectively a mass parameter and the strength of the field's self interaction. By choosing $\mu^2 < 0$ the minimum of the potential is not unique and reaches a non-zero field strength

$$< \Phi^\dagger \Phi >= v^2 = \frac{|\mu^2|}{\lambda} > 0. \tag{1.17}$$

A field configuration in the region around the vacuum is obtained by performing an expansion around a particular choice of the vacuum state

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}, \tag{1.18}$$

where the field $h(x)$ is a quantum fluctuation away from the vacuum state and corresponds to the physical Higgs boson field. Since $h(x)$ is the only gauge invariant degree of freedom in Φ, the symmetry breaking sector gives rise to only one new particle, the scalar Higgs boson. The mass of the Higgs boson is given by $m_H = \sqrt{2\lambda v^2}$. The gauge vector boson fields $W_\mu^\pm$ and Z_μ acquire masses

$$m_W = \frac{1}{2} v \, |g_2| \qquad\qquad m_Z = \frac{1}{2} v \sqrt{g_1^2 + {g_2}^2}. \tag{1.19}$$

Through the spontaneous symmetry breaking mechanism, the $W^\pm$ and Z^0 gauge bosons and the Higgs boson obtain a mass, but the fermions fields remain massless. Masses for the fermions can be generated by adding gauge invariant couplings between the fermions and the Higgs boson field to the Lagrangian density. Such couplings are also known as Yukawa couplings and require the introduction of new coupling constants, which are related to the fermionic masses and are therefore free parameters of the theory.

The Yukawa couplings are also related to the Cabibbo-Kobayashi-Maskawa (CKM or quark mixing) matrix. This matrix arises from the fact that the mass (or strong force) eigenstates of the quarks are not the same as their weak force eigenstates. The CKM matrix provides the transformation from the strong force eigenstates to the weak force eigenstates

$$\begin{pmatrix} d^{weak} \\ s^{weak} \\ b^{weak} \end{pmatrix}_L = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L . \tag{1.20}$$

The elements of the matrix $V_{qq'}$ are related to the probability of a transition from a quark q to a quark q'.

1.1.3 Challenges for the Standard Model

During the last 40 years the Standard Model of particle physics was extensively tested. The only significant deviation from its initial prediction is the existence of non-zero neutrino masses. There is however no conceptual problem to incorporate this recent observation into the theory. Currently, there is still one particle predicted by the Standard Model that is not observed yet, the Higgs boson. Therefore, there is no confirmation of the electroweak symmetry breaking mechanism. In addition, there are other fundamental questions that remain unanswered, like for instance the 19 free parameters with arbitrary values. With the recent observation of non-zero masses for the neutrinos, at least 7 additional free parameters are added for the neutrino sector namely three masses, three mixing angles and at least one phase. Also an explanation for the existence of exactly three fermion generations is missing. One can also wonder why the typical scale for electroweak physics, $m_Z \sim 10^2\,\text{GeV}$, is much smaller than the fundamental scale of gravity, or Planck mass scale $m_{Pl} \sim 10^{19}\,\text{GeV}$. The Standard Model does not foresee new physics up to the Planck mass scale, a problem which is referred to as the hierarchy problem. The hierarchy problem requires an extreme fine-tuning of the constants in the theory, imposing questions on the naturalness of the theory. Connected with the hierarchy problem, another question is posed, namely if the three forces described by the Standard Model can be unified and characterized by a single coupling constant at higher energy scales. Furthermore, it is difficult to combine the Standard Model quantum field theory with general relativity.

The many open questions, the undiscovered Higgs boson and the lack of a candidate for dark matter in cosmology lead to the believe that the Standard Model of particle physics is only an effective field theory, valid up to some physical energy cut-off scale. Several proposals to extend the Standard Model are made in the past decades, which predict new physics phenomena at the TeV scale and provide answers to some of the open questions. Many models as extension of the Standard Model are based on supersymmetry, which invokes new particles with identical quantum numbers as the Standard Model particles except for their spin that differs by half a unit. Since these so-called superpartners are not yet observed the supersymmetry must be broken. Supersymmetry is able to provide an answer to the hierarchy problem, has candidates for dark matter, unifies the three gauge interactions of the Standard Model at high energies and includes a natural mechanism for electroweak symmetry breaking. In other models, the Standard Model Higgs boson is replaced by a fermion condensate induced by a new strong interaction near the TeV scale. Some extensions of the Standard Model introduce extra dimensions of space for various reasons, for instance to explain the fundamental forces, with the exception of gravity (Kaluza-Klein theory) or to explain the weakness of gravity (Randall-Sundrum model).

To formulate an answer to the open questions and to confirm or exclude the existence of new physics phenomena at the TeV scale, new collider experiments at higher energies are required.

1.2 Top quark physics

The observation of the top quark in 1995 [8, 9] by the CDF [10] and D0 [11] experiments at the Tevatron collider [12], marks the beginning of the study of the top quark sector of the Standard Model. Until recently, only the Tevatron collider was able to reach the high energies needed to produce the heaviest particle in the Standard Model [6]. Therefore, all experimental knowledge from direct measurements about top quark physics is obtained by the two Tevatron experiments, which both analyzed up to $5\,\mathrm{fb}^{-1}$ of proton anti-proton collisions at 1.96 TeV.

An overview of the key properties of the top quark together with the status of their experimental measurements is given in Section 1.2.1. Emphasis is put on the current precision of the combined top quark mass measurements, since this is an important aspect for the calibration methods presented in this thesis. The top quark sector can be used to test theories beyond the Standard Model that predict changes in the distributions of observable properties of collision events in which top quarks appear. In addition, the precise determination of the top quark mass provides an indirect constraint on the Higgs boson mass. These two aspects are discussed in Section 1.2.2. In view of the recent start of the data taking at the Large Hadron Collider and hence the expected abundant production of top quark events, one can also envisage the usage of top quark events for calibration purposes. Some examples are summarized in Section 1.2.3.

1.2.1 Key properties of the top quark

The top quarks are produced either in pairs (top quark pair production) by gluon fusion and quark annihilation and thus via the strong interaction, or via the electroweak interaction in single top quark production. The lifetime of the top quark is less than $10^{-24}\,\mathrm{s}$ or about 20 times shorter than the timescale for the strong interaction [13]. Therefore, the top quark decays before it is able to hadronize. Because the top quark is the only quark with this property, it is also the only quark that can be studied as a bare quark. According to the Standard Model, which assumes three generations of fermions, the top quark decays predominantly into a W boson and a b quark (99.8 %). The $|V_{tb}|$ element in the CKM matrix in equation 1.20 is measured from the production cross section of single top quarks at the Tevatron experiments. Using the measured single top quark production cross section corresponding to a top quark mass of $170\,\mathrm{GeV/c^2}$ and a theoretical cross section normalization given by $\sigma^{\mathrm{theory}} = 3.14 \times |V_{tb}|^2\,\mathrm{pb}$, the combined direct measurement results in $|V_{tb}| = 0.91 \pm 0.08$ [14].

In this thesis top quark pair production is considered. Top quark pair or $\mathrm{t\bar{t}}$ events decay mainly as $\mathrm{t\bar{t}} \rightarrow \mathrm{bW\bar{b}W}$. They are usually categorized according to the decay mode of the W bosons, which decay either hadronically with a corresponding branching ratio $\mathcal{B}(W \rightarrow q\bar{q}) \simeq 2/3$ or leptonically with a corresponding branching ratio $\mathcal{B}(W \rightarrow l\bar{\nu}_l) \simeq 1/3$. The decay mode in which the two W bosons decay leptonically is referred to as the fully-leptonic decay and similarly, when both W bosons decay hadronically

[6] At the moment of writing, the Large Hadron Collider started operating at even higher center of mass energies, but no observation of the top quark has been claimed yet.

the decay mode is denoted as fully-hadronic. The decay mode in which one of the two W bosons decays leptonically and the other hadronically is the semi-leptonic or lepton+jets decay channel. Since the tau lepton is able to decay hadronically, top quark pair decays involving tau leptons are often considered as a separate category. The branching ratio for the decay of the W boson in leptons of a certain generation is the same for the three lepton generations. Therefore, the probability that a top quark pair decays semi-leptonically involving leptons of either the first, second or third generation is calculated for each lepton flavour as $2 \times 2/3 \times 1/3 \times 1/3 = 4/27$ [7] or about $\sim 14.8\,\%$ of the top quark events. For the studies presented in the next chapters, the semi-muonic $t\bar{t}$ decay is considered $t\bar{t} \rightarrow bW\bar{b}W \rightarrow bq\bar{q}\bar{b}\mu\bar{\nu}_{\mu}$.

Another important property of the top quark is its mass. In Figure 1.1 the combined top quark mass measurement of the Tevatron experiments is shown. The top quark

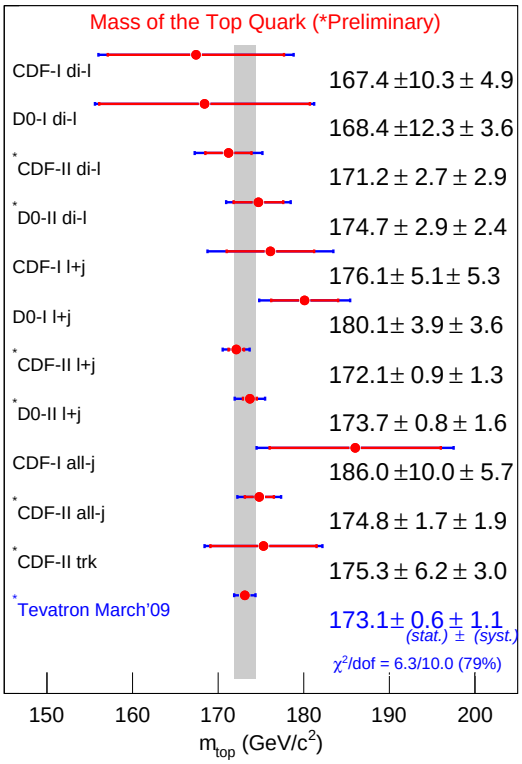


Figure 1.1: The combined CDF and D0 top quark mass measurement.

mass is measured to be $m_t = 173.1 \pm 0.6(stat) \pm 1.1(syst)\,\mathrm{GeV}/c^2$. When adding the uncertainties in quadrature, the total uncertainty is below $1.3\,\mathrm{GeV/c^2}$ and hence a precision of about $0.7\,\%$ is obtained [15]. The top quark mass measurement is already dominated by the systematic uncertainty. Therefore, efforts are ongoing to improve and unfold the sources of systematic uncertainty. One of the main systematic uncertainties is the jet energy scale (of the b quark jets). The statistical uncertainty will be further

[7] The factors 2/3 and 1/3 are related to the probability for respectively a hadronic decay of the first W boson and a leptonic decay of the second W boson, while the factor 2 takes the interchange of the first and second W boson into account. A last factor 1/3 splits the inclusive leptonic decay of the W boson into the three exclusive lepton flavours.

reduced when data sets with higher integrated luminosities become available. Three main different measurement techniques are deployed to measure the top quark mass at the Tevatron, namely the template method, the matrix element method and an ideogram or kinematic fitting method [16]. These three different techniques are shortly addressed:

- **Template method**
  The idea behind this method is to use observables that are correlated to the mass of the top quark as an estimator for this mass. An estimate of the top quark mass is obtained by comparing the differential distribution of the reconstructed observable in data with different expected template distributions for various assumed values of the top quark mass. The template distributions are obtained from simulated signal and background events.

- **Ideogram method**
  The precisely measured W boson mass can be used to constrain the measured $t\bar{t}$ event kinematics. A kinematic fitting technique can be used to force the W boson mass hypothesis in the $t\bar{t}$ decay on an event by event basis before reconstructing the top quark mass spectrum. The kinematic fit provides an event by event probability $P(m_t)$ that the reconstructed top quark candidate has a mass m_t. The top quark mass is estimated from these probabilities.

- **Matrix element method**
  This method yields the most precise results in the semi-leptonic channel at the Tevatron experiments. For each reconstructed top quark event the method employs all information to construct an event by event likelihood as a function of the reconstructed top quark mass. Rather then using mass constraints in the kinematic fit to the hypothesis of m_t, the full matrix element is calculated for different values of m_t. A transfer function is used to relate the final-state momentum configuration to the measured quantities in the detector. The total likelihood of the sample is obtained as the product of the individual event likelihoods. The method uses significantly more computation time that the template method, but the available information is optimally used. Usually assumptions are made to simplify the determination of the matrix elements in order to reduce the computation time.

While the statistical precision improves when more information is used, the ideogram and matrix element methods are more sensitive to the uncertainty on the jet energy scale. To reduce this uncertainty, the jet energy scale can be measured in situ, through the hadronic decay of the W boson present in the top quark decay. Currently, the most precise results are obtained with the semi-leptonic decay channel using the matrix element method.

1.2.2 Top quark physics as a probe for new phenomena

The top quark mass is related to the Higgs boson mass through radiative corrections. Therefore, a precise measurement of the top quark mass together with the precisely

measured electroweak parameters in the Standard Model, provides an indirect measurement of the Higgs boson mass. This is shown in the left plot of Figure 1.2. The 68 % contour of the χ^2 fit to the electroweak data with the exception of the top quark mass measurement, assuming the Standard Model to be correct, is shown as a function of the top quark mass and the Higgs boson mass. The measured top quark mass obtained from the Tevatron experiments is also shown together with the region of the Higgs boson mass that is already excluded by direct searches [17]. The right plot in Figure 1.2 shows the result of the χ^2 fit when the top quark mass measurement is included. The region for a Higgs boson mass $m_H < 114\,\mathrm{GeV/c^2}$ was excluded by the

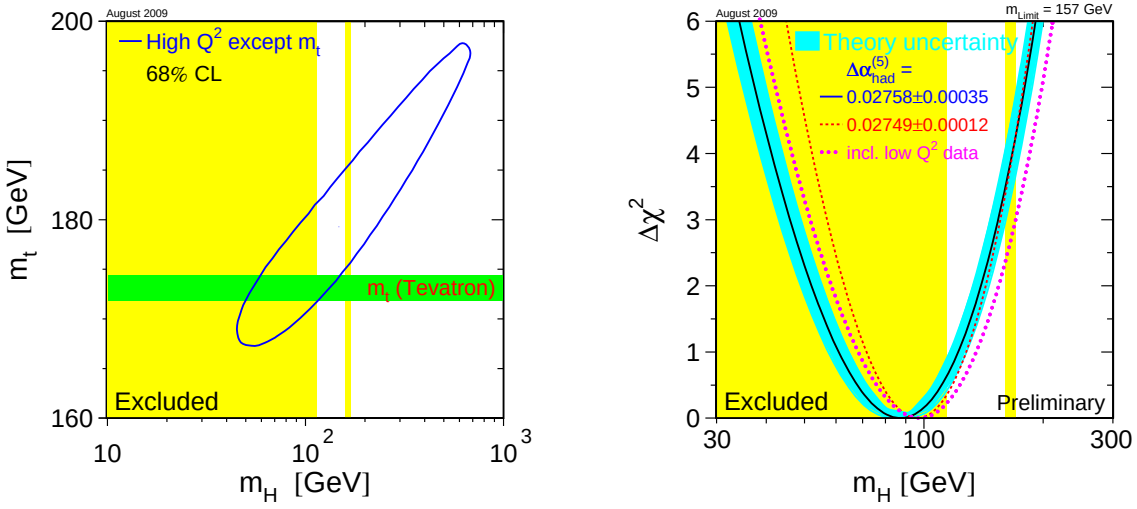


Figure 1.2: The 68 % contour from a Standard Model fit to the electroweak data shown as a function of the top quark mass and the Higgs boson mass (left). The Tevatron result for the top quark mass is also shown as a 68 % contour. The $\Delta\chi^2$ as a function of the assumed Higgs boson mass resulting from fits to data assuming the Standard Model is correct (right). The exclusion regions from direct searches are shown.

LEP experiments, while recently the region $160 < m_H < 170\,\mathrm{GeV/c^2}$ is excluded by the direct searches at the Tevatron experiments. The solid black line indicates the central value while the blue band corresponds to the theoretical uncertainty. The inclusion of the low Q^2 NuTeV data, which might show a slight deviation compared to other experiments, is shown by the dotted line. The precision data depend on the extrapolation of the coupling constant characterizing the strength of the electromagnetic interaction from its measured value at low energy up to high energy. In particular, an uncertainty is associated with the contribution of low energy QCD to the extrapolation. The effect on the χ^2 fit of using a different value for the low energy QCD, denoted by $\Delta\alpha^{(5)}_{had}$, is indicated by the dashed line. From the precision electroweak measurements the mass of the Standard Model Higgs boson is constrained to be lower than about $157\,\mathrm{GeV/c^2}$ (one-sided 95 percent confidence level upper limit including both the experimental and the theoretical uncertainty). The limit is increased to $186\,\mathrm{GeV/c^2}$ when the lower limit of the LEP experiments is taken into account.

The Higgs boson mass can be constrained by the top quark mass because of the large Yukawa coupling between the two elementary particles. Due to this large coupling it is believed that the top quark may play a special role in the mechanism of spontaneous symmetry breaking or is special in another way. Therefore, top quark physics is considered to provide a window on physics beyond the Standard Model. New physics phenomena may alter the top quark couplings, modify the top quark production and decay or affect the top quark properties. In many theories heavy resonances are predicted that decay almost exclusively into a pair of top quarks affecting the top quark pair invariant mass distribution [18].

Another example is coming from the possible existence to include a fourth generation of fermions [19]. The pair production of heavy top-like t' quarks would then mimic the signal of a top quark pair event and induce an effect in the top quark pair invariant mass distribution.

Supersymmetry requires a superpartner for every particle. After a soft breaking of the supersymmetry these superpartners obtain masses heavier than their Standard Model counterpart. However, the superpartner of the top quark, namely the scalar top or stop quark can be lighter than the top quark itself [20]. The reason why it escaped detection so far is due to the lower pair production cross section for the scalar particle compared to its fermionic superpartner. The decay of the pair produced stop quarks would be similar to the final state of pair produced top quarks, but the lepton branching ratio would be significantly enhanced over the Standard Model prediction due to the presence of chargino's and neutralino's.

In many theories beyond the Standard Model, the existence of an additional Higgs doublet is required and multiple physical Higgs particles are therefore predicted, including two charged Higgs bosons [21]. The charged Higgs boson can appear in the top quark decay $t \rightarrow H^+ b$, provided the mass of the H^+ boson is sufficiently small. If the mass of the charged Higgs boson is however sufficiently large, the charged Higgs boson could decay as $H^+ \rightarrow t\bar{b}$. In both cases the yield of top quark events would be different and therefore the study of top quark events can be used as a probe for charged Higgs bosons.

It is worth mentioning in this context that top quark events are a background for many searches beyond the Standard Model at hadron colliders and hence the understanding of top quark physics is crucial for the discovery of new physics phenomena.

1.2.3 The top quark as a calibration tool

When the Large Hadron Collider described in Chapter 2, is operating at its design luminosity and center of mass energy, about one top quark pair will be produced per second. Apart from the detailed measurements of the top quark properties, the huge amount of events will allow to use top quark events for calibration and commissioning purposes. In particular semi-leptonic $t\bar{t}$ decays are of interest since they require the use of about all reconstruction techniques to obtain the kinematics and to identify the objects in the observed final state. As an example, the accuracy of missing transverse energy calculations can be studied with these events. The trigger efficiencies can be measured from data by using for instance jet triggers to measure the efficiency of isolated lepton triggers. Assuming the Standard Model prediction $\mathcal{B}(t \rightarrow bW) \sim 1$,

the heavy flavour content of $t\bar{t}$ events is well predicted, which allows to calibrate and measure the efficiency of b-tagging algorithms and jet energy corrections from the produced top quark pair events [22–27]. Both the detector commissioning as well as the measurement of b-tagging efficiencies and jet energy corrections from data are crucial for the searches for new physics phenomena. Hence, methods are developed to obtain a fully commissioned detector and well calibrated reconstructed objects in the final state. This thesis is dedicated to the measurement of the jet energy corrections from top quark pair events. The method will use the inverse of the ideogram method used to measure the top quark mass at the Tevatron. Rather then using the observed jet energy scales as an input for the ideogram method to obtain the most probable value of the top quark mass, the top quark mass measured at the Tevatron will be used as an input to estimate the most probable jet energy corrections that fit the data at the Large Hadron Collider.

The true worth of an experimenter consists in pursuing not only what he seeks in an experiment, but also what he did not seek.

Claude Bernard

Chapter 2

Experimental setup

The predictions of the Standard Model of particle physics are in agreement with the many precision measurements performed over the past decades. Despite the long series of successes, there is one building block for which experimental evidence is still missing. This building block is related to the mechanism of electroweak symmetry breaking, a process through which all the elementary particles receive their masses. The main physics motivation to build the Large Hadron Collider (LHC) is the search for phenomena that could give insight into this mechanism. The LHC is the highest-energy accelerator in the world and allows searches for physics at the TeV scale.

In Section 2.1 the design of the LHC is discussed together with its physics programme. One of the detectors at the LHC is the Compact Muon Solenoid (CMS) experiment, described in Section 2.2. For a successful operation of the experiments at the LHC, dedicated computing facilities are essential and specific structures were developed as discussed in Section 2.3.

2.1 The Large Hadron Collider

The LHC is a proton-proton accelerator built at CERN (European Organization for Nuclear Research) in the existing 26.7 km long tunnel between 50 and 175 meters underground that housed the Large Electron Positron Collider until 2000. The accelerator complex is located on the border between Switzerland and France between the Geneva lake and the Jura mountains. Beside the collisions of protons, the LHC is also designed to collide heavy ions. After the approval of the LHC project in 1994 many technological innovations were needed to accommodate the design criteria. Since November 2009, the LHC is the world's highest-energy particle accelerator. In March 2010, two proton beams collided for the first time at a center of mass energy of 7 TeV.

The design and machine layout are described in Section 2.1.1 while the different experiments and their physics programme are discussed in Section 2.1.2.

2.1.1 The LHC design and operation

The study of physics at the TeV scale requires the highest possible beam energy. With a design beam energy of 7 TeV at the LHC, the beam energy is seven times higher than the beam energies reached at any other hadron collider experiment. Additionally, also the design luminosity of $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ is two orders of magnitude higher than the other hadron collider experiments. The production rate per second for a physics process of interest is given by the product of the luminosity $\mathcal{L}$, and the production cross section of the physics process σ_{process}

$$N = \mathcal{L}\sigma_{\text{process}}. \tag{2.1}$$

Therefore, to study phenomena with a low production cross section, the luminosity of the collider should be as high as possible. The luminosity for head-on collisions is given by

$$\mathcal{L} = \frac{f_{\text{rev}} N_p^2 k}{4\pi\sigma_x\sigma_y}, \tag{2.2}$$

where f_{rev} is the revolution frequency, N_p the number of particles in the colliding bunches, k the number of bunches and σ_x, σ_y the transverse size of the bunches at the collision point. To obtain a high luminosity, many particles in each bunch are needed and a maximal number of bunches, while the transverse size of the bunches in the interaction point should be minimal.

The decision to build the LHC in the existing circular LEP tunnel was based on cost saving arguments [28]. For circular accelerators, dipole magnets are crucial to keep the collision particles on track. One of the effects when particles in the ultra-relativistic energy regime are bent, is synchrotron radiation. The particles lose during each revolution an amount of energy corresponding to

$$\Delta E \propto \frac{E^4}{Rm^4}, \tag{2.3}$$

where R is the radius of the accelerator, m the mass of the particle and E the energy of the particle in the revolution. Since the radius of the accelerator was fixed by the choice of using the former LEP tunnel, the amount of energy loss due to synchrotron radiation when operating at higher energies can only be limited by using particles with a higher mass then electrons. Therefore, it was decided to use protons, since their mass is about 2000 times higher compared to the mass of electrons. The decision to build a proton proton collider instead of a proton anti-proton collider such as the Tevatron collider was driven by the wish to reach a high luminosity, because one is not limited by difficulties in producing large amount of anti-protons. For a proton anti-proton collider, the same magnetic field is used to keep the particles in orbit. This is not possible when using two proton beams. Given the limited diameter of the tunnel of about 4 m, two separate rings with dipole magnets for the counter-rotating beams was not an option. Hence, a special design of "two-in-one" dipole magnets was adopted [29]. Additionally, super conducting magnets are required to obtain the high magnetic field of 8.33 T necessary to keep the 7 TeV beams on track. A schematic view of the cross section of a LHC dipole magnet is shown in Figure 2.1. The cold mass or iron yoke around the two apertures in which the proton beams circulate is operated at 1.9 K in

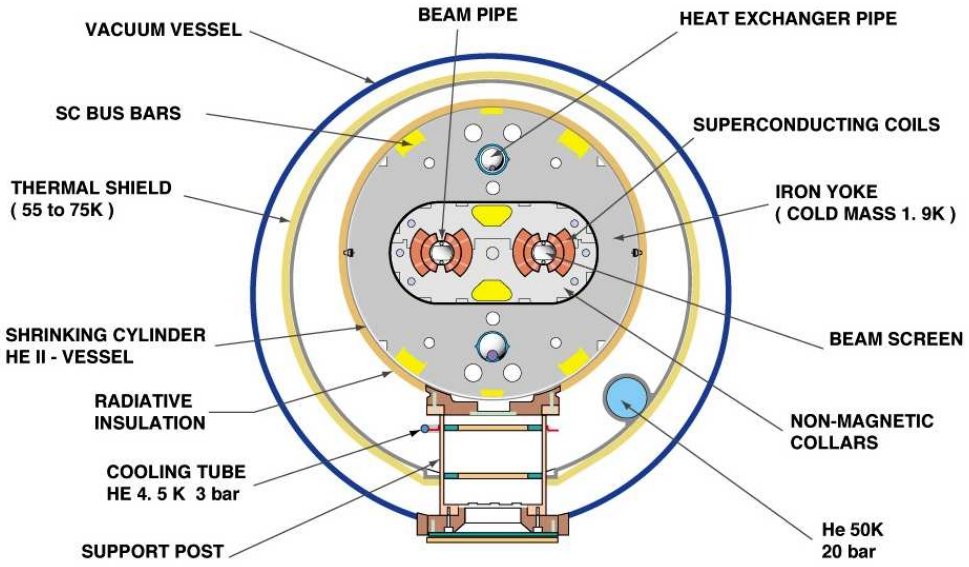


Figure 2.1: Schematic view of the cross section of a LHC dipole magnet.

superfluid Helium. In the two apertures non-magnetic collars are placed that contain the superconducting coils surrounding the beam pipes. The total length of one dipole magnet is about 15 m and its total mass is about 27.5 ton. The LHC contains 1232 superconducting dipoles through which a current of about 12 kA is sent to obtain the 8.33 T magnetic field.

Before the proton beams are injected in the LHC, they go through the complete pre-accelerator complex of CERN, shown in Figure 2.2. The protons are first accelerated

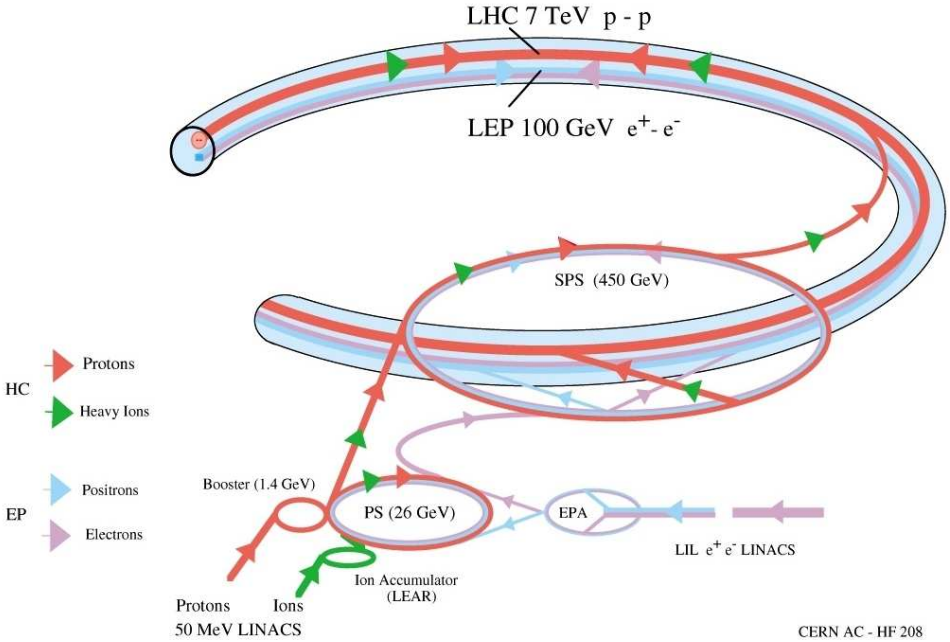


Figure 2.2: The injection chain of the LHC accelerator complex.

to an energy of 50 MeV with a linear accelerator and subsequently fed to the Booster which produces the first bunches and brings the protons to an energy of 1.4 GeV. These bunches of protons are then injected in the Proton Synchrotron (PS), which in turn accelerates the protons to an energy of 26 GeV and delivers bunches 25 ns spaced in time to the Super Proton Synchrotron (SPS). The SPS accelerates the protons to an energy of 450 GeV and injects 2808 bunches in the LHC. In total there is space for 3654 bunches separated by 25 ns in the LHC, but a few beam gaps are present. A 3 μs gap is foreseen in the bunch pattern to be able to dump the beam in one revolution, corresponding to the time needed for a rise in the magnetic field of the beam-dump magnets. Some additional $< 1\,\mu$s gaps exist corresponding to the rise time for the magnets that take care of the injection at different levels, for instance between the PS and SPS or between the SPS and LHC. The LHC is then responsible for the acceleration of the 450 GeV protons delivered by the SPS to the desired beam energy of 7 TeV. Initially, the LHC is operating at a beam energy of 3.5 TeV. The time needed to fill the LHC beams is about seven minutes and the beams circulate between a few hours to about ten hours, until the luminosity decreased too much and the beams are dumped.

The tunnel geometry in which the LHC is installed, consists of eight arcs and eight straight sections, and is shown in Figure 2.3. On four of the eight straight sections

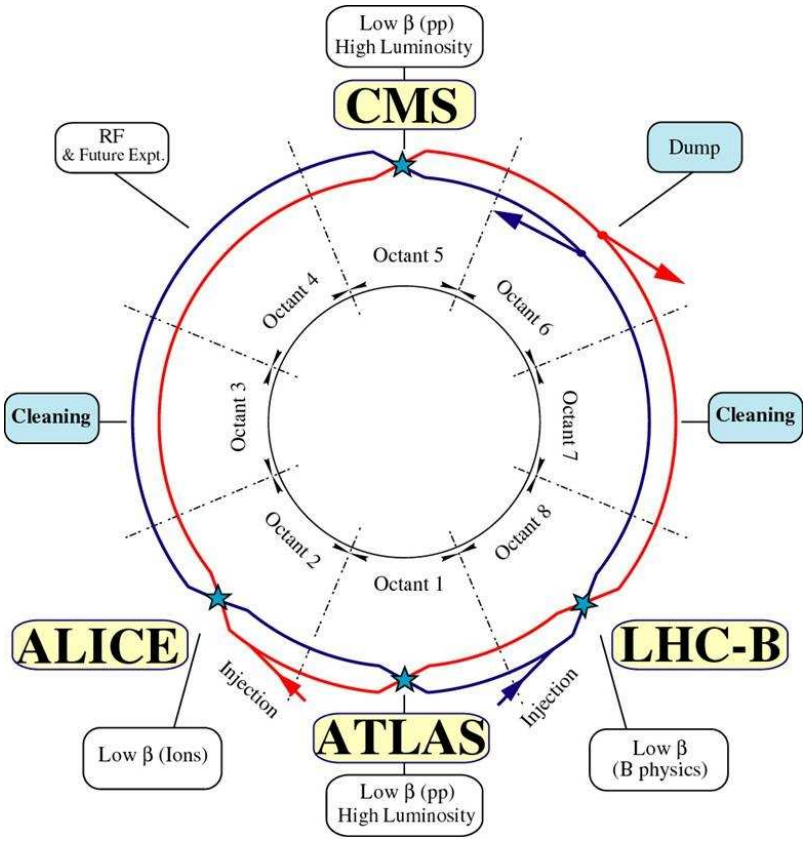


Figure 2.3: Schematic layout of the Large Hadron Collider.

the beams cross each other, while the remaining four straight sections do not have beam crossings. The β function is an important function in accelerator physics and of particular interest in the interaction points where the beam crossing takes place. It is

defined as the envelope around all the trajectories of the beam particles oscillating in the successive quadrupoles. In the interaction points, the β function should be small, consistent with focussed beams, to enhance the luminosity. The β function is therefore closely related to the transverse width of the beams in Equation 2.2. In the figure the location of the four main LHC experiments at the four interaction points is also shown.

2.1.2 The LHC physics programme and experiments

The largest part of the physics programme at the LHC will be conducted with the two general purpose detectors ATLAS [30] and CMS [31]. One of the main objectives of the physics programme is the search for the Higgs boson that is predicted to be responsible for the electroweak symmetry breaking mechanism as described in Chapter 1. In the same chapter the possible existence of new physics at the TeV scale was also discussed. The search for supersymmetric particles, new massive vector bosons and extra dimensions is clearly part of the physics programme to be conducted with ATLAS and CMS. Apart from the searches for new phenomena, the programme also aims at a deeper understanding of the already discovered Standard Model particles, such as the top quark. Additionally, with the LHCb experiment [32], precise measurements of CP violation and rare decays are performed. The LHC is also able to collide heavy ions, such as for example lead ions. The ALICE [33] experiment is designed to study the lead collisions in order to observe the state of hot nuclear matter (quark-gluon plasma). CMS and ATLAS will also study both CP violation and heavy ion collisions. There are two more experiments, TOTEM [34] and LHCf [35]. The TOTEM experiment is designed to measure the total proton-proton cross section with an accuracy of 1 %. Also elastic scattering and diffractive proton collisions will be studied with TOTEM. The LHCf experiment studies the forward production of neutral particles in the proton collisions at low angles. The results are then used as input to model cosmic ray air showers within the Earth atmosphere.

When the LHC operates close to the design luminosity, the total amount of data accumulated each year corresponds to an integrated luminosity L

$$L = \int \mathcal{L} dt \tag{2.4}$$

of about $100\,\mathrm{fb}^{-1}$. During the current run in 2010 and 2011, the LHC operates at lower luminosities and at a center of mass energy of 7 TeV. The aim is to accumulate about $1\,\mathrm{fb}^{-1}$ of data during this run. The physics programme consists in re-establishing the precise observation of Standard Model particles and all searches of new particles and phenomena possible with a low integrated luminosity.

2.2 The Compact Muon Solenoid experiment

At a center of mass energy of 14 TeV, the total proton-proton cross section is about 100 mb. Therefore, at the design luminosity of the LHC, the CMS and ATLAS experiments will observe about 10^9 events per second. This very high event rate puts stringent requirements on the design of the experiments. Due to the short time interval of 25 ns

between the bunch crossings, a fast online event selection process (trigger) is needed to reduce the event rate to about 100 events per second which can be stored and used for offline analysis. At the design luminosity, about 20 inelastic collisions are expected in the same bunch crossing. In order to be able to disentangle the detector signals belonging to the same interaction, detectors are needed with a high granularity and a good time resolution. A last requirement on the detectors is the radiation hardness of the detector material and the front-end electronics. In particular the subdetectors close to the beam pipe will be suffering from high radiation levels.

2.2.1 Overall detector concept

In Figure 2.4 a schematic layout of the CMS experiment is shown. The overall concept

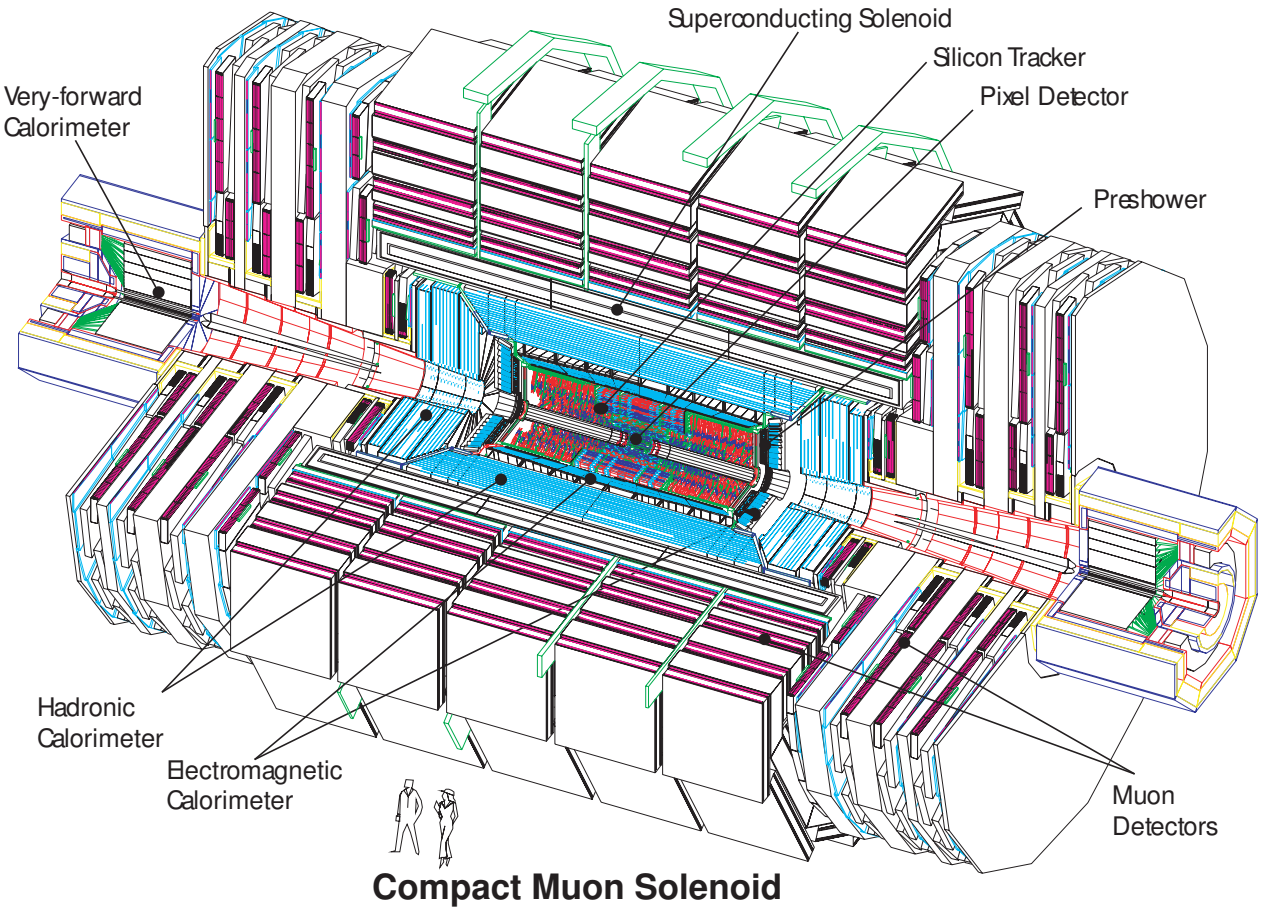


Figure 2.4: Schematic layout of the Compact Muon Solenoid experiment.

on which the CMS detector is based, relies on the magnetic field configuration needed for the precise measurement of the momenta of muons. A large bending power and therefore a large magnetic field is required. The superconducting solenoid developed for the CMS detector has a length of 12.5 m and a diameter of 6.3 m with an inner diameter of the coil of 5.9 m. The magnet weighs 220 tons and is designed to produce a field of 4 T at a nominal current of about 19 kA, corresponding to a stored energy of

2.6 GJ. The obtained magnetic field strength is 3,8 T. The CMS experiment consists of a central barrel part and two endcaps. The full silicon-based inner trarcking system is located close to the interaction region and consists of a pixel detector surrounded by the main silicon strip tracker. The fully active scintillating crystals-based electromagnetic calorimeter and the brass/scintillator sampling hadron calorimeter are located within the superconducting solenoid. The magnetic field lines are closed by the iron return yoke, between which muon detectors are placed. The iron return yoke has a total weight of 10000 tons. To make the detector more hermetic, a very forward calorimeter is placed in the endcaps along the beampipe.

Compared to the dimensions of the ATLAS experiment, the CMS detector is considered compact because it is half the size. The CMS detector has a length of 21.6 m and a diameter of 14.6 m. The total weight of the CMS experiment is 12500 tons. The following coordinate conventions are used in the CMS collaboration. With the origin of the coordinate system centered at the nominal interaction point in the experiment, the y-axis points vertically upwards. The x-axis is chosen to point towards the center of the LHC and the z-axis is chosen along the beam direction in such a way that a righthanded coordinate system is obtained. The azimuthal angle ϕ is measured from the x-axis to the y-axis in the $(x \times y)$ plane. The polar angle θ is measured from the z-axis. The transverse momentum and energy are obtained using the components transverse to the beam direction and are therefore calculated in the $(x \times y)$ plane. A variable that is commonly used at hadron colliders is the pseudo-rapidity defined as $\eta = -ln\tan(\theta/2)$. This variable is Lorentz invariant for boosts along the beam direction.

The CMS detector is developed to meet the goals of the LHC physics programme, for which a good object identification is crucial. A high momentum resolution for charged particles and the possibility to determine the charge of particles with a high transverse momentum is required. For the reconstruction of the (interaction) vertices a highly granulated inner tracking system is needed. Furthermore, to detect particles that do not interact with the experiment, full geometric coverage is required to reconstruct the transverse energy of the escaping particles from a transverse energy balance. In the following sections the layout of the different subdetectors in the CMS experiment is discussed.

2.2.2 Inner tracking system

The full silicon-based inner tracking system has a length of approximately 5.4 m and a radius of about 1.1 m. It consists of a pixel detector close to the interaction region and the surrounding silicon strip detector.

A schematic layout of the pixel detector is shown in Figure 2.5. The pixel detector consists of three barrel layers and two endcap disks on each side. The three barrel layers with a length of about 53 cm in total length are located between 4 cm and 11 cm from the interaction point. The two endcap disks on each side are extending from 6 to 15 cm. The size of the individual pixels is $100 \times 150\,\mu m^2$. The spatial resolution on a single hit is measured to be about $10\,\mu$m for the $r \times \phi$ measurement and $20\,\mu$m for the measurement along the z-axis.

Figure 2.6 shows 1/4 of the longitudinal view of the layout of the silicon strip

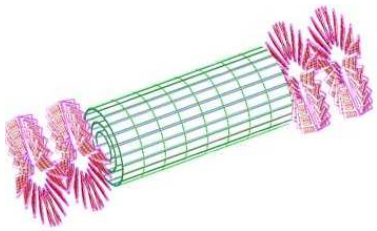


Figure 2.5: Schematic layout of the silicon pixel detector.

tracker. The barrel part of the silicon strip detectors are placed at distances of 20 cm to 110 cm from the beampipe. Two parts can be distinguished, a Tracker Inner Barrel (TIB) consisting of four layers with an extent to $|z| < 65$ cm and the Tracker Outer Barrel (TOB) containing six layers located between $-110 < z < 110$ cm. At each side of the four TIB layers, there are three minidisks, the Tracker Inner Disks (TID). The Tracker Endcap (TEC) is placed at each side of the TOB and consists of nine disks in each endcap covering the region $120 < |z| < 280$ cm. To keep the occupancy low and depending on their distance from the interaction point, the detector modules in the different barrel layers and disks have different sizes and contain different number of strips. The TIB provides a spatial resolution on a single hit of about $23 - 34\,\mu$m for the $r \times \phi$ direction and $230\,\mu$m for in z, while this increases to respectively $35 - 52\,\mu$m and $530\,\mu$m for the TOB.

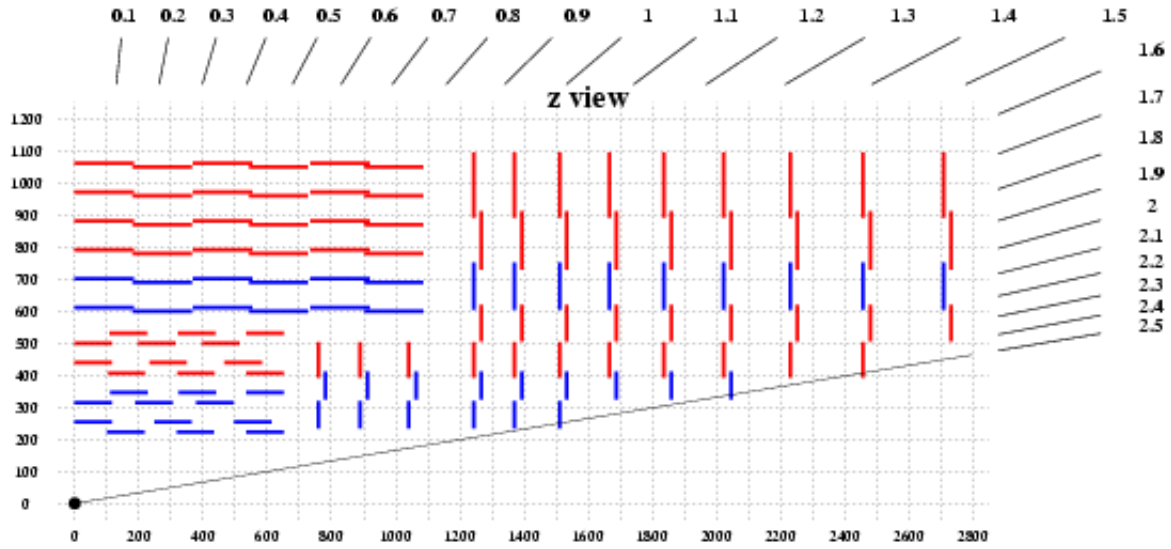


Figure 2.6: Schematic layout of the silicon microstrip detector.

The inner tracking system consists in total of 66 million pixels and 9.6 million silicon strips. The complete tracker system is cooled to an operating temperature of about -20 °C. The inner tracking system provides coverage up to $|\eta| \approx 2.5$.

Track reconstruction

The track reconstruction using the CMS detector consists of several steps. Hits are reconstructed by clustering strip and pixel signals produced by charged particles and estimating their position and uncertainty. Using the reconstructed tracker hits, four main steps can be distinguished to determine the helix trajectories of the tracks: seed generation, trajectory building or pattern recognition followed by the fitting of the final track [36].

In the seeding stage initial track candidates are provided for the full track reconstruction. The initial track candidates or seeds define the initial five parameters of the helix trajectory and their uncertainty. Seeds are obtained by combining pairs of hits in the silicon pixel or silicon strip detector. Due to the low occupancy and excellent position resolution, the seeds obtained with the pixel detector are better than seeds reconstructed by making use of the strip detector only. The best seeds in the endcap region are however obtained from the combined information of the pixel and microstrip detector. Seeding efficiencies of 99 % are reached [37].

The pattern recognition is based on the Kalman Filter approach [38]. The initial trajectory parameters from the seeds and their uncertainties are extrapolated outwards to the next layer. To reduce the number of possibilities, only compatible layers are considered by taking into account the uncertainty on the parameters of the trajectory. The Kalman Filter is a succession of alternating prediction and filtering steps. Hits are assigned to the extrapolated track based on the χ^2 value between the predicted and measured positions. After a new hit is assigned to the trajectory, the trajectory parameters and their uncertainties are updated. Ambiguities can arise when several tracks share one or more hits. To avoid double counting of tracks, the ambiguities are resolved based on the fraction of shared hits between the two trajectories.

In the last stage, the final estimate of the parameters of each helix track are obtained by fitting the collection of hits assigned to a trajectory using a least-squares fit in two steps. An inside-out fit is performed first to remove biases of the seeding and finding stages, followed by a smoother outside-in fit yielding the final best estimates of the track parameters at the vertex.

A resolution on the transverse momentum p_T better than 1 % is obtained for tracks with $p_T \leq 10\,\mathrm{GeV/c}$ in the central region $\eta < 1$ of the tracker [39]. At higher transverse momenta, the relative resolution becomes worse. In the forward region lower track reconstruction efficiencies and worse resolutions are observed, due to the lower hit resolution in the forward region.

2.2.3 The CMS calorimeter system

The calorimeter of CMS consist of an homogeneous electromagnetic calorimeter (ECAL) for the precise measurement of the energy of photons and electrons and a sampling hadron calorimeter (HCAL) surrounding the ECAL for the measurement of the energy of charged and neutral hadrons. The energy of a particle reaching the calorimeter subsystems is measured from the particle shower it initiates in these subdetectors.

A schematic layout of the ECAL is shown in Figure 2.7. The barrel section (EB) and the endcaps (EE) contain together about 76000 lead tungstate ($PbWO_4$) scintillating

crystals. Lead tungstate was chosen for the high stopping power and radiation hardness.

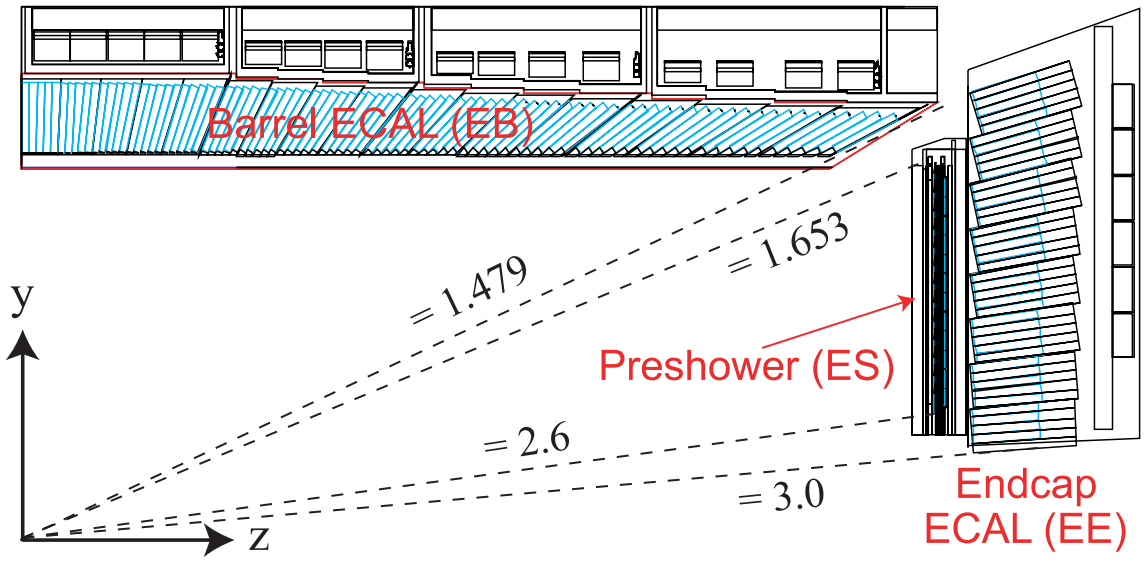


Figure 2.7: Longitudinal view of a quarter of the ECAL layout.

In addition, lead tungstate crystals have a fast response, since 80 % of the light is emitted within 25 ns. The light emitted by the crystals is detected by photodiodes. The barrel section covers the range in pseudo-rapidity of $0 < |\eta| < 1.479$. The crystals in the barrel have a length of 230 mm and point toward the interaction region, but are tilted by 3° in both θ and ϕ to minimise the energy loss of particles traversing the EB exactly between two crystals. The crystals cover 0.0174 radians (or 1°) in both $\Delta\theta$ and $\Delta\phi$, which corresponds to a cross section of $22 \times 22\,\mathrm{mm}^2$. The ECAL endcaps cover a range in pseudo-rapidity of $1.479 < |\eta| < 3.0$ and are located at a distance $|z| = 314\,\mathrm{cm}$. The crystals have a cross section of $28.6 \times 28.6\,\mathrm{mm}^2$ and a length of 220 mm. The length of both the barrel and endcap crystals correspond to about 25 radiation lengths. The intrinsic energy resolution of the ECAL measured in test beams can be parametrized as

$$\left(\frac{\sigma(E)}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2, \tag{2.5}$$

where S is the stochastic term, N the noise and C the constant term. When the energy E is given in GeV, $S = 2.8\,\%$, $N = 0.12$ and $C = 0.30\,\%$ [31]. This corresponds to an energy resolution better than 0.5 % for 100 GeV particles. In front of the endcaps, between the tracker and the EE the preshower detectors are located. The preshower is mainly used for the identification of π^0 particles that decay into two photons, but in addition also the electron identification and both the electron and photon position determination are improved.

The hadron calorimeter is important for the measurement of hadron jets and the missing transverse energy. The barrel section of the HCAL (HB) is located between the ECAL and the solenoid. Due to the restricted space, it can happen that the complete hadronic shower is not absorbed by the HB. Therefore, an outer hadron calorimeter

(HO) is placed outside the solenoid to complement the HB. The endcaps of the HCAL (HE) provide coverage up to $|\eta| = 3.0$ and are complemented by the forward HCAL (HF), which extends the pseudo-rapidity coverage up to $|\eta| = 5.2$. The HCAL is a sampling calorimeter, which means that layers of absorber material producing the particle shower are alternated with layers of active material measuring the energy of the particles produced in the shower. As absorbing material brass is chosen, while for the active material about 70000 plastic scintillators tiles are used. The Figure 2.8 shows the segmentation of the HCAL into calorimeter towers. The HB detector of the

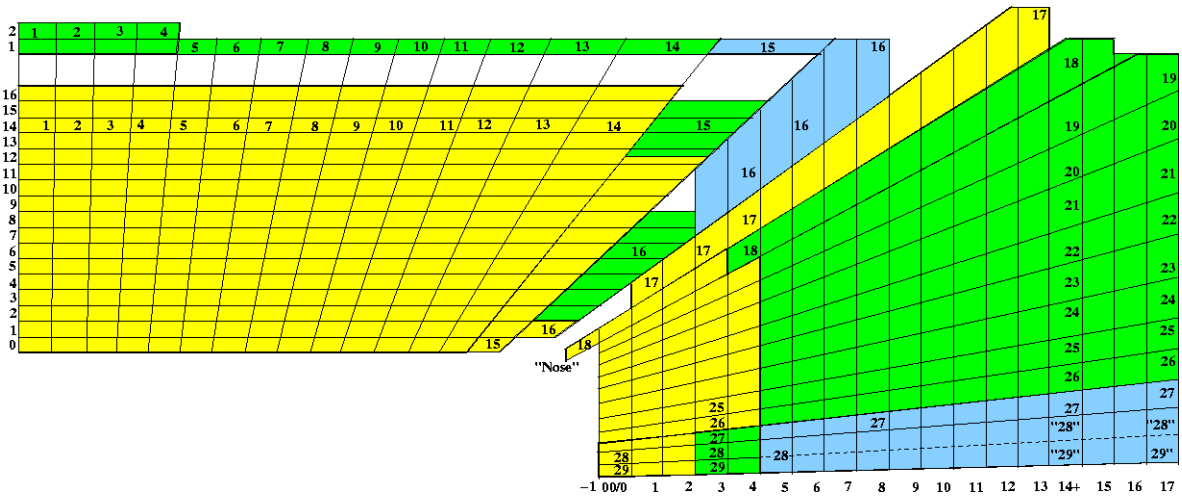


Figure 2.8: Longitudinal view of a quarter of the HCAL layout.

HCAL covers pseudo-rapidity ranges up to $|\eta| = 1.4$. The towers have a segmentation of $\Delta\phi \times \Delta\eta = 0.087 \times 0.087$ (i.e. $\Delta\phi = 5°$). The endcaps provide coverage for pseudo-rapidities between 1.3 and 3.0. In the region between $1.74 < |\eta| < 3.0$ the segmentation is more coarse with $\Delta\phi = 10°$ and $\Delta\eta$ ranging between 0.09 and 0.35. The HO part covers $|\eta| < 1.26$ and is constructed from iron as absorber material complemented with scintillator tiles following the segmentation of the barrel muon system. With the HO the effective thickness of the HCAL is increased to 10 interaction lengths. For the HF steel is used as absorber material and as an active medium quartz fibers are used, which are more persistent for the higher levels of radiation expected in the forward region. The hadronic energy resolution of the barrel HCAL and ECAL combination is parameterized as

$$\left(\frac{\sigma(E)}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + C^2, \tag{2.6}$$

where S corresponds to a stochastic term and C to a constant. With E measured in GeV, the values of S and C are measured to be $S = 0.847 \pm 0.016$ and $C = 0.074 \pm 0.008$ [40]. The energy resolution in the endcaps is similar to that in the barrel. The performance of the HCAL is often determined by calculating the resolution on the transverse energy of the reconstructed jets. The resolution on the transverse energy of the jets is an important item for this thesis and is discussed in Chapter 4.

2.2.4 The muon system

The muon system of the CMS detector is located in the iron return yoke of the experiment to profit from the strong magnetic return field for momentum and charge measurements. Only particles escaping detection and muons reach this part of the detector without significant energy loss. The full muon system makes use of three different types of gaseous detectors and consists of four layers of muon stations both in the barrel and endcaps. A layout of the muon system is shown in Figure 2.9. The barrel

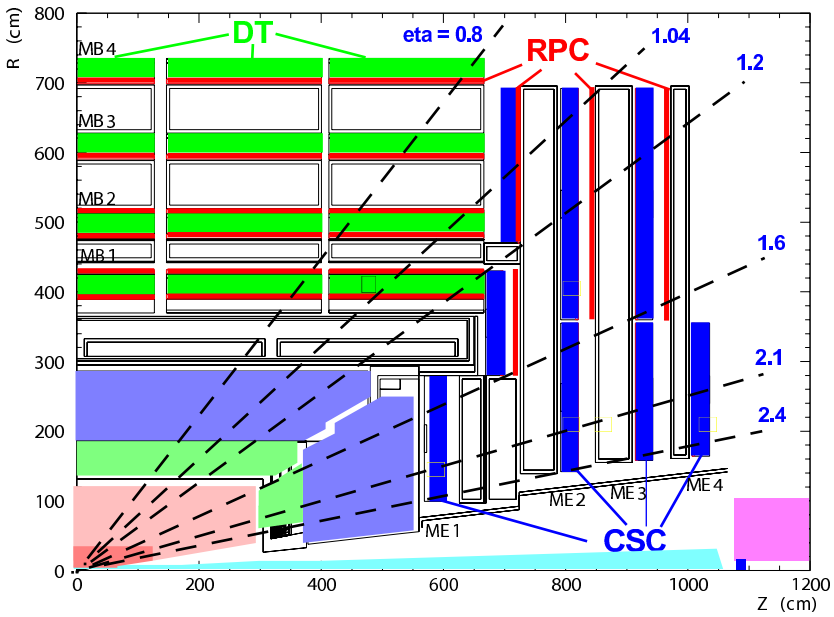


Figure 2.9: Longitudinal view of a quarter of the muon system layout indicating the location of the three different detector types in the muon system.

part of the muon system provides coverage for pseudo-rapidities up to $|\eta| = 1.2$, while the endcaps extend the coverage to $|\eta| = 2.4$. In the barrel four layers or stations of Drift Tubes (DT) and Resistive Plate Chambers (RPC) are placed. In the endcap the stations are formed by placing Cathode Strip Chambers (CSC) and RPC's together in concentric rings. RPC's were chosen for their fast response with a good time resolution of about 1 ns, but a coarser position resolution than the DT's or CSC's. RPC's are therefore used to determine the correct bunch crossing. Both in the barrel and endcap regions, the other types of muon detectors are assisted by the RPC's. In the barrel region, for the first two layers both before and after the DT's, RPC's are placed. In the endcap region the RPC's cover pseudo-rapidities up to $|\eta| = 1.6$. The DT's and CSC's are used to determine precisely the position of the muon traversing the detector. The DT's are used in the barrel region since the rate of particles is expected to be low. A final spatial resolution on a single hit of 100 μm is obtained with the DT's, while the CSC's provide a spatial resolution of the order of 200 μm. The angular resolution on the ϕ direction is typically 10 mrad with the CSC's and about 1 mrad when measuring the direction with the DT's.

2.2.5 Online event selection process

When the LHC bunch crossing rate is 40 MHz and when the LHC operates at its design luminosity this corresponds to about 10^9 proton-proton interactions per second in the CMS detector. However, only the data from about 100 interactions per second can be written to storage for analysis purposes. Therefore, an online event selection process or trigger is developed to select the interesting events and reject about 99.99975 % of the bunch crossings. Since the time between two bunch crossings is only 25 ns, the decision to keep or reject an interesting interaction must be taken as quickly as possible. However, reading out all the channels from the CMS detector takes already more time than the time between two bunch crossings. Therefore, the trigger system consists of two physical levels. The Level-1 (L1) trigger implemented on dedicated hardware and a High-Level Trigger (HLT) based on software for offline reconstruction. In most proton collisions soft hadrons will be produced with low transverse momenta (p_T), while events from processes of interest contain photons, leptons or jets with a high transverse momentum, transverse energy (E_T) or missing transverse energy due to particles escaping detection. The different trigger streams are therefore defined based on the reconstruction and identification of these physics objects. For the analysis performed in this thesis the trigger decision based on muons is of importance, hence the focus is on the trigger decisions based on the information of the muon system.

The Level-1 trigger

The total time in which a L1 trigger decision is made is designed to be 3.2 μs, corresponding to the time during which 128 bunch crossings happen. The complete detector information is kept in pipeline memories with a depth of 128 bunch crossings. The L1 decision is made based on the information provided by synchronous working processing elements. Each processing element is designed to take less than 25 ns to complete and works in a parallel structure. Because the largest part of the 3.2 μs is used for the transport of signals from the detector to the services cavern housing the trigger logic and the transport of the L1 trigger decision back to the detector, the time allocated to the decision of the L1 trigger is less than 1 μs. The L1 trigger reduces the rate of 40 MHz to 100 kHz, or 50 kHz when the LHC is operating at lower luminosities.

For the L1 trigger the information of the muon detectors and calorimeters is used. The decision to keep or reject an interaction is defined by the so-called trigger conditions, i.e. the presence of one or more objects above predefined p_T or E_T thresholds. In total up to 128 trigger conditions can be implemented in parallel.

The High-Level Trigger

The HLT is designed to reduce the data rate further to about 100 Hz, which is the maximal rate that can be accepted taking into account the available bandwidth and an event size of about 1 MB. The HLT software uses more detailed information and employs sophisticated reconstruction algorithms to make the final selection of events to be archived. Different HLT streams are developed based on energetic jets, missing transverse energy, isolated leptons etcetera. The HLT software is installed on a farm of several hundreds of processors each processing the information of one full event at

the time. The software used for the HLT is the same as the software used for the offline analysis. In order not to waste resources, the HLT decision should be made as soon as possible. Therefore, the event is reconstructed in different steps. In each step of one HLT stream more sophisticated and time-consuming algorithms or detector information are used provided the event is still selected after the previous step. The muon trigger streams in the HLT are implemented in different steps. The muons are first reconstructed using only the muon system after which the muon track candidates are propagated to the inner silicon tracker. If an isolated muon trigger condition is implemented, the HLT isolation algorithms discriminate on the basis of the energy deposits in the calorimeters and the total transverse momentum of the tracks in the inner tracker in a cone around the direction of the muon [1].

2.3 Distributed computing in CMS

The event rate of proton collisions recorded by the CMS experiment results in several Pbytes of data that need to be archived for offline processing and analysis purposes. The CMS collaboration consists of more than 3000 collaborators distributed over 40 countries that need access to the archived data. Therefore the CMS computing model is based on the principle of distributed computing and analysis. All the experiments at the LHC make use of the so-called Worldwide LHC Computing Grid (WLCG) [41] to manage data access and workload. WLCG groups together smaller Grid networks such as for example EGEE (Europe) and OSG (USA).

2.3.1 Hierarchical tiered structure

The distributed computing model is based on Grid techniques, i.e. computing resources that are geographically distributed and interconnected. A hierarchical structure of three tiers of computing facilities is put in place. The sites are interconnected through high-speed networks and each computing center is part of a certain tier level with specific functionalities.

- **The Tier-0 (T0) center at CERN**
  The T0 located at CERN is used for the data export from the CMS detector and performs the first pass reconstruction. A system, the CMS Analysis Facility (CAF) exists at the T0 and is used for very high priority analyses such as calibration and trigger activities.

- **The Tier-1 (T1) centers**
  The reconstructed data is distributed from the T0 to the T1s. Currently there are 7 T1 centers of which five are located in Europe, one in the USA and one in Asia. The T1s store a fraction of the raw data recorded by the CMS detector as well as simulated proton collisions produced at the T2s. The computing resources at the T1 centers are used for the (re)reconstruction of data.

[1] More information on the reconstruction of muons and the definition of isolation can respectively be found in Section 4.2 and Section 5.1.2.

- **The Tier-2 (T2) centers**
  About 50 T2 sites distributed over the world currently exist, providing the computing resources for the CMS collaboration to perform data analyses. The production of simulated proton collisions is also performed at the T2s.

Dedicated software for data and workload management [42] is well established as well as software for users to facilitate running analysis jobs on the Grid [43].

2.3.2 Data transfer links

The distributed computing model can only work successfully if data can be efficiently transported between the different tier centers around the world. In 2007 a dedicated taskforce was created within the CMS collaboration with as a main task the Debugging of Data Transfer (DDT) links between the centers [44–46]. Data transfer connections or links between the different centers are regularly tested by injecting so-called LoadTest transfers, followed by a coordinated effort between the site administrators and the DDT taskforce to debug possible transfer issues. A good quality label is assigned to a data transfer link if it reaches an average transfer rate of 20 MB/s over 24 hours for T1 to T1 and T1 to T2 data transfer links. For T2 data transfer links to a T1, this average rate is reduced to 5 MB/s. Only links with a good quality label are used in the CMS computing infrastructure. For successful operation, the links to the T0 from the T1s and vice versa need to have a good quality. Additionally, the data transfer links between the T1s are required to be operational. All the T1s should be able to transfer data to any T2, while a T2 center is considered commissioned and thus useful for distributed computing when it has at least one fully operational link to transfer data to its regional or associated T1 center. The total number of links to be commissioned is thus close to 500.

2.3.3 Deployment of CMS software

In order to use the tier centers, the CMS software, CMSSW, needs to be installed at these centers. CMSSW is used for the HLT, offline reconstruction, physics analysis and the production of simulated proton collisions. For most tier centers, the deployment and removal of software is managed centrally by two deployment teams, one for OSG based centers and one for EGEE-like tier centers. Grid jobs are used to install and remove software remotely. With a special Virtual Organization Membership Service (VOMS) role the job has a high priority in the queueing system of the tier center and writing privileges to the software area. Once a release is installed, a tag is set on the site to publish the availability of the release. These tags are used by the workload management systems to distribute analysis or production jobs that require a specific CMSSW version over the Grid. Since no interactive access to the tier centers is available, the installation and removal jobs must be robust against possible failures to avoid a broken software area on a tier center. When a CMSSW release is installed in the software area, it can be used by all the worker nodes of the computing farm (computing element) of the tier center.

When a new tier center is commissioned a first setup of the software area is performed. During this first setup, the underlying operating system is checked for all required packages. It is then possible to install the CMS software, which is packaged using the RedHat Package Manager (RPM), independent of the underlying operating system used at the tier center. Each CMS software release consists of about 80 packages, corresponding to a total of about 5 Gb of disk space needed for each release. About 10 to 20 CMSSW releases co-exist in the software area of each tier center.

New software releases are regularly developed to include new reconstruction algorithms or for instance to update the detector alignment. At least one new CMSSW version is released every week. However, it can happen that only one version is released in a period of two weeks, or up to five versions are releases in a few days depending on the input the software development team is receiving. For instance when a production round of simulated proton collisions starts or during the initial data taking period, every few days new software versions become available. The installation is requested by the CMS software management and followed by the manually creation of the deployment jobs which are then sent to the different tier centers. An example of the time needed to install a certain software release is shown in Figure 2.10. The time axis starts

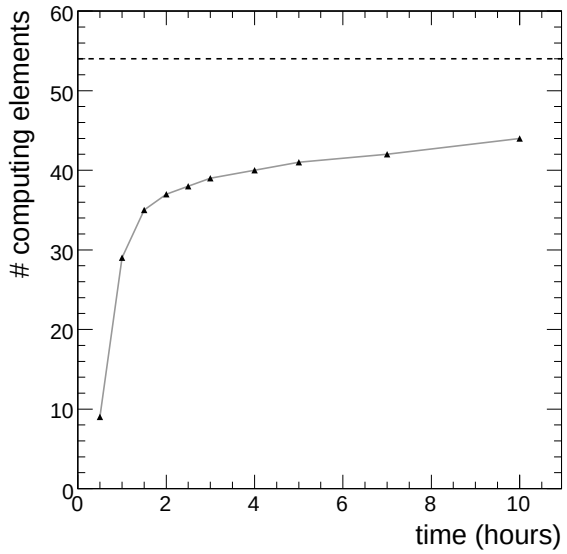


Figure 2.10: Time needed to deploy CMSSW_3_1_4 on the EGEE sites, in total 54 computing elements (dashed line).

at the moment the software management warned the deployment team, while the first jobs were submitted 20 minutes later. A tier center has usually one computing farm or computing element, although some tier centers have two computing elements with independent software areas. It takes about an hour to submit the jobs to the whole list of EGEE tier centers, in total 54 computing elements. After 3.5 hours, the software release is available at about 80 % of the computing elements. At 5 tier centers problems occurred during the installation, these problems involve for instance network problems or the wrong mapping of the site for the owner of the installation job such that writing permissions are not obtained. Tier centers that were in maintenance are unavailable

and the software could not be installed there, therefore the maximum indicated with the dashed line can not be reached.

Recently some efforts were done to automate the installation procedure. The installation jobs are automatically submitted when a new release is announced. Human intervention is however still needed to follow up job failures at 10-20 % of the tier centers.

Models are to be used, not believed.

H. Theil

Chapter 3

Simulating collision events

Particle accelerators like the Large Hadron Collider produce collisions that are registered by a detector, for instance the CMS detector, and stored for physics analyses. Before the machine becomes operational, particle collisions are simulated by event generator programs and subsequently fed into a detector simulation. Event generators aim to both describe and predict the physics phenomena in collisions from the hard interaction until the observed hadrons in the final state. Event generation and detector simulation are vital for the design of the detector and trigger, and to develop diverse analysis strategies, while the real collisions are needed to unravel the fundamental interactions. Eventually, the discovery of new phenomena will extend our knowledge of the first principles and the phenomenological modelling of the physics in collisions, leading to even better event generators. A strong interplay is thus present between the real and simulated proton collisions at the Large Hadron Collider.

In this chapter, we will focus on how the event generation and detector simulation is performed by the CMS collaboration, with an emphasis on the aspects that are needed for the physics studies in the next chapters. It is worth to mention that the event reconstruction framework and the physics studies are developed in such a way that they are identical for both the real and the simulated proton collisions. A factorization of the different steps in the event generation process is possible, as described in Section 3.1. Because this thesis is based on the observation of top quark pairs at the LHC, the theoretical predictions for the top quark pair production cross section are summarized in Section 3.2. In Section 3.3, the kinematic properties of the top quarks produced with different event generators are compared. Section 3.4 elaborates on how the stochastic interaction between the generated particles and the detector is simulated and discusses the software framework in the CMS collaboration for the production of simulated event samples. Finally, the physics analysis model is discussed in Section 3.5.

3.1 Generating $\mathrm{pp} \to \mathrm{X}$ events

The generation of a $\mathrm{pp} \to \mathrm{X}$ collision event starting from the incoming protons at the LHC until the decay of the hadrons that form the experimentally observable final state X, can be factorized into different steps as illustrated in Figure 3.1. The event generation chain is a stochastic process and can be summarized as follows in chronological

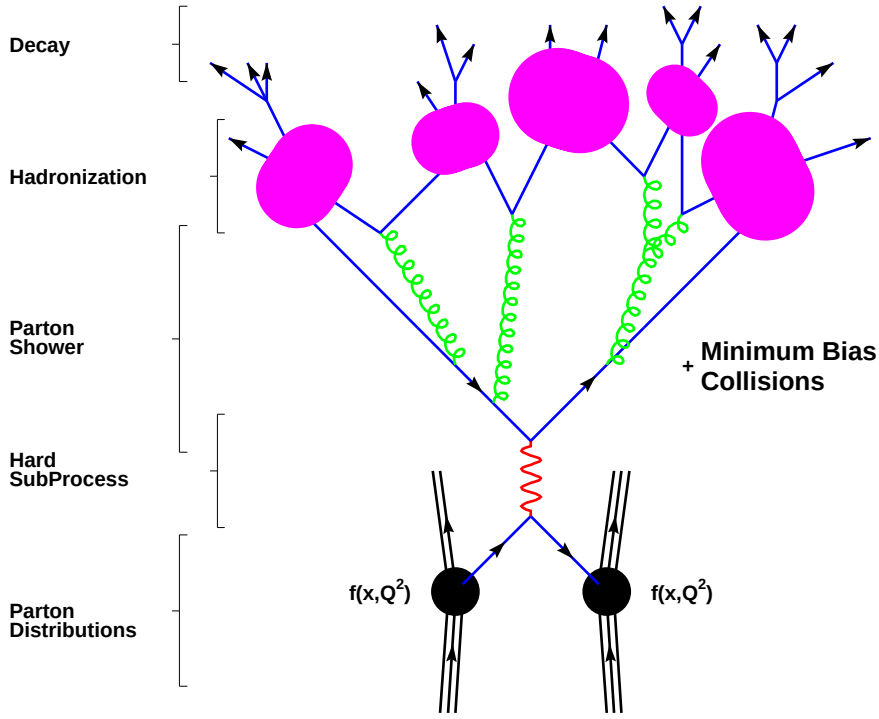


Figure 3.1: Illustration of the general structure of the event generation chain for hadron collisions [47].

order:

- Two protons, composed of partons (i.e. quarks and gluons), approach each other on a collision course. A parton in each proton is resolved with a certain probability density $f(x, Q^2)$ which depends on the momentum fraction x carried by the parton in the proton and the momentum transfer Q^2 (cfr. Section 3.1.1).

- The two resolved partons interact and produce the hard process of interest or the hard interaction (cfr. Section 3.1.2). In case a resonance with a short life time is produced, such as a top quark or W boson, the decay is viewed as a part of the hard process.

- Both the partons before the hard interaction as well as the partons produced in the hard interaction can branch into other partons, which can branch themselves, thus resulting in a shower of secondary partons. In the case this radiation happens from the initial partons before the hard interaction it is called Initial State Radiation (ISR). When the radiation is from the partons produced in the hard interaction it is called Final State Radiation (FSR). The perturbative method to include radiation in the event generation is controlled by the coupling constants α (with $\alpha < 1$) and is discussed in Section 3.1.3.

- The running of the strong coupling constant α_S induces the phenomenon of colour confinement and therefore prevents the generated partons to be observed in iso-

lation. By using non-perturbative phenomenological models the partons are fragmented and the colour charges are grouped into hadrons. Section 3.1.5 elaborates on the phenomenological modelling of the hadronization or fragmentation process. After the hadronization, further decays are simulated.

- After the two partons are resolved in the proton, the proton remnants are not colour neutral and will thus radiate and hadronize themselves. The simulation of the so-called underlying event is described in Section 3.1.6. In this section also the modelling of interactions between other partons in the colliding protons (i.e. multiple interactions) is discussed.

For the simulation of the different steps in the event generation chain described above many software programs have been developed. Monte-Carlo techniques are used to reproduce the stochastic nature of each step in the event generation chain. Diverse tools for the simulation of specific steps, e.g. for the generation of the hard process of interest, and general-purpose generators exist. An example of generators that are capable of simulating the complete event generation chain are `PYTHIA` [48] and `HERWIG` [49]. Some commonly known generators specialised in the calculation of the hard interaction are for instance `ALPGEN` [50] and `MadGraph/MadEvent` [51, 52]. The last type of generators often needs to be interfaced with the general-purpose ones to complement the simulation of the hard interaction. This was made possible thanks to the Les Houches Accord in 2001 [53] in which a "baseline" interface was defined. Even with a uniform interface a matching is needed between the partons produced in the hard interaction and the partons produced by radiation to avoid double counting. This topic is discussed in Section 3.1.4.

3.1.1 Factorization theorem and parton distribution functions

According to the QCD factorization theorem [54], the hadronic cross section for the production of a final state X from the collision of two protons, $\sigma(\mathrm{pp} \to \mathrm{X})$, can be written as a function of the differential subprocess cross section of the partons a and b to produce the final state X weighted with the parton distribution functions (pdf's):

$$\sigma_{\mathrm{pp}\to\mathrm{X}} = \sum_{a,b} \iint dx_a dx_b f_a(x_a, Q^2) f_b(x_b, Q^2) \hat{\sigma}_{ab\to X} \tag{3.1}$$

The parton distribution functions $f_i(x_i, Q^2)$ represent the probability that a parton i in the proton is resolved with a momentum fraction x_i at a scale Q^2. The sum runs over all contributions from the different initial partons (flavour, spin, colour, ...), while integrating over the momentum fractions of the partons x_a and x_b. The partonic cross section $\hat{\sigma}_{ab\to X}$ can be expanded in the strong coupling constant α_S so that a perturbative calculation of the total partonic cross section is possible. The parton distribution functions are obtained by means of global fits to experimental data from deep inelastic scattering, Drell-Yan and jet processes. Different collaborations, such as CTEQ [55] and MRST [56], provide updates when new data and/or theoretical predictions become available. The parton distribution functions used in the production of the Monte-Carlo simulated event samples for this thesis is the CTEQ6L1 set. Some

of the parton distribution functions are shown in Figure 3.2 [57] for a scale Q =350 GeV, corresponding to the invariant mass of $t\bar{t}$ system. The gluon distribution function is

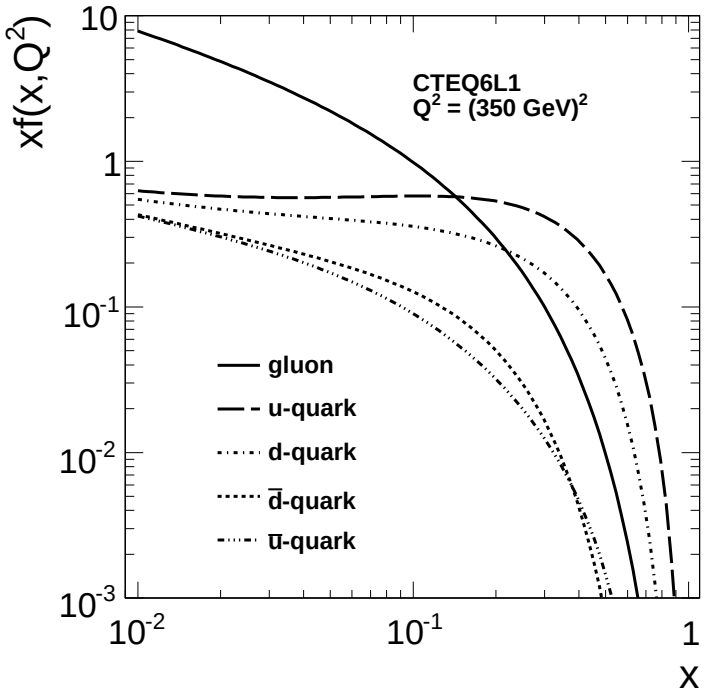


Figure 3.2: Different parton distribution functions from the CTEQ6L1 release are shown for a Q^2-value of $(350\,\text{GeV})^2$ [57].

dominant at low x-values, while the parton distribution functions of the valence quarks in the proton are dominant at high x-values. Due to the high center-of-mass value for proton collisions at the LHC, relatively low x-values will be probed and thus top quark pairs are dominantly produced through gluon fusion.

To calculate the uncertainty on the parton distributions two different techniques exist, the Lagrange Multiplier and the Hessian techniques [58, 59]. The Hessian method is often used as it is able to provide an uncertainty on the parton distribution functions for the cross section of any process. In the Hessian technique, a matrix with dimensions equal to the number of free parameters in the pdf fit needs to be diagonalized. In the case of the CTEQ pdf's, this results in 20 orthonormal eigenvector directions. To construct the 40 "error" pdf's, each eigenvector is evaluated in the "+" and the "−" direction where the value of the fit parameter is deviated from its estimated value by respectively +1 and −1 standard deviation. The final systematic uncertainty on a measured variable X is then obtained by summing over the 20 couples of "error" pdf's:

$$\Delta X = \frac{1}{2}\sqrt{\sum_{\text{eigenvector}\,i} (X_{+,i} - X_{-,i})^2} \tag{3.2}$$

where $X_{+,i}$ is the value of X using the "error" pdf of eigenvector i in the "+" direction.

3.1.2 Simulation of the hard interaction

In the previous section the factorization theorem was introduced, leading to the insight that the key element in obtaining the hadronic cross section is the calculation of the differential partonic cross section or the cross section of the hard interaction $\hat{\sigma}_{ab\to X}$. Consider for instance the hard interaction $u\bar{u} \to Z \to d\bar{d}$. The differential cross section $\hat{\sigma}_{u\bar{u}\to Z\to d\bar{d}}$ of this process can then be written as

$$d\hat{\sigma}_{u\bar{u}\to Z\to d\bar{d}} = \frac{1}{2s}|\mathcal{M}|^2 \frac{d\cos\theta d\phi}{8(2\pi)^2} \tag{3.3}$$

with θ and ϕ the decay angles of the intermediate Z boson, s the center-of-mass energy squared and $\mathcal{M}$ the matrix element of the process. To generate an event, on one hand the matrix element needs to be calculated and on the other hand, the phase space, defined by θ and ϕ needs to be sampled. The calculation of the matrix element, up to a certain order of the strong coupling constant α_S, forms the core of the simulation of the hard interaction. Usually virtual loops are not included in the calculation and the computation is performed for a fixed number of partons in the final state. Most event generators provide matrix elements in Leading Order (LO). A so-called K-factor is often used to rescale the LO cross section, which contains the relative strength of the Next-to-Leading Order (NLO) correction to the lowest order cross section. It is obtained by dividing the NLO and LO values of the cross section obtained from theoretical calculations. However, one should be careful when using the K-factor as it may depend on the region of phase space (θ, ϕ) that is studied and on the parton distribution functions that have been used for both the LO and NLO calculations. Also a dependence on the factorization scale exists [60].

In the general-purpose event generators, like for instance `PYTHIA`, the analytical formula to calculate the matrix element for a large number of processes is implemented. Therefore, often only the lowest order matrix element is available and additional hard partons in the final state are generated by parton shower algorithms (cfr. Section 3.1.3). However, also generators explicitly based on Matrix Elements for specific processes have been developed, which are able to generate additional partons in the final state within the matrix element up to a certain order. Examples of such generators, which are used in the CMS collaboration in particular for the study of top quark physics, are `ALPGEN` and `MadGraph/MadEvent`.

The `ALPGEN` matrix element generator [50]

The `ALPGEN` matrix element generator has been developed to study multi-parton Standard Model processes in hadronic collisions. There are two modes to generate events with `ALPGEN`. In the first mode, weighted events are produced by performing the calculation of the matrix element relative to the selected hard processes. This mode allows for parton-level studies for which high statistics are needed. In the second mode, unweighted events are produced and can subsequently be processed through parton showering and hadronization programs such as `PYTHIA`. Colour, spin, flavour and kinematical information are stored so the parton shower program can access all the needed information.

The `MadGraph/MadEvent` matrix element generator [51, 52]

`MadGraph` is a matrix element generator calculating the Feynman diagrams for all the relevant subprocesses of a process specified by the user. Unweighted events and cross sections are obtained by feeding the output of `MadGraph` to `MadEvent`. Colour and spin information are stored and the output of `MadEvent` can be given to a parton showering program. The limiting factor is related to the maximum number of QCD particles in the final state of the subprocess and currently this limit is ten thousand diagrams. An example close to the practical limit is for instance the simulation of $W + 5$ jets events. Apart from the Standard Model processes, many different models beyond the Standard Model have been implemented. Furthermore, new models can be easily implemented due to the high flexibility. Multi-CPU clusters can be used to obtain the relevant Feynman diagrams as the architecture of the calculation code is intrinsically parallel.

3.1.3 Parton showering

Parton showering programs aim to describe the physics of radiating partons in the initial and final state and to relate the few partons produced in a hard interaction to the partons after an evolution to a lower scale Λ_{QCD} where the perturbative parton shower approach breaks down ($\alpha_S \approx 1$). Three different types of branchings exist. For the first and second one, a gluon is emitted either from a quark ($q \to qg$) or a gluon ($g \to gg$), while the third type is the splitting of a gluon into a quark and anti-quark ($g \to q\bar{q}$). If the partons produced in the final state are well-separated in phase space and as long as loops (additional emitted partons that are not present in the final state) are not taken into account, matrix element generators provide an accurate description. Divergences in the cross section calculation do however occur for $2 \to n$ processes when the partons are collinear (collinear singularities) or when the energy of one of the partons vanishes (soft singularities). Therefore, the extension of the calculations to the soft and collinear regions is usually handled by the parton showering programs rather than by a matrix element formalism. At the energy scales of the hard interaction, a perturbative evolution of the parton fragmentation functions is possible using the DGLAP (Dokshitzer-Gribov-Lipatov-Altarelli-Parisi) formalism [61–63].

Consider a mother parton a that branches into two daughter partons b and c. Daughter b takes a fraction z of the momentum of the parton a, while daughter c takes the remaining $1 - z$. Let us introduce the scale-dependent parameter t,

$$t \;=\; ln(Q^2/\Lambda^2_{QCD}) \quad\Longrightarrow\quad dt \;=\; \frac{dQ^2}{Q^2}, \tag{3.4}$$

with Λ_{QCD} the scale at which α_S has unit strength (perturbative QCD breaks down). The differential probability for a parton a to branch is then given by

$$d\mathcal{P}_{a\to bc} = \frac{\alpha_S}{2\pi}\mathcal{P}_{a\to bc}(z)dtdz = \frac{\alpha_S}{2\pi}\frac{dQ^2}{Q^2}\mathcal{P}_{a\to bc}(z)dz \tag{3.5}$$

where $\mathcal{P}_{a\to bc}$ are the splitting functions, defined as

$$\begin{aligned}\mathcal{P}_{q\to qg}(z) &= C_F\frac{1+z^2}{1-z}\ ,\\ \mathcal{P}_{g\to gg}(z) &= N_C\frac{(1-z(1-z))^2}{z(1-z)}\ ,\\ \mathcal{P}_{g\to q\bar{q}}(z) &= \frac{n_f}{2}(z^2+(1-z)^2)\end{aligned} \tag{3.6}$$

where $C_F = 4/3$ is the colour factor, $N_C = 3$ is the number of colours and n_f is the number of quark flavours. Successive emissions are obtained by combining these equations in several steps to allow the daughters to branch further and thus to develop a shower. For initial state radiation the evolution is performed backwards. This means that when a parton b enters the hard interaction, the program tries to reconstruct what happened before using a conditional probability, defined as the probability that a branching $a \to bc$ occurred at a smaller Q^2 scale, under the condition that the parton b is present at the Q^2 scale of the hard interaction.

After each emission the virtuality Q^2 of the mother (i.e. branching) parton decreases and therefore α_S increases. The showering is stopped at a lower cutoff scale, often taken around $Q^2 = 1\,\mathrm{GeV}^2$, to avoid ending up in the non-perturbative region where α_S diverges or dQ^2/Q^2 in equation 3.5 becomes singular. Despite the cutoff scale, branching probabilities above unity are still possible, therefore the branching probabilities are weighted with the so-called Sudakov form factor [64]. If we consider time-evolution, we can write the probability that no branching occurred up to a time T as follows

$$\mathcal{P}_{\text{no branching}}(0 < t \leqslant T) = \lim_{n\to\infty}\prod_{i=0}^{n-1}\mathcal{P}_{\text{no branching}}(T_i < t \leqslant T_{i+1}) \tag{3.7}$$

with $T_i = (i/n)T$ and $0 \leqslant i \leqslant n$. Taking into account the conservation of total probability, the previous equation can also be written as

$$\begin{aligned}\mathcal{P}_{\text{no branching}}(0 < t \leqslant T) &= \lim_{n\to\infty}\prod_{i=0}^{n-1}(1-\mathcal{P}_{\text{branching}}(T_i < t \leqslant T_{i+1}))\\ &= \exp\left(-\lim_{n\to\infty}\sum_{i=0}^{n-1}\mathcal{P}_{\text{branching}}(T_i < t \leqslant T_{i+1})\right)\\ &= \exp\left(-\int_0^T\frac{d\mathcal{P}_{\text{branching}}(t)}{dt}dt\right)\end{aligned} \tag{3.8}$$

The time-evolution can be translated into a Q^2 evolution. As time evolves, the Q^2 scale becomes smaller and smaller. After weighting the DGLAP equation with the Sudakov form factor, equation 3.5 becomes

$$d\mathcal{P}_{a\to bc} = \frac{\alpha_S}{2\pi}\frac{dQ^2}{Q^2}\mathcal{P}_{a\to bc}(z)dz\,\exp\left(-\sum_{b,c}\int_{Q^2}^{Q^2_{max}}\frac{dQ'^2}{Q'^2}\int\frac{\alpha_S}{2\pi}\mathcal{P}_{a\to bc}(z')dz'\right) \tag{3.9}$$

where the sum in the Sudakov form factor runs over the possible branchings. For initial state radiation, the splitting function in the exponent is weighted by the parton distribution function at the relevant scale.

In the simulation of the samples used for the physics studies presented in this thesis, the parton showering available in PYTHIA is applied. To understand the effect of the parameter tuning in PYTHIA on the physics estimators, simulated samples have been produced with different parameter values. The considered parameters are listed below with their default values, while the variation of the values is specified together with the produced event samples in Section 3.4. The values and their variation are based on a study that was performed by the CMS collaboration for the estimation of theoretical uncertainties [65].

- Λ_{QCD} or parameters PARP(61), PARP(72) and PARJ(81) in PYTHIA
  This is the general QCD scale parameter used in the evaluation of the running strong coupling α_S and used in the DGLAP equations through the parameter t defined in equation 3.4. The amount of initial state radiation is tuned with the parameter value PARP(61), while the value of PARP(72) corresponds to the amount of final state radiation, except in the decay of resonances, when the value of PARJ(81) is used instead. The default value used in the physics analysis is 0.25 GeV for the three parameters PARP(61), PARP(72) and PARJ(81).

- Scaling of α_S or the parameter PARP(64) in PYTHIA
  In the space-like shower evolution (ISR), the squared transverse momentum of the evolution scale $k_\perp^2 = (1-z)Q^2$ is multiplied by PARP(64) for use as a scale in α_S and the parton distribution functions. The default value of this parameter is set to 1.0.

- $Q^2_{max,ISR}$ or the parameter PARP(67) in PYTHIA
  The Q^2 scale of the hard interaction is multiplied by the value of this parameter to define the maximum parton virtuality allowed for the showering of the initial state radiation. The default value of the parameter PARP(67) is set to 4.0.

3.1.4 Matching the parton shower with the matrix element

In the case a matrix element generator is interfaced with a parton showering program, such as PYTHIA, one has to be careful to avoid double counting of radiation. The problem of double counting appears from the fact that a $(N+1)$-jet event can be obtained in two ways: either from a $(N+1)$-parton matrix element calculation or from an N-parton matrix element calculation where an additional hard parton was radiated at a large angle during the parton showering. Several schemes exist to perform a matching between the parton showering and the matrix element calculation, such as the CKKW scheme [66, 67] and the MLM scheme [68, 69]. As both the ALPGEN and MadGraph/MadEvent matrix element generators are interfaced with the parton showering in PYTHIA using the MLM scheme, we focus the discussion on this approach.

For the MLM matching scheme, events with different parton multiplicities are generated by MadGraph/MadEvent or ALPGEN with certain requirements on the kinematics of the partons in the final state. Partons should have a transverse momentum p_T^{parton} above

a threshold p_T^{min} and should be within the pseudo-rapidity acceptance $|\eta^{parton}| < \eta^{max}$. In addition, each parton i should be separated from another parton j in the $(\eta \times \phi)$ plane by requiring $\Delta R_{i,j} = \sqrt{\Delta\phi_{ij}^2 - \Delta\eta_{ij}^2} > \Delta R^{min}$. The events are then showered with PYTHIA and a jet finding algorithm (cfr. Chapter 4) is applied with a jet radius parameter R_{jet}^{min}. Additionally, the jets are required to have a minimum transverse energy E_T^{min}. The parameters R_{jet}^{min} and E_T^{min} are related, but not necessarily identical to the parton level parameters, respectively ΔR^{min} and p_T^{min}. The procedure continues by associating every generated parton to one and only one of the jets, by taking the hardest parton (highest p_T^{parton}) and by selecting the jet closest in $(\eta \times \phi)$ space. If the distance between the jet axis and the parton is smaller than R_{jet}^{min}, the parton and jet match. The matched jet and parton are then removed from the list and the next parton is matched. The event is fully matched if each parton has a matched jet. In the case no jet is found that matches to the parton, the event is rejected. Events with collinear partons will thus be rejected as they are not able to generate separate jets. Also events that have a very soft parton are removed as such a parton will not be able to generate a jet. A fully inclusive sample is obtained by applying this procedure to the parton level with increasing multiplicities. In the highest partonic multiplicity, events with extra jets are not rejected. The distribution of the physical observables obtained from the fully inclusive sample should be independent of the cuts applied on p_T^{min} and ΔR^{min}.

Also for the CKKW matching scheme both a separation criterion between any two partons in the final state and a transverse momentum criterion on these partons are defined. The matrix elements are then reweighted using Sudakov form factors and coupling weights. Parton showers with a transverse momentum exceeding the transverse momentum criterion are vetoed.

A more ambitious approach tries to perform the matching accurately up to next-to-leading order. The MC@NLO program [70] is the most commonly known example of such an approach. Therefore the NLO matrix element corrections to an N-parton final state need to be calculated, including the $(N+1)$-parton real corrections and the N-parton virtual corrections. Analytically the calculation is performed how a first branching in a shower of the N-parton matrix element would populate the final state of the $(N+1)$-parton phase space. The calculated shower expression is then subtracted from the $(N+1)$-parton matrix element showering to obtain the "true" $(N+1)$ events. The difference is considered as being part of the N-parton final state. Now that the N-parton and $(N+1)$-parton event classes have been defined, full showering is added to both the N-parton and $(N+1)$-parton final states.

3.1.5 Fragmentation of the partons

At short distances, the QCD perturbation theory is valid and can be formulated in terms of quarks and gluons. However, at long distances the strong coupling constant α_S diverges and perturbation theory breaks down. After the parton showering, the partons are transformed in colourless hadrons by a process called hadronization or fragmentation. Because fragmentation is not yet understood from first principles, different phenomenological models have been developed to simulate the process. Cur-

rently, mainly two different approaches are used, with on one hand string fragmentation and on the other hand cluster fragmentation. In PYTHIA the hadronization is simulated using the Lund string model [71]. String modelling is based on the physical picture of a colour flux tube being stretched between two partons q and $\bar{q}$ moving apart. The energy stored in the flux tube is assumed to increase linearly with the separation between the two partons with different charge (the short-distance Coulomb term is neglected). The string constant κ, which corresponds to the amount of energy per unit length, is estimated to be $\kappa \approx 1\,\mathrm{GeV/fm}$. Due to the linearly increasing potential energy stored in the flux tube, the string may break while producing a new $q'\bar{q}'$ pair. In this way two new colour singlets $q\bar{q}'$ and $q'\bar{q}$ are formed. Further breakings may occur provided that the invariant mass of the new colour singlets is high enough. The process is continued until only on-mass-shell hadrons remain. This procedure is described by the Schwinger mechanism based on quantum tunneling [72], resulting in a probability proportional to

$$\exp\left(\frac{-\pi m^2}{\kappa}\right)\exp\left(\frac{-\pi p_T^2}{\kappa}\right) \tag{3.10}$$

for pair production, with p_T the transverse momentum and m the mass of the quarks in the created pairs. Since the probability decreases with the mass of the produced quarks, heavy quark production is suppressed, $u : d : \mathrm{s} : c \approx 1 : 1 : 0.3 : 10^{-11}$. Therefore, charm and heavier quark production is neglected in the fragmentation process. As the string fragmentation process should be independent of the direction of the fragmentation on the string (the quark end or the anti-quark end), the Lund fragmentation function should be symmetric. The function is defined as

$$f(z) \propto \frac{(1-z)^a}{z}\exp(\frac{-b(m_h^2 + p_{T,h}^2)}{z}) \tag{3.11}$$

with m_h and $p_{T,h}$ respectively the mass and transverse momentum of the created hadron. The parameters a and b are tuned to experimental observations. The fragmentation function $f(z)$ is used to obtain stochastically the longitudinal momentum of the created hadrons, where z is the fractional momentum of the hadron $q\bar{q}'$ that is split off, while $(1-z)$ is the fractional momentum for the remainder of the string $q'\bar{q}$. While these functions are fine for u, d, s quarks, experimental data indicate the need for harder fragmentation functions for heavy flavours. The fragmentation function used for that purpose is mostly the Peterson function [73]

$$f(z) \propto \frac{1}{z\left(1 - \frac{1}{z} - \frac{\epsilon_q}{1-z}\right)^2}\ , \tag{3.12}$$

where ϵ_q is a free parameter, expected to scale like $\epsilon_q \propto 1/m_q^2$ and depends on the quark flavour.

The hadronization process in PYTHIA is thus configured using five parameters, of which the values are given in Table 3.1. The default values [48] that have been used in the simulation of the fragmentation are close to the values given in [65] and consistent with the tuning from the LEP experiments [74–76].

After the hadronization process, a significant fraction of the produced particles is unstable and need to be decayed by the simulation program. Therefore, all particles

parameter	PYTHIA	value
$p_{x,h}$ and $p_{y,h}$ widths	PARJ(21)	0.36 GeV/c
Lund a	PARJ(41)	0.3
Lund b	PARJ(42)	0.58 $\mathrm{GeV}^{-2}\mathrm{c}^2$
Peterson ϵ_c	PARJ(54)	-0.05
Peterson ϵ_b	PARJ(55)	-0.005

Table 3.1: The default values of the relevant PYTHIA hadronization parameters used in this thesis.

need to be included with their proper mass distributions and decay properties. In PYTHIA all particles with a proper lifetime τ such that $c\tau < 10\,\mathrm{mm}$ are allowed to decay. Dedicated programs for the decay of particles exist, such as TAUOLA package, which is used for the generation of tau lepton decays including spin polarization. The package is interfaced to the general-purpose event generator PYTHIA.

3.1.6 Beam remnants and multiple interactions

The so-called underlying event is made up of two components. One contribution comes from the beam remnants, i.e. the remainder of the protons from which a parton has been resolved for the hard interaction and that are consequently not colour neutral anymore. The beam remnants are therefore colour-connected to the hard interaction. A primordial transverse momentum $k_\perp$ is usually assigned to the resolved parton and the remnant of the proton is taking to be its recoil. This area is not at all understood from first principles, in particular how the energy is shared and how the colour connection between the beam remnants and the resolved partons is happening. The primordial $k_\perp$ distribution is taken as a Gaussian during the simulation with PYTHIA and the width of this distribution $\sigma(k_\perp)$ can be tuned. Additionally, an upper threshold $k_\perp^{max}$ for the primordial $k_\perp$ distribution can be applied. The relevant values of the primordial $k_\perp$ distribution from the Tune D6T [48] obtained with Tevatron data are quoted in Table 3.2.

Another contribution to the underlying event comes from multiple interactions. It is possible that several partons from each beam are involved in separate $2 \to 2$ scatterings. However, the probability to have several scatterings with high transverse momentum in the same bunch crossing is small, since the cross section drops rapidly with increasing p_T. A lower cut-off scale needs to be introduced since the dominant $2 \to 2$ QCD cross sections are divergent for $p_T \to 0$. This divergence is appearing because the assumption has been made that the quarks and gluons are free at very small distances, a phenomenon referred to as asymptotic freedom. However, the confinement of colour into hadrons of a finite size has not been taken into account. Since the proton is colour neutral, while the individual partons in the proton are carrying a colour charge, colour screening is present. To describe Tevatron data, an empirical lower cut p_T^{min} is introduced of about 2 GeV/c. Below this threshold the perturbative cross section is

taken to be zero or at least strongly dampened. Since there is a strong dependence of the p_T^{min} in the collision energy, the tuning has been performed with data from different experiments. In the simulated events samples used in the analysis this dependency is originally included from a fit to the UA5 and Tevatron data [65, 77, 78].

Multiple interactions are stochastically simulated in PYTHIA assuming a varying impact parameter and a hadronic matter overlap consistent with a double Gaussian matter distribution given by

$$\rho(r) \propto \frac{1-\beta}{a_1^3} \exp\left(-\frac{r^2}{a_1^2}\right) + \frac{\beta}{a_2^3} \exp\left(-\frac{r^2}{a_2^2}\right) \tag{3.13}$$

This corresponds to a distribution with a small core region, of radius a_2 and containing a fraction β of the total hadronic matter, embedded in a larger hadron of radius a_1. The values of p_T^{min}, the minimum transverse momentum for multiple interactions to occur, and the regularization scale p_T^0, for the continuous dampening of the cross section, which are used in the simulation have been determined at a reference energy scale $\sqrt{s}^{\text{ref}}$ taken to be equal to the center of mass energy studied at the Tevatron experiment. The energy rescaling is assumed to be proportional to $(\sqrt{s})^{x_{scale}}$, where x_{scale} was tuned to Tevatron data. Table 3.2 summarizes the values of the various parameters, as obtained from Tune D6T, which have been used for the simulation of the underlying event. Recently, new tunes became available [79]. However, it must be noted that

parameter	PYTHIA	value
$\sigma(k_\perp)$	PARP(91)	2.1 GeV/c
$k_\perp^{max}$	PARP(93)	15 GeV/c
β	PARP(83)	0.5
a_2/a_1	PARP(84)	0.4
p_T^{min}	PARP(81)	1.9 GeV/c
p_T^0	PARP(82)	1.8387 GeV/c
$\sqrt{s}^{\text{ref}}$	PARP(89)	1.96 TeV
x_{scale}	PARP(90)	0.16

Table 3.2: The values of the relevant PYTHIA6.4 Tune D6T underlying event parameters [48].

the extrapolation of the energy dependence of the multiple parton interactions from $\sqrt{s}$ =1.96 to $\sqrt{s}$ =10 TeV has large uncertainties, which can only be constrained in the future by the data produced by the LHC.

In the phenomenological modelling of multiple interactions a lot of progress has been made in the past decades [80]. However, the colour connection between all the outgoing partons coming from different interactions is one of the areas that is poorly understood. An observable that is sensitive to the colour connection is the dependency

of the average transverse momentum of charged particles on their multiplicity. Obtaining this quantity in an experiment is one of the first studies that is performed. During the recent data taking periods at the LHC at center-of-mass energies of 0.9 and 2.36 TeV the observable has been measured and is found to be consistent with current predictions [81]. This provides some confidence in the hypotheses made in the Monte Carlo based simulation used for this thesis.

3.2 Top quark pair production cross section

During the last years, progress has been made in the theoretical calculation of the top quark pair production cross section at the center-of-mass energies of the Tevatron and the LHC. Currently, approximations of the next-to-next-to-leading order (NNLO) predictions are available [82]. Usually NLO cross sections are first calculated and the NNLO soft-gluon corrections in the $q\bar{q}$ and gg channels are added. The NLO cross section for the LHC at 14 TeV calculated with the NNLO MRST parton distribution functions is

$$\sigma^{NLO}_{pp\to t\bar{t}}(14\,\mathrm{TeV}, m_t = 172\,\mathrm{GeV/c^2}, \mathrm{MRST}) = 900^{+110+11}_{-107-12}\,\mathrm{pb} = 900^{+111}_{-108}\,\mathrm{pb}, \tag{3.14}$$

while with the NNLO CTEQ parton distributions it is

$$\sigma^{NLO}_{pp\to t\bar{t}}(14\,\mathrm{TeV}, m_t = 172\,\mathrm{GeV/c^2}, \mathrm{CTEQ}) = 854^{+102+27}_{-100-29}\,\mathrm{pb} = 854^{+106}_{-104}\,\mathrm{pb}. \tag{3.15}$$

The first uncertainty is due to the dependence of the factorization scale, while the second uncertainty is related to the parton distribution functions itself. The NNLO approximation becomes for the MRST parton distribution functions

$$\sigma^{NNLOapprox}_{pp\to t\bar{t}}(14\,\mathrm{TeV}, m_t = 172\,\mathrm{GeV/c^2}, \mathrm{MRST}) = 968^{+80}_{-52}\,\mathrm{pb}, \tag{3.16}$$

while with the CTEQ parton distributions it is

$$\sigma^{NNLOapprox}_{pp\to t\bar{t}}(14\,\mathrm{TeV}, m_t = 172\,\mathrm{GeV/c^2}, \mathrm{CTEQ}) = 919^{+76}_{-55}\,\mathrm{pb} \tag{3.17}$$

where the uncertainty is the total systematic uncertainty. The difference in the total uncertainty on the cross section between the NLO calculation and the NNLO approximation is almost entirely related to the dependence on the factorization scale, which is much reduced when taking into account the NNLO corrections. The uncertainty related to the parton distribution functions is smaller at the LHC compared to the Tevatron. This can be explained by the fact the LHC will operate at lower x values, and thus in a range where the parton distributions functions are better known compared to the range of x values for Tevatron [83]. Also, the dependency of the cross section on the top quark mass is smaller at the LHC than at the Tevatron, because the production of top quarks is further from its threshold. For the same reason, the dependency on the factorization scale μ, which corresponds to the resolution with which the (anti-)proton is probed, is smaller at the LHC than at the Tevatron. The factorization scale is set equal to the top quark mass m_t and the uncertainty is obtained by varying the factorization scale over $m_t/2 < \mu < 2m_t$. In references [82, 83] the top quark pair

production cross section was calculated at a center-of-mass energy of 10 TeV. With the MRST parton distribution functions the NNLO approximate cross section is

$$\sigma_{pp\to t\bar{t}}^{NNLOapprox}(10\,\mathrm{TeV}, m_t = 172\,\mathrm{GeV/c^2}, \mathrm{MRST}) = 446^{+33}_{-25}\,\mathrm{pb}, \tag{3.18}$$

while with the CTEQ parton distributions it is

$$\sigma_{pp\to t\bar{t}}^{NNLOapprox}(10\,\mathrm{TeV}, m_t = 172\,\mathrm{GeV/c^2}, \mathrm{CTEQ}) = 415^{+34}_{-29}\,\mathrm{pb} \tag{3.19}$$

The exact NLO cross section using MRST parton distribution functions is 414 ± 53 pb (the uncertainties are added in quadrature), while for the CTEQ parton distributions the cross section becomes 385 ± 51 pb. Recently, also the calculation of the top quark pair production cross section in the NNLO approximation at 7 TeV was performed and can be found in [84]. Compared to 10 TeV the cross section is reduced by a factor of about 2.4. For this thesis the center of mass energy of 10 TeV and the value of the NLO cross section is used.

3.3 Event generators for top quark pair production

As different event generators for top quark pair production exist, it is interesting to compare the properties of the top quarks produced by the different generators. Ideally, they should predict the same physics although internally the matrix element calculations are approached in different ways. Three generators are compared: `PYTHIA`, `ALPGEN` and `MadGraph/MadEvent`. The leading-order cross sections given by these three generators are respectively 240, 200 and 320 pb. Although this difference might be partially explained by the different top quark masses used in the three generators (respectively 172.4, 175 and 170.9 $\mathrm{GeV/c^2}$) [1], the difference is also related to the matrix element approach and the matching between the matrix element and parton showering. Figure 3.3 shows some properties of the generated top quark pair, the distributions have been normalized. The difference between the invariant mass of the $t\bar{t}$ system produced by the different generators is related to the different top quark masses used in the generation. The invariant mass of the $t\bar{t}$ system from the `PYTHIA` generator reaches lower values because the top quark has been generated with a non-zero width unlike what has been done in the other two generators. The distribution of the pseudo-rapidity of the (anti-)top quark reveals a difference between `PYTHIA` on one hand and `ALPGEN` and `MadGraph/MadEvent` on the other hand. The (anti-)top quarks generated with `PYTHIA` are produced slightly more central (η^{top} closer to zero). This effect is shown to be correlated with the difference in their transverse momentum. The (anti-)top quarks produced with `PYTHIA` are harder (higher p_T values) and the discrepancy becomes larger when considering the transverse momentum of the $t\bar{t}$ system. The origin of this effect is due to the difference in the matrix element generation, while `PYTHIA` uses $2 \to 2$ matrix element approximations, `MadGraph/MadEvent` and `ALPGEN` are able to generate matrix elements with extra partons in the final state. These extra partons take part of the transverse momentum available in the final state and the (anti-)top quarks thus

[1] At Next-to-Leading Order and around $m_t = 172\,\mathrm{GeV}/c^2$ at a center of mass energy of 14 TeV, for each addition of $1\,\mathrm{GeV}/c^2$ to the top quark mass, the cross section is reduced by 25 pb [82].

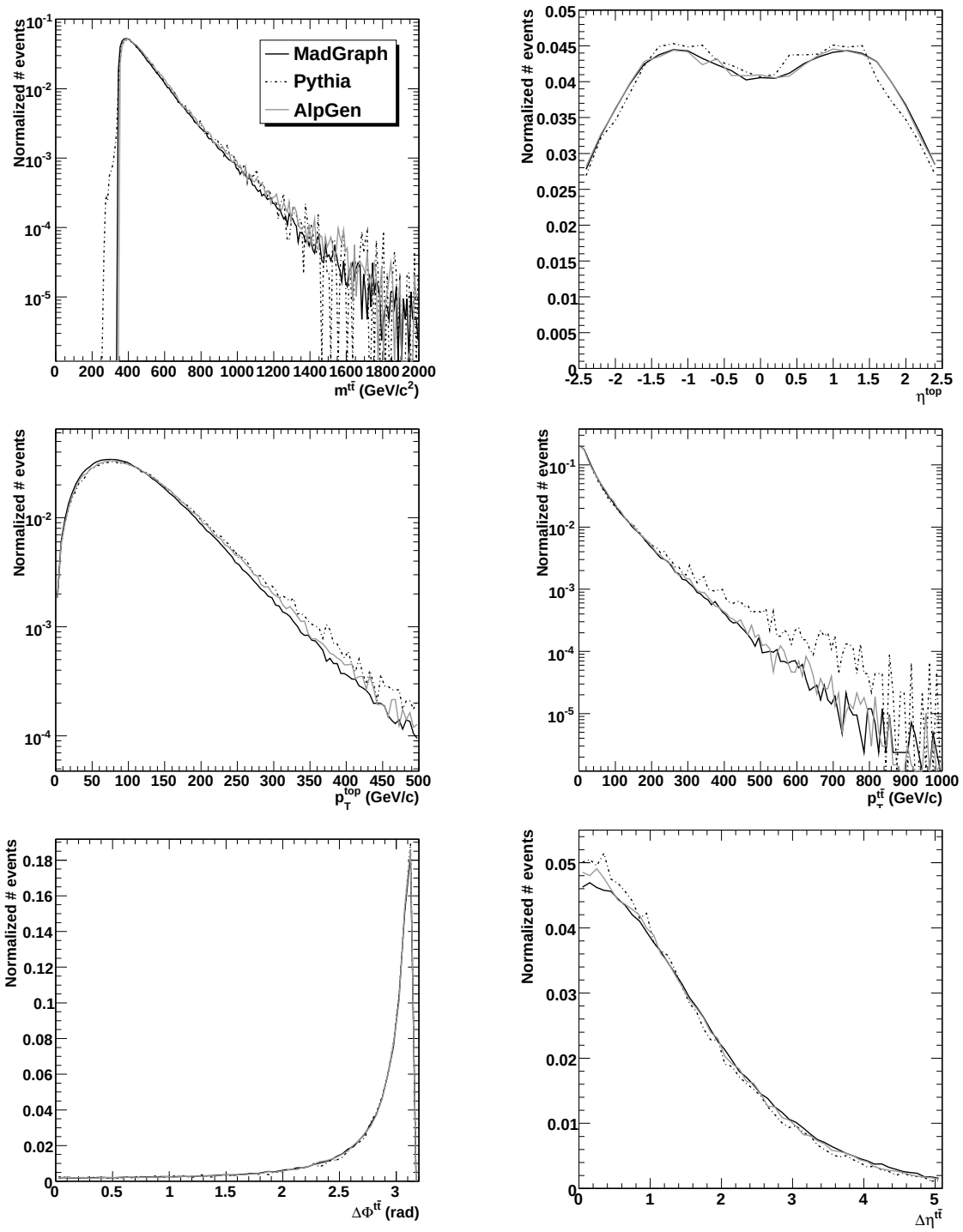


Figure 3.3: For three different generators different properties of the generated top quark pair are shown. The upper plots show the invariant mass of the $t\bar{t}$ system and pseudo-rapidity of the (anti-)top quark, while the plots in the middle show the transverse momentum of the (anti-)top quark and the $t\bar{t}$ system and the plots on the bottom show the difference in azimuthal angle and pseudo-rapidity between the top and the anti-top quarks.

end up with slightly lower transverse momenta. The difference between the azimuthal angle of the top and anti-top quarks and the absolute value of the difference between the pseudo-rapidity of the top and anti-top quarks are similar between the different event generators. The difference in azimuthal angle is peaked at π indicating that the top quark pair is produced back-to-back in the transverse plane. The absolute value of the difference in pseudo-rapidity is peaked at zero, which is consistent with top quark pairs that are boosted along the z-axis and thus end up with the same sign for the pseudo-rapidity. This also explains the two peaks in the pseudo-rapidity distribution of the (anti-)top quark. If the top quark pair would not be boosted along the beam line, a peak around zero would be observed instead of the dip.

In Figure 3.4, the effect of the value of the transverse energy E_T^{min} for the matching between matrix element (calculated with the `ALPGEN` generator) and the parton showering (performed with `PYTHIA`) is shown. No significant differences are observed for the different variables, indicating the stability of the resulting simulated events versus the matching threshold.

The pseudo-rapidities and transverse momenta of the generated top and anti-top quarks were also compared. Within the statistical uncertainties, the distributions of the pseudo-rapidities and transverse momenta of the generated top and anti-top quarks are compatible for the different generators.

3.4 Simulating events in CMS

After the event generation described in Section 3.1, the next step in the simulation of an event is the possible interaction of the final state particles with the detector and the simulation of the digitization phase in the registration of the observed event. A summary of the simulation of the detector response is given in Section 3.4.1. As large groups of physicists are interested in the same simulated event samples for their studies, the requests and simulation of the samples is coordinated centrally within the CMS collaboration, as discussed in Section 3.4.2. In the same Section, the list of simulated event samples needed for the physics studies in the next chapters, is provided with all relevant information.

3.4.1 Simulation of the detector response

In order to obtain realistic and observable final state physics objects from the simulated event, all relevant interactions when particles propagate through the detector need to be simulated. Both the effects of the interaction of the particles with the material in the detector need to be modelled as well as the simulation of the readout electronics. The material of the detector does not only include the material of the subdetectors (active and passive layers), but also the cables, support structures and the magnet. A full detector simulation is based on the `GEANT4` toolkit [85], which can achieve a high level of detail and precision. The total CPU-time that is typically needed to simulate a fully observable $pp \rightarrow X$ event is about 3 minutes and depends on the number of generated final state particles. The largest part of the CPU-time is used by the simulation of the material effects on the particles' trajectory, while the simulation of the readout elec-

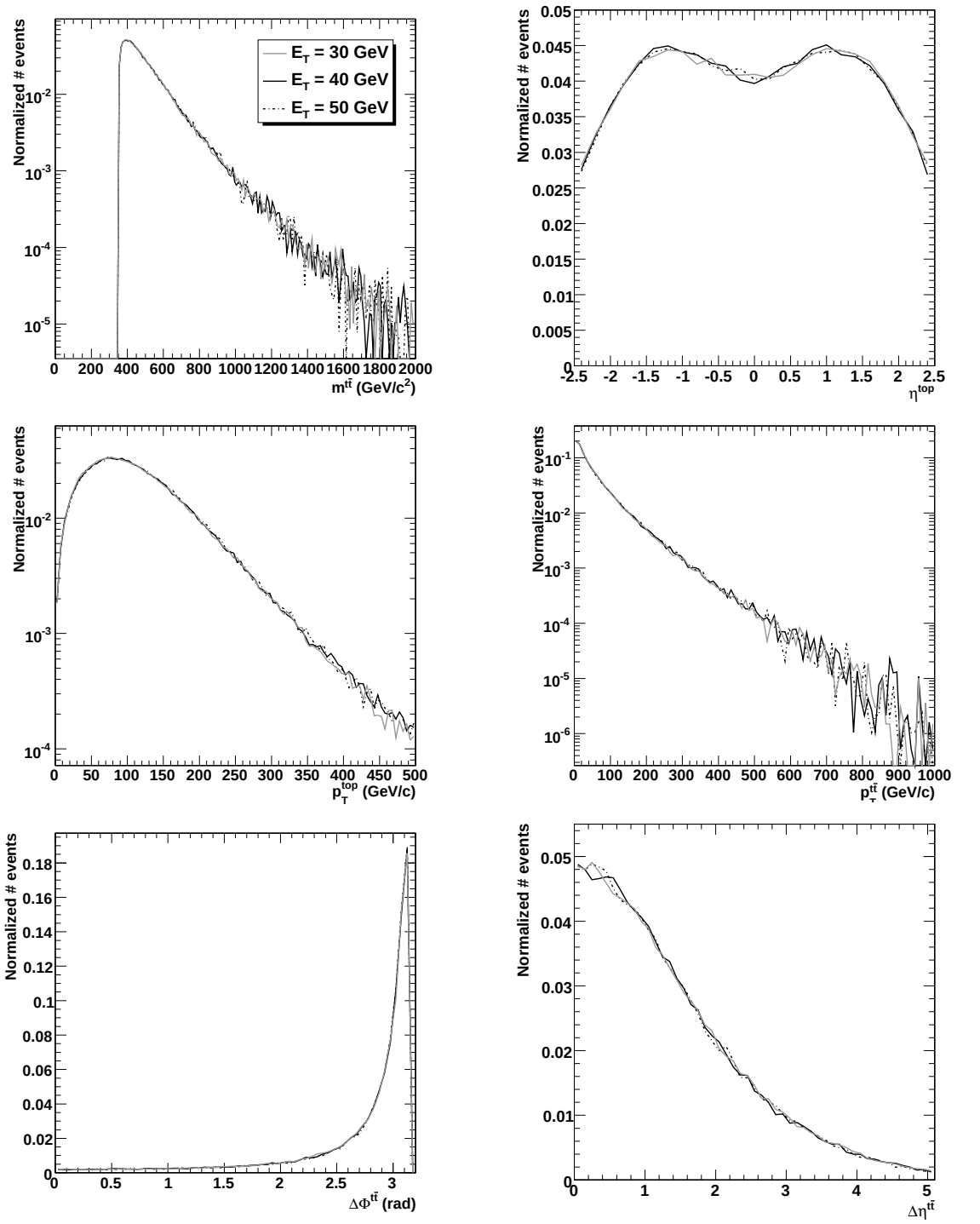


Figure 3.4: For three different matching thresholds the different properties of the generated top quark pair are shown. The upper plots show the invariant mass of the $t\bar{t}$ system and pseudo-rapidity of the (anti-)top quark, while the plots in the middle show the transverse momentum of the (anti-)top quark and the $t\bar{t}$ system and the plots on the bottom show the difference in azimuthal angle and pseudo-rapidity between the top and the anti-top quarks.

tronics is typically twenty times smaller. Of course, when very large event samples are needed or when a physicist quickly wants to check if a certain analysis is feasible, the relatively long CPU-time indicate that a faster simulation might be needed. Therefore, many experiments including the CMS collaboration, developed a fast simulation [86], which is tuned to and validated with full simulation results. To make the simulation faster, a simplified detector geometry is used for the tracking devices. The ECAL is considered as if it was a homogeneous medium after which the energy deposits are sliced longitudinally. The energy response of hadrons is derived from the full simulation of single pions and the energy is smeared using parametrized longitudinal and lateral shower profiles. Muons are not propagated until the muon chambers, the calorimetric response is tabulated and the response of the muon chambers is parametrized on samples of fully simulated single muons, to reproduce efficiencies and resolutions. The simplified detector geometry and the usage of parametrizations, give a gain of more than two orders of magnitude in CPU-time with respect to the full simulation.

3.4.2 Production of simulated event samples

The CMS software, CMSSW, contains both the full `GEANT`4-based simulation code as well as the fast simulation code. The production of large simulated samples, needed for detector performance studies, software validation and physics analysis, is carried out on the Grid infrastructures. A number of production teams share the available Tier1 and Tier2 sites to produce the workflows, which have been registered in a ProdRequest system by the physics groups. A production coordinator organizes, approves and assigns workflows to the different production teams based on the availability and reliability of the Tier1/Tier2 sites. Each production team makes use of the CMS ProdAgent multi-daemon workflow management tool [87]. ProdAgent interacts with the Data Management System (DMS) to discover which data needs to be processed or to register produced data. The DMS consists of the Dataset Bookkeeping System (DBS), which describes the existing data (files, blocks, datasets), and a Data Location Services (DLS), which keeps track of the location of the data. Production jobs are submitted by the operator of a production team to the Grid Workload Management System (WMS) which decides on the basis of a set of requirements to which Grid site the jobs are assigned. The jobs then enter sites through a computing element (CE), which distributes them between worker nodes (WN). Produced data are staged out into the local storage system. The CMS data transfer system, PhEDEx, assembles the produced files and transports the data to the analysis sites. Since it has been found that dealing with small files is inadequate for transfer and storage, a merge step has been implemented, to produce larger blocks of files [88]. The data is published into the data transfer system database and into the global DBS/DLS instance to make them available for transfer and to the collaboration for analysis. The components of ProdAgent are loosely coupled python daemons that cooperate and communicate through a mysql database. The work is split into these components that encapsulate specific functionalities. Scaling is then achieved by running in parallel any number of ProdAgent instances, which each make use of a local ProdAgent mysql database. The aim of the production operations [88, 89] is to deliver the largest possible amount of simulated events to the collaboration by optimally using the available distributed

computing resources. Therefore, the automation of the event production chain and operational expertise are key points. Other important aspects are the availability of the needed software release for the simulation and the validation of the configuration and production system before a large scale event simulation is started. After each production round optimizations are implemented.

The production of simulated event samples is done in so-called production rounds. During a certain production round, the same software and similar configurations are used for the production of the different event samples or workflows. The datasets that have been used for the physics study in this thesis are part of the Summer08 and Fall08 production round, while Winter09 datasets have been used to study systematics. The $t\bar{t}$ samples, produced with different event generators are summarized in Table 3.3. The

process	generator	m_{top} (GeV/c^2)	σ^{LO} (pb)	# events	$\mathcal{L}$ (fb^{-1})
$t\bar{t}$	PYTHIA	172.4	242	103 k	0.4 (0.25)
$t\bar{t}$+jets (20 GeV)	MG/ME	170.9	317	947 k	3.0 (2.3)
$t\bar{t}$0j excl. (30 GeV)	ALPGEN	175	118	149 k	1.3
$t\bar{t}$1j excl. (30 GeV)	ALPGEN	175	61.8	66 k	1.1
$t\bar{t}$2j excl. (30 GeV)	ALPGEN	175	20.6	30 k	1.5
$t\bar{t}$3j excl. (30 GeV)	ALPGEN	175	5.2	16 k	3.1
$t\bar{t}$4j incl. (30 GeV)	ALPGEN	175	1.6	11 k	6.9
$t\bar{t}$0j excl. (40 GeV)	ALPGEN	175	133	142 k	1.1
$t\bar{t}$1j excl. (40 GeV)	ALPGEN	175	52.1	56 k	1.1
$t\bar{t}$2j excl. (40 GeV)	ALPGEN	175	13.6	21 k	1.5
$t\bar{t}$3j excl. (40 GeV)	ALPGEN	175	2.7	19 k	7.0
$t\bar{t}$4j incl. (40 GeV)	ALPGEN	175	0.6	13 k	21.7
$t\bar{t}$0j excl. (50 GeV)	ALPGEN	175	144	151 k	1.0
$t\bar{t}$1j excl. (50 GeV)	ALPGEN	175	43.7	48 k	1.1
$t\bar{t}$2j excl. (50 GeV)	ALPGEN	175	9.3	16 k	1.7
$t\bar{t}$3j excl. (50 GeV)	ALPGEN	175	1.5	17 k	11.3
$t\bar{t}$4j incl. (50 GeV)	ALPGEN	175	0.3	11 k	36.7

Table 3.3: Overview of the centrally produced (Summer08/Fall08) full GEANT-4 simulated $t\bar{t}$ samples. MG/ME indicates the MadGraph/MadEvent generator. The number between brackets for the MadGraph/MadEvent and ALPGEN samples is the cut on the transverse momentum of the jets for the matching between matrix element and parton showering. For the MadGraph/MadEvent and PYTHIA samples also the integrated luminosity using the theoretical NLO cross section of 414 pb is given between brackets.

LO cross sections that are given in the table are the cross sections as obtained from the event generators themselves. As can be seen from the table, the cross section varies between about 200 pb for ALPGEN and 317 pb for MadGraph/MadEvent. The difference is related to the different handling of the processes in the different event generators and the different top quark masses that were used, as explained in Section 3.1. In the top quark physics group of the CMS collaboration it was decided to use NLO cross sections whenever they are available, and the values are summarized in [90]. For the matching between the ALPGEN matrix element generator and the parton showering in PYTHIA different thresholds on the transverse energy E_T^{min} of the jets are applied.

In Table 3.4 the simulated datasets are listed that are considered in this thesis as background processes for top quark pair production. Also these datasets were produced using the full GEANT4-based simulation. For the different single top channels, NLO cross sections are given, while for the other samples the cross sections are LO. For the background processes in Table 3.4 no NLO cross section predictions are available from literature. In the next chapters LO cross sections are thus used for these processes, while NLO cross section predictions are used for the processes for which they are available. The different $\hat{p_T}$ ranges for the QCD process generated with MadGraph/MadEvent correspond to the allowed transverse momentum as obtained in the rest frame of the hard interaction during the simulation of the events. The relevant simulated event

process	generator	σ^{LO} (pb)	# events	$\mathcal{L}$ (fb^{-1})
single top				
s-channel (leptonic decay)	MG/ME	1.6 (σ^{NLO})	12 k	(7.5)
t-channel (leptonic decay)	MG/ME	41.6 (σ^{NLO})	282 k	(6.8)
tW-channel	MG/ME	29 (σ^{NLO})	169 k	(5.8)
W+c	MG/ME	$1.5\,10^3$	2.9 M	2
V+QQ (V=W/Z and Q=b/c)	MG/ME	290	1 M	3.3
QCD ($\hat{p_T}$ bins)				
100-250 GeV/c	MG/ME	$15\,10^6$	12 M	$0.82\,10^{-3}$
250-500 GeV/c	MG/ME	$4\,10^5$	5 M	$13\,10^{-3}$
500-1000 GeV/c	MG/ME	$14\,10^3$	5 M	0.33
1000-∞ GeV/c	MG/ME	370	1 M	2.9
$pp \to \mu + X$ ($p_T^\mu > 15\,GeV/c$)	PYTHIA	$122\,10^3$	6 M	$49\,10^{-3}$

Table 3.4: Overview of the centrally produced Summer08/Fall08 full GEANT4 simulated background samples. MG/ME indicates the MadGraph/MadEvent generator. The values for the cross sections of the different single top channels are obtained from [91–94].

samples produced with the fast simulation of the CMS collaboration are given in Table 3.5. A K-factor of 1.14 for the W+jets and Z+jets processes was used to scale from

process	generator	m_{top}(GeV/c^2)	σ^{NLO} (pb)	# events	$\mathcal{L}$ (fb^{-1})
nominal $t\bar{t}$+jets sample					
$t\bar{t}$+jets (20 GeV)	MG/ME	170.9	414	3 M	6.7
ME/PS E_T^{min} threshold					
$t\bar{t}$+jets (10 GeV)	MG/ME	170.9	414	1 M	2.6
$t\bar{t}$+jets (40 GeV)	MG/ME	170.9	414	2 M	4.4
factorization scale					
$t\bar{t}$+jets (Q^2 down)	MG/ME	170.9	414	1.2 M	2.8
$t\bar{t}$+jets (Q^2 up)	MG/ME	170.9	414	1.2 M	2.8
nominal $t\bar{t}$ sample					
$t\bar{t}$	PYTHIA	172.4	414	4 M	9.7
ISR/FSR					
$t\bar{t}$+jets (more ISR/FSR)	PYTHIA	172.4	414	2.0 M	4.9
$t\bar{t}$+jets (less ISR/FSR)	PYTHIA	172.4	414	2.1 M	5.1
W+jets	MG/ME		$45.6\,10^3$	94 M	2.1
Z+jets	MG/ME		$4.2\,10^3$	8 M	1.8

Table 3.5: Overview of the centrally Winter09 and privately produced samples using fast simulation. MG/ME indicates the MadGraph/MadEvent generator. (Scaled) NLO cross sections are used to calculate the available integrated luminosity for each sample. For the W/Z+jets samples the LO cross section is taken from [95] and the K-factor of 1.14 to scale the LO cross section to NLO is derived from [96].

LO to NLO cross sections as obtained from [96]. The PYTHIA parameters that have been introduced in Section 3.1.3 were altered to produce the samples with higher and lower Q^2 scale, with values summarized in Table 3.6. The PYTHIA event samples were privately produced, since such samples were not available from the central production. The configuration file of the centrally produced samples was used and the relevant parameters have been added or changed to increase/decrease the ISR and FSR. The PYTHIA parameters and their values are listed in Table 3.7.

parameter value	nominal	Q^2 down	Q^2 up
PARP(64)	1.0	0.25	4.0
PARP(72)(GeV)	0.25	0.5	0.125

Table 3.6: The values of the PYTHIA parameters that have been altered for the production of the systematics samples with a varying Q^2 scale.

parameter value	nominal	more ISR/FSR	less ISR/FSR
PARP(64)	0.2	1.0	0.2
PARP(61)(GeV)	0.25	0.35	0.25
PARP(72)(GeV)	0.25	0.35	0.25
PARJ(81)(GeV)	0.25	0.35	0.25
PARP(67)	4.0	4.0	2.5

Table 3.7: The values of the PYTHIA parameters that have been altered for the production of the systematics samples with an increased or decreased amount of ISR and FSR.

3.5 CMS Event Data Model

The software of the CMS collaboration CMSSW, is based on the Event Data Model (EDM) [97]. Additionally, services exist that are needed by the simulation, calibration, alignment and reconstruction modules. This framework aims to facilitate the development and deployment of reconstruction and analysis software. The simulation of the interaction of the generated particles with the detector produces output in the same format as one would obtain with real collisions. Furthermore, the same executable (cmsRun) and many of the plugin modules, which contain for instance the reconstruction algorithms, are used both for real collisions (detector data) as for simulated collisions (Monte Carlo data). The executable is steered by a configuration file containing the relevant information for the executable to run, such as the data that should be used as input, the modules that it should execute, the parameter settings to use for each module, etcetera. The central concept of the Event Data Model is the "Event". An Event is a C++ object container for all RAW and reconstructed data related to a particular collision. The RAW data events contain the detector data and trigger results and some higher-level objects created during HLT processing. The Event is the central object and all modules access the data through the Event. Information produced by the modules is added to the Event in separate containers and the resulting ROOT file is saved by means of a dedicated module. The collection of data from an Event coming from the detector or simulation is called the RAW data, while the collection of data after the reconstruction modules added their information to the Event is referred to as RECO data. The combination of RAW and RECO data is the full event data (FEVT).

Usually a subset of the RECO data, containing the relevant Analysis Object Data (AOD), is provided to the physics analysis users. The AOD format is much lighter in size and is sufficient for most physics studies.

In early 2007, the Top Quark Analysis Framework (TQAF) was developed by and for the CMS Top Physics Analysis Group [98], aiming at an efficient, flexible and user friendly analysis framework. The framework was efficient in preventing code duplication between physics analyses based on it, but also in performance speed. Secondly, the flexibility of the TQAF avoided to constrain physics analyses in the use of objects and tools. New modules could be easily added or adapted to changes in the underlying CMS software. The user friendliness of the TQAF made it possible for new users to start doing physics analysis, while avoiding a too steep learning curve. The TQAF contains two basic layers. TQAF Layer 1 contains final state physics objects inheriting from AOD (or RECO) to which additional information is added relevant for top quark physics analyses. The second layer of the TQAF aims in reconstructing the final state of top quark events with the objects from the first layer. Therefore, tools are provided for additional selection, kinematic fitting, jet combination, final state likelihood analysis, etc. The top quark pair topology was reconstructed in this second layer, including the full decay tree.

Based on the TQAF Layer 1, a CMS-wide analysis model has been developed, that has been called the Physics Analysis Toolkit (PAT). The goals of the PAT are essentially the same as for the TQAF Layer1: reducing the time CMS users spend in learning about computing structures and providing flexibility to the advanced user, while code duplication is avoided. The PAT candidates are, as in TQAF Layer 1, the reconstructed high-level physics objects such as jets, leptons and photons. Reconstruction variables described in the next chapter, like for instance object isolation, object identification, jet energy correction factors, b-tagging information for jets, tracks associated to a jet, matched generator or trigger object information is added to the PAT candidates, which inherit from the RECO or AOD reconstructed objects. Relevant information is embedded or references are used and the unneeded information is dropped. In this way the event size is reduced from 500 kB/event (100 kB/event) in RECO (AOD) format to 16 kB/event in PAT format [99, 100].

The TQAF Layer 1 was replaced by the PAT and the TQAF Layer 2 evolved from 2007 until now based on the PAT. For the simulated collisions, the TQAF Layer 2 reconstructs the generated hard interaction in which the top quark pairs were produced. The initial four-momenta of the partons in the hard interaction are kept as well as the four-momenta of the the daughters from the top quark decay. Also access to the partons from ISR and FSR is ensured. Based on the generator information, filters have been constructed based on the decay topology (fully hadronic, semi-leptonic or fully leptonic). A second important class in the TQAF Layer 2 is the "TtSemiLeptonicEvent" class [101] containing the relevant high level analysis information for $t\bar{t}$ event topologies in the semi-leptonic decay channel. Several so-called "hypotheses" for the jet-parton assignment have been implemented based on kinematic fitting, generator information, event shapes, Multi Variate Analysis (MVA) techniques such as likelihood ratio variables, etc. The usage of unified code facilitates the comparison of physics results and reduces significantly the time spent in implementing and debugging code.

Therefore, the structures mentioned above have been developed and are used for the physics analysis in the next chapters.

We have found a strange footprint on the shores of the unknown. We have devised profound theories, one after another, to account for its origins. At last, we have succeeded in reconstructing the creature that made the footprint. And lo! It is our own.

Sir Arthur Eddington

Chapter 4

Reconstruction of physics objects

The proton collisions at the LHC will provide diverse particles in the final state, basically all Standard Model particles (and hopefully others) will be produced. Most of these particles will decay or fragment in the CMS detector and both the initial particles as well as their decay products will interact with the different subdetectors they traverse, leading to a huge amount of electronic signals. The first step to study the physics phenomena of the observed collisions is to combine the electronic signals into objects reflecting the final state particles that were produced in the collision. This reconstruction of the final state particles is performed in different steps. In a first stage the electronic signals of a particular subdetector are combined to form "raw objects" that are used to reconstruct the high level "physics objects", usually from the combined information of several subdetectors.

In hadron collisions the majority of final state particles are quarks and gluons. Due to confinement, these coloured partons are not directly observable and will fragment into a collimated bunch of stable hadrons, a so-called hadronic jet. It are these stable particles that are detected by the calorimeters and, if they carry electric charge, by the tracking system. The precise reconstruction of the four-vectors of the initial partons from the observed fragments is one of the major challenges in experimental particles physics. Apart from quarks and gluons, also leptons will be produced in the final state. By combining the information from the calorimeters with the information of the tracking system, it is possible to reconstruct electrons and photons. The CMS detector has an extensive muon system surrounding all other subdetectors. Muons are first reconstructed locally in the muon system after which the information of the tracking system is added. The transverse energy of particles that cross the experiment undetected, like for instance neutrinos, can be computed from the total transverse energy balance, provided that the detector is hermetic.

In Chapter 2 an overview of the different subdetectors of the CMS experiment was given. This chapter will elaborate on how the output of the electronic signals registered by the subdetectors are combined in order to reconstruct the high level physics objects in the final state of interest. The analysis in the following chapters is

based on the $pp \rightarrow t\bar{t} \rightarrow bW\bar{b}W \rightarrow bq\bar{q}\bar{b}\mu\bar{\nu}_{\mu}$ process. The focus will therefore lie on the reconstruction of jets, muons and missing transverse energy, while the reconstruction of electrons, photons and taus can be found elsewhere [36, 102–104]. In Section 4.1 an overview is given of the different algorithms and their performance to reconstruct jets within the CMS detector. Details are given on the identification techniques of jets orginating from heavy bottom quarks. The reconstruction of the muons is discussed in Section 4.2. Based on simulated events, the resolutions on the direction of the muon and its transverse energy are shown. Although it will not be used in the analysis to measure the jet energy scale corrections (Chapter 6 and 7), the reconstruction of the missing transverse energy is summarized in Section 4.3.

4.1 Jet reconstruction

Different methods are described in 4.1.1 to cluster input objects in order to reconstruct the four-momentum of the original parton from a hard scattering. The origin of the bias on the direction of the reconstructed jets with respect to the direction of the partons of which they originate is discussed in 4.1.2. Section 4.1.3 elaborates on the performance of the jet-finding algorithms. Jets originating from a b quark have a distinct signature and a review of possible methods to identify these b-jets is given in 4.1.4. On top of the jet reconstruction a calibration of the jet energies is often required and the CMS collaboration opted for a factorized approach described in Section 4.1.5. In Section 4.1.6 the resolution on the direction and the transverse energy of the reconstructed jets is determined.

4.1.1 Jet-finding algorithms and jet types

In the past years many jet algorithms have been developed and tuned to speed up the jet reconstruction process, while keeping in mind the main goal of providing an optimal correspondence between the partons and the stable particles in which they fragment. Furthermore it is important that the jet algorithm is insensitive to the addition of soft radiation (infrared safe) and collinear splitting of particles (collinear safe). An infrared safe algorithm yields the same list of jets in an event independent of the presence of soft radiation. A collinear safe algorithm should find the same list of jets independent if there are two collinear particles or if they are merged into a single particle carrying the same total energy.

In order to compare the performance of jet algorithms in different situations one needs to be sure that the jet clustering algorithm is specified in a non-ambiguous way, as proposed in [105]. First the jet definition, consisting of the jet algorithm together with its parameters and the recombination scheme (the scheme to recombine the four-momenta of the clustered objects) should be given. Secondly the kind of objects used as input for the jet algorithm should be specified. Following this proposal first the jet clustering algorithms applied in the reconstruction software of the CMS experiment are discussed below and secondly the input objects for the different types of jets are specified.

Jet finding algorithms

Two main categories of jet-finding algorithms can be distinguished, cone-type algorithms and sequential clustering algorithms. They cluster the observed objects in the detector differently.

- **Cone-type algorithms** are based on the idea that during the fragmentation of a parton, the hadrons form a collimated shower. The stable objects can thus be clustered by defining a cone of a certain radius R in $(\eta \times \phi)$ or $(y \times \phi)$ space surrounding all trajectories of the objects coming from the same parton. There are two types of cone algorithms, those that are based on seeds to define an initial trial cone and the seedless algorithms.

  - **The Iterative Cone (IC) algorithm**
    The initial trial cone is constructed around a seed direction, usually an object with a transverse energy above a predefined threshold E_T^{seed}. This makes the algorithm infrared and collinear unsafe because when soft radiation with a transverse energy exceeding the seed threshold ($E_T > E_T^{seed}$) is added, one can end up with another set of jets for the event. The advantage of this algorithm is that it is fast and therefore often used to trigger on jets. The input objects are first ordered in ascending transverse energy. The object with the highest transverse energy is used as a seed and compared to the seed threshold E_T^{seed}, taken to be 1 GeV for the CMS version of this algorithm. If it passes the threshold, the E_T recombination scheme is used to cluster the input objects in a trial cone around the direction of the seed. In this recombination scheme the transverse energy of the so-called proto-jet is calculated as the sum of the transverse energy of its constituents: $E_T = \sum_j E_{T,j}$. The direction of the proto-jet is obtained from the directions of the constituents. For the pseudo-rapidity this is $\eta = \sum_j E_{T,j}\eta_j / \sum_j E_{T,j}$ and for the azimuthal angle $\phi = \sum_j E_{T,j}\phi_j / \sum_j E_{T,j}$. The obtained direction of the proto-jet is compared with the direction of the initial proto-jet and it is called stable if it changes by less then 0.001 compared to the direction of the initial trial cone in the $(\eta \times \phi)$ space. If this is the case, the clustered objects are removed from the list of input objects and the proto-jet is added to the list of stable jets. If the proto-jet is not stable, the input objects that enter the cone with the new direction of the proto-jet are also clustered. This procedure is repeated until all input objects exceeding the seed threshold have been used. Apart from the seed threshold the IC algorithm has a second parameter, namely the cone radius. In the CMS experiment jets are reconstructed with a cone radius $R = 0.5$ (IC5).

  - **The Seedless Infrared Safe Cone (SC) algorithm**
    The SC algorithm used in CMS is an external plugin [106] that is part of the FastJet package [107]. The usage of the FastJet package makes it possible to use exactly the same algorithm in different experiments which facilitates the comparison of the obtained results. The main idea behind the algorithm is to search for all stable cones by adding the four-momenta of

the input objects in a trial cone and comparing the direction of that proto-jet with the direction of the clustered constituents. Like in the Iterative Cone algorithm, a proto-jet is again called stable when the direction does not change compared to the summed direction of its constituents. If the proto-jet is not stable, the procedure is repeated by clustering all the input objects within a cone around the summed direction of the constituents. The advantage of searching all stable cones is that the algorithm is infrared safe because no seeds are used, which means that the direction of every constituent is used to construct the initial trial cone. Of course this leads to the disadvantage is that the computation time can increase quickly with the number of input objects to be clustered. The computation time is reduced by considering for every pair of objects within a distance of $2R$ from each other, an $(y \times \phi)$-circle with radius R enclosing them. All stable cones containing this pair of objects can be calculated from the circles that are found by moving around the initial circle while making sure that its content remains unchanged [108]. These stable cones are then called proto-jets and at this stage input objects can be shared between the different proto-jets. Therefore a splitting and merging technique is applied. For this the proto-jet i with the highest transverse momentum $p_{T,i}$ is taken and the proto-jet j with the highest transverse momentum $p_{T,j}$, sharing input objects with proto-jet i, is identified. The scalar sum of the transverse momenta of the objects that are shared $p_{T,shared} = \sum_k p_{T,k}$ is then divided by the $p_{T,j}$ of the softest jet. This value is compared to the overlap threshold of 0.75. If the $p_{T,shared}/p_{T,j}$ is greater then 0.75, the two proto-jets are merged and the new proto-jet is added to the list of proto-jets. In the other case, each shared input object is assigned to the jet who's axis is closest and the resulting two new proto-jets are added to the list of proto-jets. When a proto-jet does not share any input object with other proto-jets, it is added to the list of jets. The procedure is repeated until no proto-jets and no input objects are left. For this algorithm jets have been reconstructed with cone radii of R =0.5 (SC5) and R =0.7 (SC7) in $(y \times \phi)$ space.

- **Sequential clustering algorithms** like the **inclusive longitudinally invariant k_T algorithm** [109] are also part of the FastJet package [107]. The inclusive longitudinally invariant k_T algorithm makes use of the distances between any two input objects i and j and between any input object i and the beam. These distances are defined as:

$$d_i = k_{\mathrm{T},i}^2 \qquad \text{and} \qquad d_{ij} = \min\left\{k_{\mathrm{T},i}^2, k_{\mathrm{T},j}^2\right\} \Delta R_{ij}^2/D^2, \tag{4.1}$$

where $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ is the distance in $(y \times \phi)$ space between the two input objects, $k_{T,i}$ is the transverse energy of input object i and D reflects somewhat the radius of the jet. The clustering algorithm proceeds by searching for the minimum distance among all distances. In case this minimum is d_{ij}, the four-momenta of object i and j are summed to form a new object. The new object is added to the list of input objects while the two original ones, i and j, are removed. In the case the minimum is d_i, the object i is moved to the

list of jets. The process is continued until all input objects are clustered. The algorithm is infrared and collinear safe. Jets are reconstructed with parameters D =0.4 (KT4) and D =0.6 (KT6).

Input objects and jet-types

The traditional strategy is to reconstruct the four-momenta of the original partons from the energy deposits in both the hadronic and the electromagnetic calorimeter, taking into account the finer granularity of the latter. The energy deposits of each HCAL cell and the in $(\eta \times \phi)$-space geometrically associated ECAL crystals are combined into a single calorimeter tower. In the barrel region the signals from 5x5 ECAL crystals are added to the signal of one HCAL cell. The geometry of the calorimeters outside the barrel region is somewhat more complex, but also in that case several electromagnetic crystals are assigned to each HCAL cell. The granularity of the calorimeter towers is shown in Figure 4.1. Electronic noise is suppressed by neglecting cells or crystals that do not pass the "Scheme B" energy thresholds [110] shown in Table 4.1. In addition to the threshold on the individual cells a threshold on the calorimeter towers is applied of $E_T = 0.5\,\mathrm{GeV}$ to reduce the energy contribution coming from the underlying event.

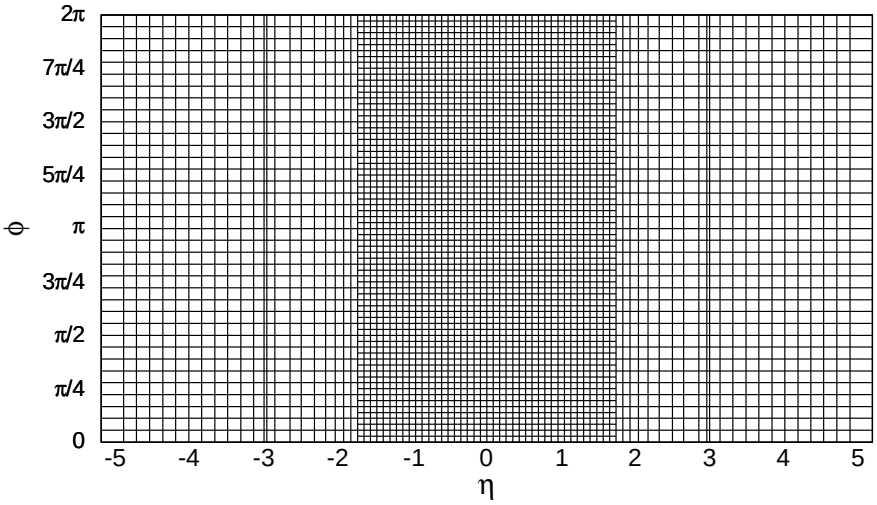


Figure 4.1: Layout of the calorimeter tower geometry in the $(\eta \times \phi)$ space.

Calorimeter cell thresholds (GeV)					
	HB	HO	HE	$\sum$EB	$\sum$EE
Scheme B	0.90	1.10	1.40	0.20	0.45

Table 4.1: "Scheme B" energy thresholds for the suppression of noise in the calorimeter cells. The notations $\sum$EB and $\sum$ EE refer to the sum of the crystals of the electromagnetic calorimeter associated with the same tower in the barrel and in the endcaps respectively, while HB, HO and HE were defined in Chapter 2.

.

Beside the traditional calorimeter jets, there are several other types of jets deployed in the CMS experiment, depending on what input objects are used for the jet reconstruction algorithm. The different jet-types are defined below [111].

- **Calorimeter jets (CaloJets) and Jet-plus-track jets (JPTJets)**
  For the reconstruction of the calorimeter jets only the calorimeter towers are used, therefore they are called CaloJets. The jet-finding algorithm uses calorimeter towers as the input objects with the direction of the tower as observed in the calorimeter. One can also assign tracks to the CaloJets based on the distance between the jet axis and the track measured at the interaction vertex [112]. The tracks associated with the jet are then projected on the surface of the calorimeter to decide for each track if it lies in the cone of the jet (in-cone tracks) or if the magnetic field has bent the track out of the cone (out-cone tracks). The resolution and the value of the transverse momentum of these JPTJets is improved by taking properly into account the transverse momenta of the in-cone and out-cone tracks associated to the jet.

- **Particle Flow jets (PFJets)**
  Another possibility consists in reconstructing and calibrating all the individual particles produced in the collision and this by using the information of the tracking system and the calorimeter towers together, a procedure called Particle Flow [113]. The reconstructed particles are then used as input objects by a jet algorithm, with the direction of the input particles taken at the interaction vertex, hence before the bending by the 3.8 T magnetic field. The resulting jets are called PFJets.

- **Jets from generated particles (GenJets)**
  In simulated proton collisions one can use the stable particles, after hadronization but before their interaction with the detector or the bending by the magnetic field, as input objects for the jet-finding algorithm and obtain so-called GenJets. By comparing the jets reconstructed with the same jet-finding algorithm at different levels of the simulation, the effect of the interaction of the particles with the detector and the magnetic field can be studied and the performance of the jet algorithm can be determined.

In this thesis CaloJets are used, but the conclusions are equivalent for JPTJets and PFJets. The definition of an optimal jet-finding algorithm does not only depend on the choice of its parameters, like for instance the cone radius or the recombination scheme, but also on the physics process under consideration. A jet-finding algorithm that is optimal for the reconstruction of multijet events might not give the same performance for the reconstruction of the partons in $t\bar{t}$ events, for which there is on average more energy deposited in the central part of the calorimeters. The average transverse energy deposited in the calorimeters as a function of the pseudo-rapidity for multi-jet events and $t\bar{t}$ events is shown in Figure 4.2. A grid based on the tower geometry has been used for the pseudo-rapidity dimension, where every bin reflect a tower as shown in Figure 4.1. As a comparison also the average transverse energy of all the calorimeter towers that are clustered in jets by the SC5 jet algorithm is shown. For $t\bar{t}$ events

there is on average more energy deposited compared to the energy deposited in multi-jet events, in particular for the central region. In the forward region, the average energy deposition is comparable, because it is related to the energy deposition of the underlying event. For both multi-jet and $t\bar{t}$ events almost all energy depositions are clustered, the observed difference between the clustered transverse energy and the total energy deposits is due to the applied thresholds on the individual calorimeter towers.

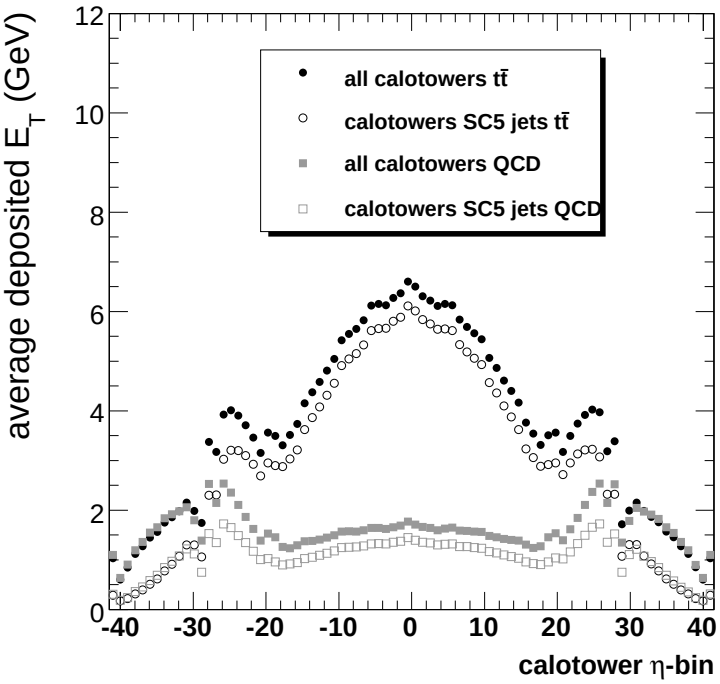


Figure 4.2: The average transverse energy deposited in the calorimeters in an (η) space reflecting the number of towers away from $\eta = 0$,for multijet and $t\bar{t}$ events.

Figure 4.3 shows the fraction of calorimeter towers assigned to jets for the different jet clustering algorithms. For multi-jet events the fraction of calorimeter towers that is clustered in a jet is clearly lower than for $t\bar{t}$ events, because on average more towers are removed due to the "Scheme B" thresholds for multi-jet events compared to $t\bar{t}$ events.

In the CMS experiment jets have been reconstructed with several jet-finding algorithms and various parameters for these algorithms. In Figure 4.4 the number of constituents for each reconstructed CaloJet and the number of CaloJets per event for $t\bar{t}$ events is shown for three algorithms. Only jets with a transverse momentum exceeding 10 GeV/c and within the tracker acceptance are considered. The number of constituents per jet is very similar for the algorithms with comparable conesizes, while the number of jets per event is very different for the considered algorithms. The discrepancy between number of reconstructed jets with the Iterative Cone algorithm or with the two other algorithms comes from the usage of a seed threshold in the jet reconstruction. Less jets are reconstructed in the event because at some point in the algorithm no more seeds are passing the seed threshold.

The jets reconstructed in the semi-muonic and fully hadronic decays of the $t\bar{t}$ events have been used in Figure 4.5, to compare the number of constituents per jet recon-

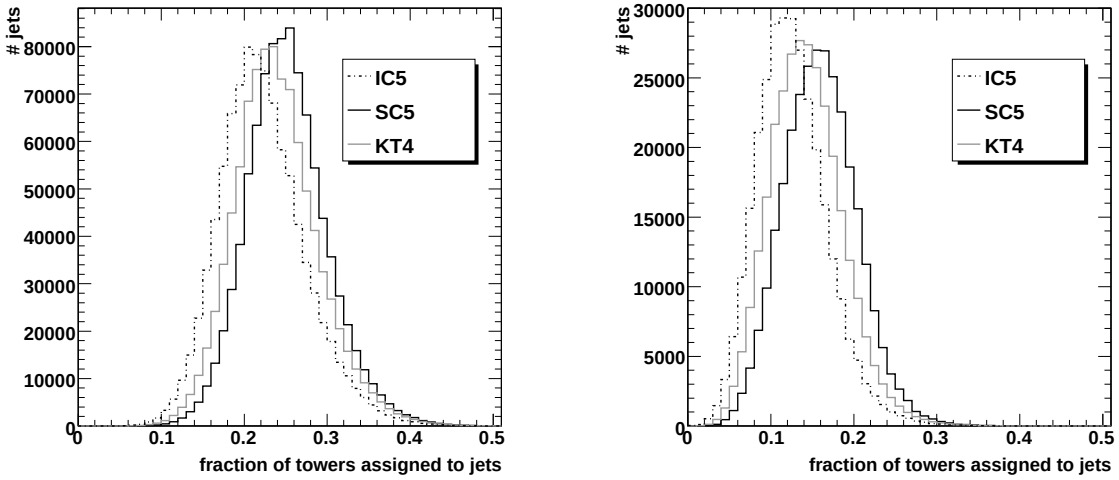


Figure 4.3: The fraction of calorimeter towers with a non-zero energy deposit that are assigned to jets reconstructed with the different jet algorithms for $t\bar{t}$ events (left) and for QCD events (right).

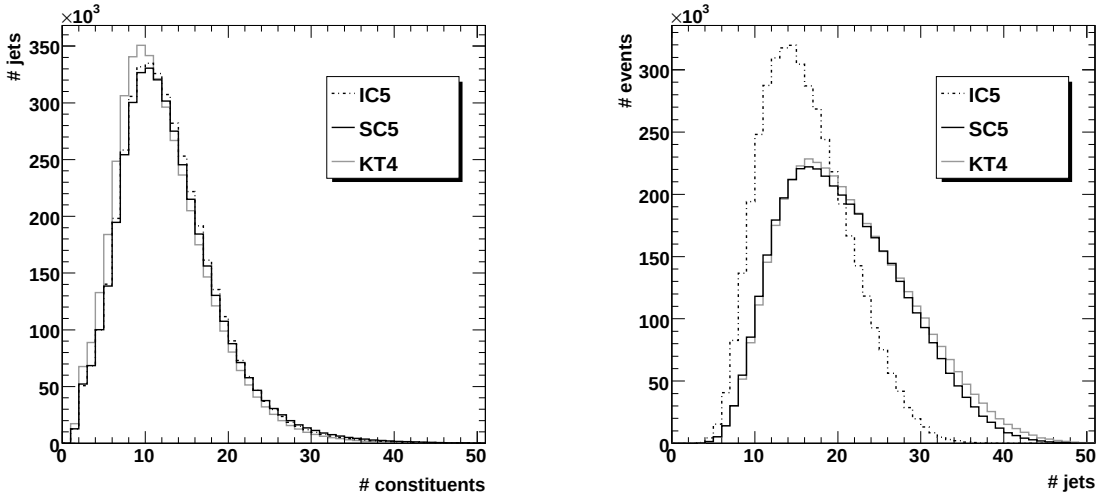


Figure 4.4: Number of constituents per CaloJet (left) and the number of jets per event (right) for the CaloJets reconstructed with three different jet algorithms in $t\bar{t}$ events.

structed by the IC5 and the KT4 algorithms with the number of constituents per jet reconstructed by the SC5 algorithm. Only jets that were reconstructed by all three jet algorithms and that are matched with the partons in the $t\bar{t}$ event better than $\Delta R_{jet,parton} = \sqrt{(\eta_{jet} - \eta_{parton})^2 + (\phi_{jet} - \phi_{parton})^2} = 0.5$ have been considered. There is a clear correlation between the number of constituents of the jets that were reconstructed by different jet algorithms, indicating that in case a jet is found by all three algorithms, mostly the same calorimeter towers were used for the jet reconstruction.

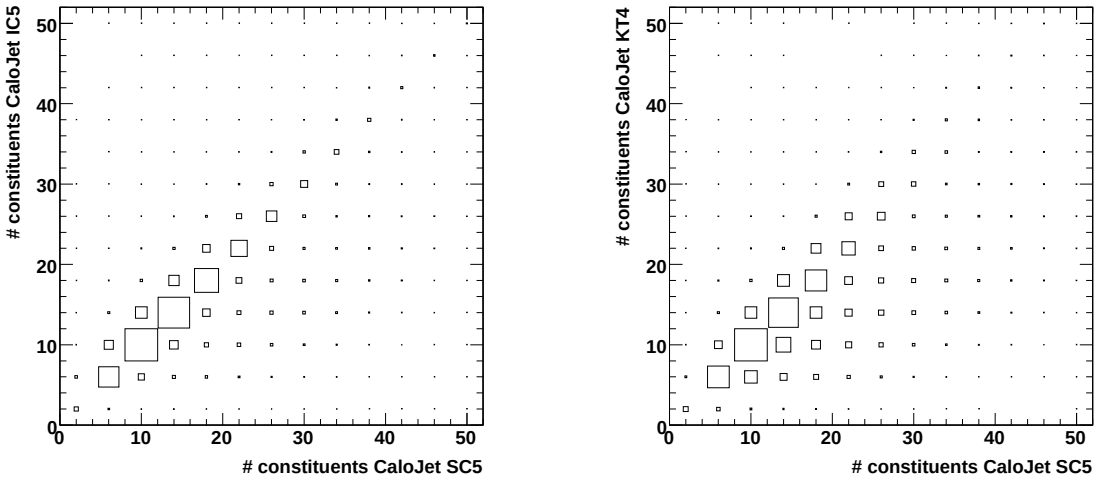


Figure 4.5: Comparison of the number of constituents of matched CaloJets in $t\bar{t}$ events reconstructed with different jet algorithms. The size of the box corresponds with the number of jets.

Some typical jet properties of jets reconstructed in the semi-leptonic and fully hadronic decays of $t\bar{t}$ events are discussed. Figure 4.6 shows the typical size of the jets reconstructed by the different jet algorithms. For each jet matching better than $\Delta R = 0.3$ to a parton, the distance in ΔR between the jet axis and its furthest constituent is given. For the jets reconstructed with the IC5 algorithm, a sharp cut-off at 0.5 is seen. Due to the splitting and merging procedure in the SC5 jet reconstruction algorithm, a long tail towards larger distances between the jet and its constituents is visible. The distribution for the KT4 jet clustering algorithm is centered around a distance of 0.4, as defined by the parameter D in the clustering procedure. In Figure 4.7 the matching efficiency as a function of the transverse momentum of the CaloJet is shown. The CaloJets with $|\eta| < 2.4$ are matched with the GenJets by requiring $\Delta R < 0.25$. The efficiency is close to 100 % for jets with $p_T > 30\,\mathrm{GeV/c}$ that were reconstructed with the IC5 or KT4 algorithm, while it exceeds 95 % for jets reconstructed with the SC5 jet algorithm. More fake CaloJets are expected for the SC5 jet algorithm because no seed threshold is applied. The reduced matching efficiency of CaloJets to GenJets for transverse momenta below 30 GeV/c is related to the magnetic field as particles with lower transverse momentum are more strongly bent and end up outside the reconstructed jet area. This is also visible in Figure 4.8, where the ratio of the number of constituents for CaloJets and matched GenJets is shown as a function of the transverse momentum of the GenJet. For higher transverse momenta the ratio

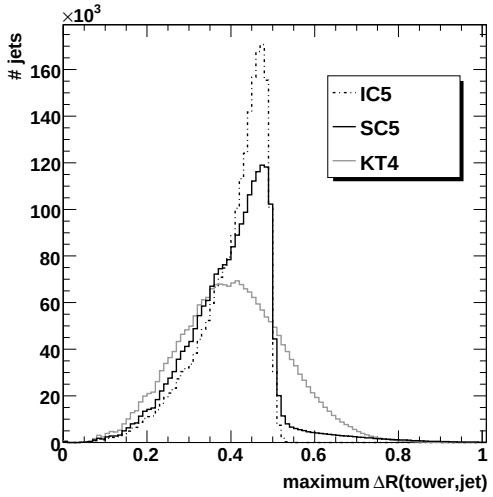


Figure 4.6: Maximum distance in ($\eta \times \phi$) space between the jet axis and its calorimeter towers.

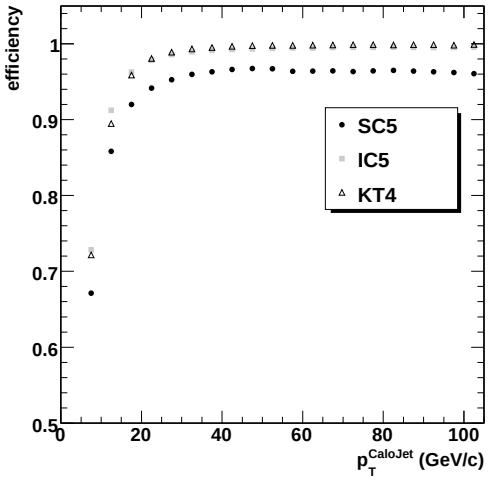


Figure 4.7: Efficiency of matching a CaloJet to a GenJet as a function of the p_T of the CaloJet.

converges to a value that is more or less constant, but below unity, indicating that several particles end up in the same calorimeter tower.

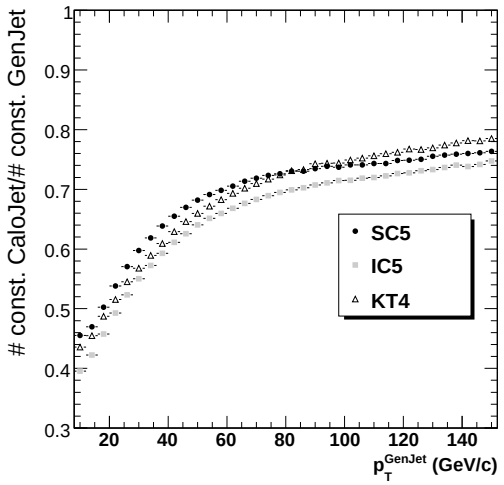


Figure 4.8: The ratio of the number of constituents for CaloJets and matched GenJets as a function of the p_T of the GenJet.

In order to compare the pseudo-rapidities and the transverse momenta of the jets that can be matched or not to the partons from which they originate, we consider only the events without initial and final state radiation. The difference between matched and unmatched jets is of importance in the $pp \rightarrow t\bar{t}$ processes and particulary in the $t\bar{t} \rightarrow bW\bar{b}W \rightarrow bq\bar{q}\bar{b}\mu\bar{\nu}_\mu$ decay that will be used for the analysis in the next chapters. To increase the statistics, also the $t\bar{t}$ final state that consists only of jets (fully hadronic decay) is considered. The CaloJets reconstructed with the SC5 jet algorithm are required to match better than $\Delta R = 0.3$ to the partons. The η and p_T distribu-

tions of the matched and unmatched jets are compared in Figure 4.9. No difference

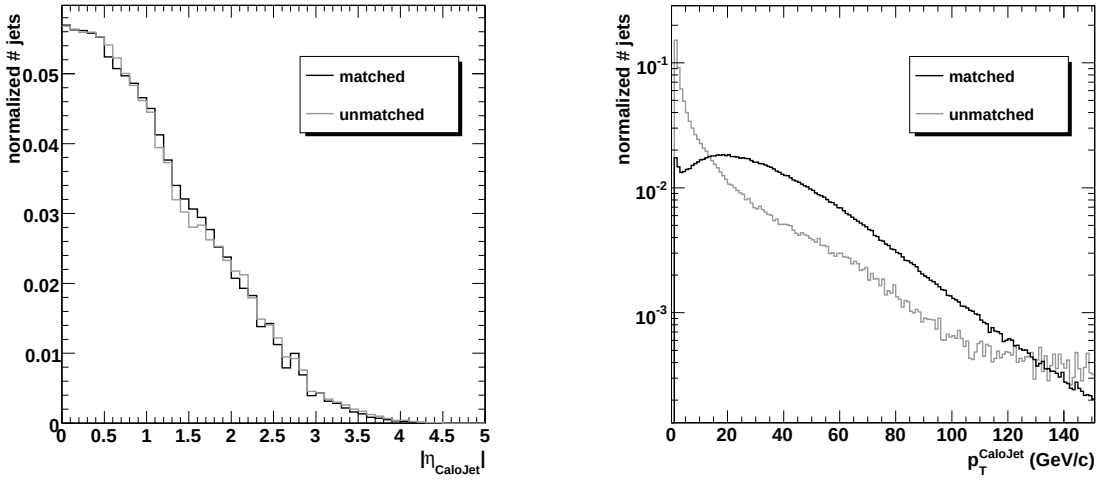


Figure 4.9: Pseudo-rapidity η (left) and transverse momentum p_T (right) of the matched and unmatched jets.

can be seen in the normalized pseudo-rapidity distributions of the matched and the unmatched CaloJets. The normalized distributions of the transverse momentum of the matched and unmatched jets show an important difference. As the unmatched jets clearly have a lower transverse momentum, these jets can be removed by applying a lower cut on the transverse momentum of the jets. The bump for matched jets with transverse momenta close to zero indicates that the matching criterion is not always perfect.

4.1.2 Angular bias of jet reconstruction

As described in the previous section the jet-finding algorithms combine the four-momenta of the input objects to reconstruct the original four-momentum of the jet. The correct reconstruction of the direction of the original parton is an important aspect for most physics studies. In Figure 4.10 the difference is shown between the polar angle of the parton and the polar angle of the CaloJet matched to this parton with $\Delta R < 0.3$, both as a function of the polar angle of the parton and the polar angle of the jet. If the direction of the jets is correctly reconstructed, this difference is expected to be zero. However, a clear deviation from zero is visible and in addition the difference is dependent on the polar angle itself. The polar angle of the jets is reconstructed further away from the beampipe than the partons from which they originate. The absolute effect reaches a maximum for polar angles corresponding to $\theta = \pi/4$ or $\theta = 3\pi/4$ and is zero for partons or jets lying in the transverse plane ($\theta = \pi/2$). The same effect is observed when looking at the difference between the pseudo-rapidity of the parton and the pseudo-rapidity of the reconstructed jet, as shown in Figure 4.11. The origin of the difference between the direction of the reconstructed jet and the direction of the parton is related to the algorithms used to combine the input objects into a jet. The jet reconstruction algorithms described in the previous section are combining the

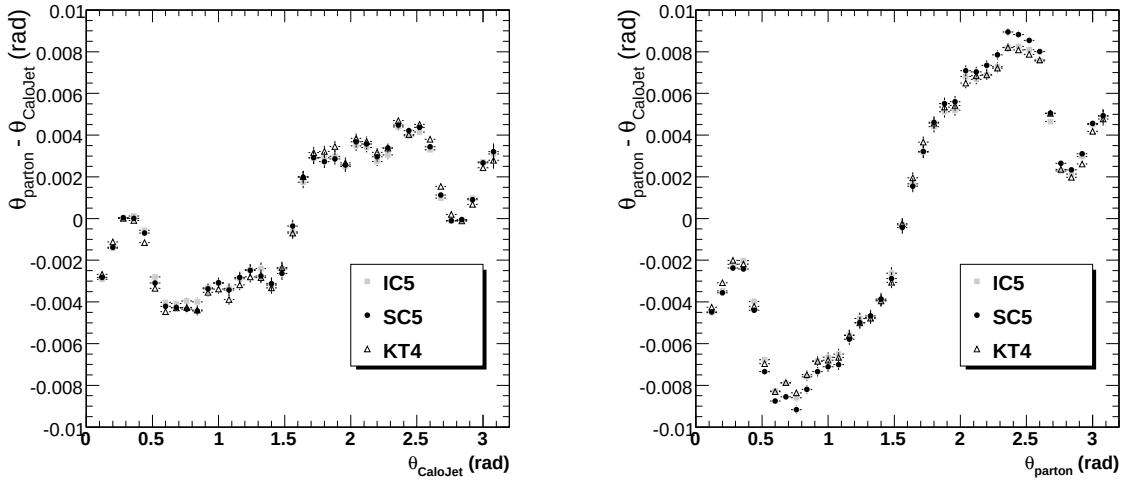


Figure 4.10: The difference between the polar angle of the parton θ_{parton} and the jet $\theta_{CaloJet}$ as a function of $\theta_{CaloJet}$ (left) and θ_{parton} (right).

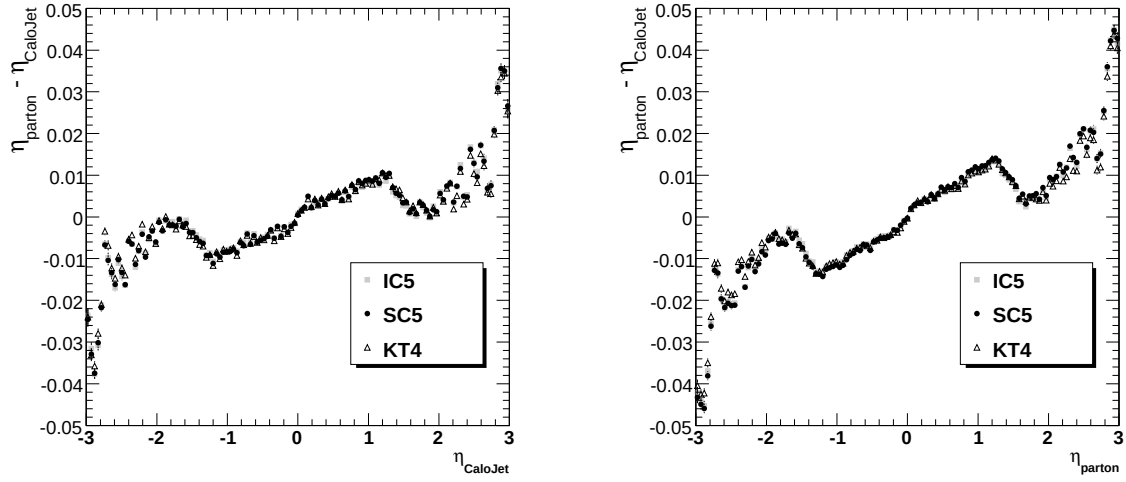


Figure 4.11: The difference between the pseudo-rapidity of the parton η_{parton} and the jet $\eta_{CaloJet}$ as a function of $\eta_{CaloJet}$ (left) and η_{parton} (right).

input objects in $(\eta \times \phi)$ space or $(y \times \phi)$ space. In the forward direction a difference $\Delta\eta$ corresponds to a smaller $\Delta\theta$ then in a less forward direction. Thus, in the case of cone-based jet reconstruction algorithms, the trial cone around the direction of an initial object has a smaller difference in θ with respect to the axis of the trial cone for the part of the cone closest to the beampipe (i.e. a more forward direction) then the difference in θ on the other side of the cone further away from the beampipe (i.e. a less forward direction). This is illustrated in Figure 4.12. While $\Delta\eta_1 = \Delta\eta_2$ holds in the

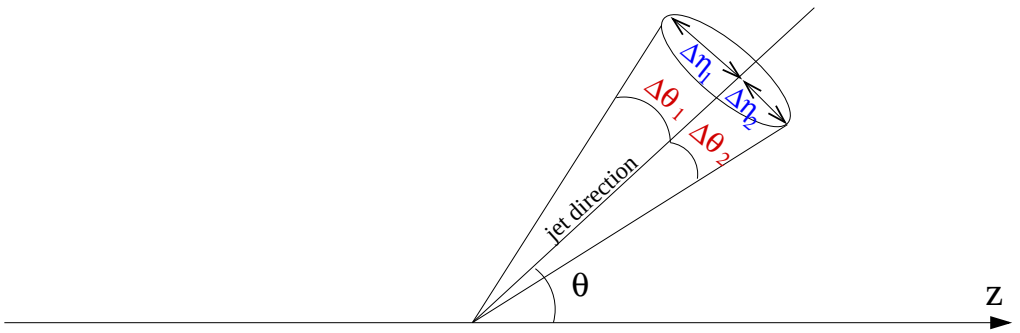


Figure 4.12: Illustration of the origin on the bias between the polar angle of the parton and the polar angle of the reconstructed jet.

pseudo-rapidity space, this becomes an inequality in the polar angle space $\Delta\theta_1 > \Delta\theta_2$. As a result, less input objects are used for the reconstruction of the jet on the "beam side" of the cone axis compared to number of input objects that are used on the other side away from the beam. Hence, when the momenta of the objects in the jet are combined, the resulting jet direction is slightly biased to the direction away from the beam. The same is true for the k_T jet clustering algorithms because the distance between any two input objects is calculated in $(y \times \phi)$ space, thus more input objects are clustered for smaller values of η compared to larger values of η.

This effect is of particular importance for high precision measurements of the mass of resonances reconstructed from jets, because the reconstructed jets can be closer to each other then the partons from which they originate, resulting in a higher reconstructed mass for the resonance. Using the information of the generated partons, a correction factor depending on the polar angle of the jet can be derived and one can correct for the effect.

Similar the difference between the azimuthal angle of the CaloJet and the azimuthal angle of the parton is shown as a function of the azimuthal angle of the parton in Figure 4.13. A small bias is visible and can be explained as originating from the magnetic field. Due to charge conservation in proton collisions, on average more positively charged particles are produced then negatively charged particles. The bias is not there when we look at the difference between the azimuthal angle of the GenJet and the azimuthal angle of the corresponding parton as shown in Figure 4.14, because GenJets are reconstructed from the particles before their interaction with the magnetic field.

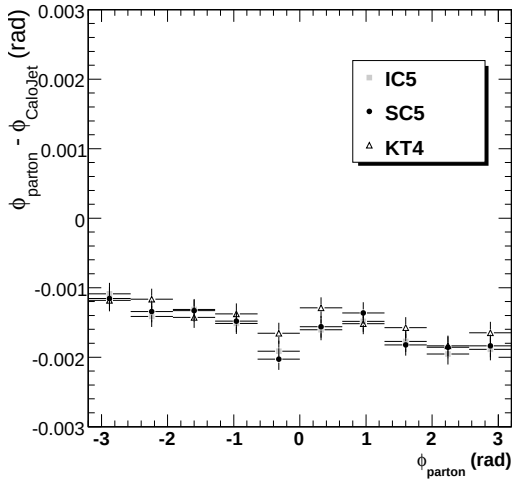


Figure 4.13: The difference between the azimuthal angle of the parton ϕ_{parton} and the CaloJet $\phi_{CaloJet}$ as a function of ϕ_{parton}.

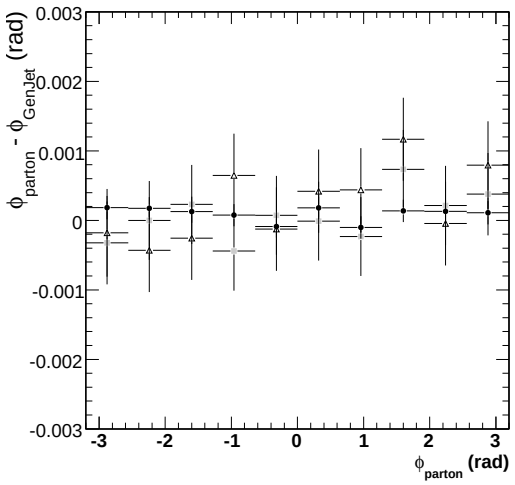


Figure 4.14: The difference between the azimuthal angle of the parton ϕ_{parton} and the GenJet ϕ_{GenJet} as a function of ϕ_{parton}.

4.1.3 Performance of SC5 jet algorithm

In this section the performance of the SC5 jet algorithm is studied because this is the jet algorithm used as benchmark example in the next chapters. The observations are however similar for the other algorithms. To check the performance of the SC5 jet algorithm, we want to know how well the partons can be matched with the GenJets and the CaloJets. Therefore, we consider the processes in which the top quarks decay semi-muonic, $t\bar{t} \rightarrow bW\bar{b}W \rightarrow bq\bar{q}\bar{b}\mu\bar{\nu}_{\mu}$, or fully hadronic, $t\bar{t} \rightarrow bW\bar{b}W \rightarrow bq\bar{q}\bar{b}\bar{q}q$. Only events for which there is no initial or final state radiation are used to avoid that more than one jet can be associated with the same parton within a specific ΔR.

The efficiency to match the partons to the GenJets as a function of either the pseudo-rapidity η or the transverse momentum p_T of the parton is shown in Figure 4.15 for different ΔR-values. A cut on the transverse momentum of the parton of $p_T \geq 20\,\mathrm{GeV/c}$ has been applied in case the efficiency is studied as a function of η, while the partons are required to be in the tracker η-acceptance if the efficiency is studied as a function of p_T. The matching efficiency of a parton to a GenJet slightly decreases as a function of the $|\eta|$, indicating that jets originating from the partons in the barrel are better reconstructed. For $\Delta R < 0.2$, the overall matching efficiency is around 82% while it increases to 99% for $\Delta R < 0.5$ reflecting the resolution on the estimation of the parton direction via the reconstructed GenJet direction. The efficiency of matching a parton to a GenJet as a function of the p_T shows that the matching efficiency is only 60 to 75% for a parton with a transverse momentum around $20\,\mathrm{GeV/c}$ when a ΔR matching of respectively 0.2 or 0.3 is applied. When $\Delta R < 0.3$ is used for the matching, the efficiency is above 90% for partons with a transverse momentum above $50\,\mathrm{GeV/c}$.

For the partons that could be matched to a GenJet, a matching of the GenJet to a CaloJet is searched to obtain the matching efficiency as a function of the η and the

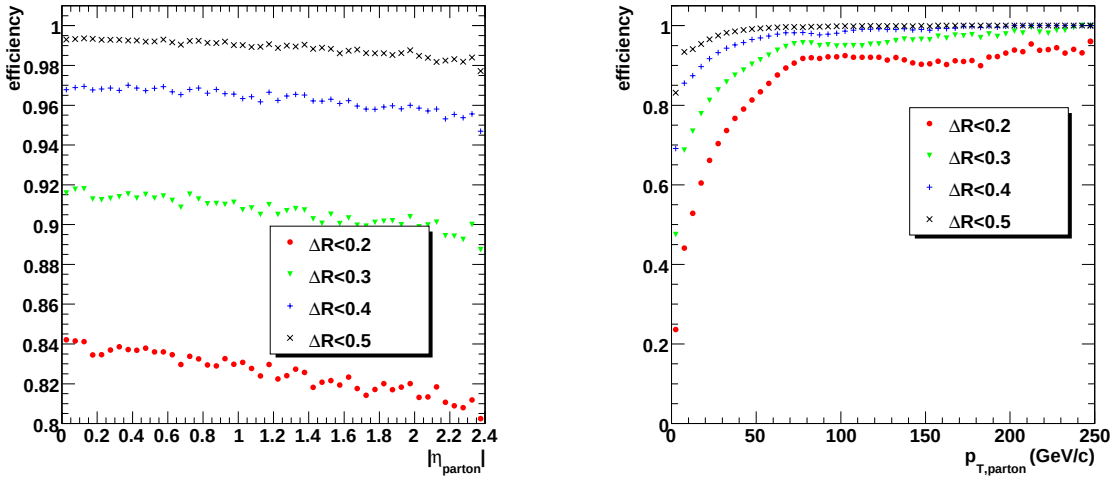


Figure 4.15: Matching efficiency of a parton to a GenJet for different ΔR criteria as a function of the η of the parton (left) and the p_T of the parton (right).

p_T of the parton as shown in Figure 4.16. When we look at the matching efficiency of a GenJet to a CaloJet as a function of the pseudo-rapidity of the parton, a dip in the efficiency is visible for $\Delta R < 0.2$ and $\Delta R < 0.3$, corresponding to the transition region of the barrel to the endcap region. Due to the magnetic field that bends particles with a low p_T outside the jet area, the matching efficiency between the GenJet and the CaloJet is lower for partons with a low transverse momentum. Also here the matching efficiency is above 90 % for partons with transverse momenta above 50 GeV/c. When studying the partons via jet reconstruction algorithms, the observation of some of the partons is lost mainly due to the fragmentation of the parton (parton→GenJet) and almost equally due to the magnetic field and detector (GenJet→CaloJet).

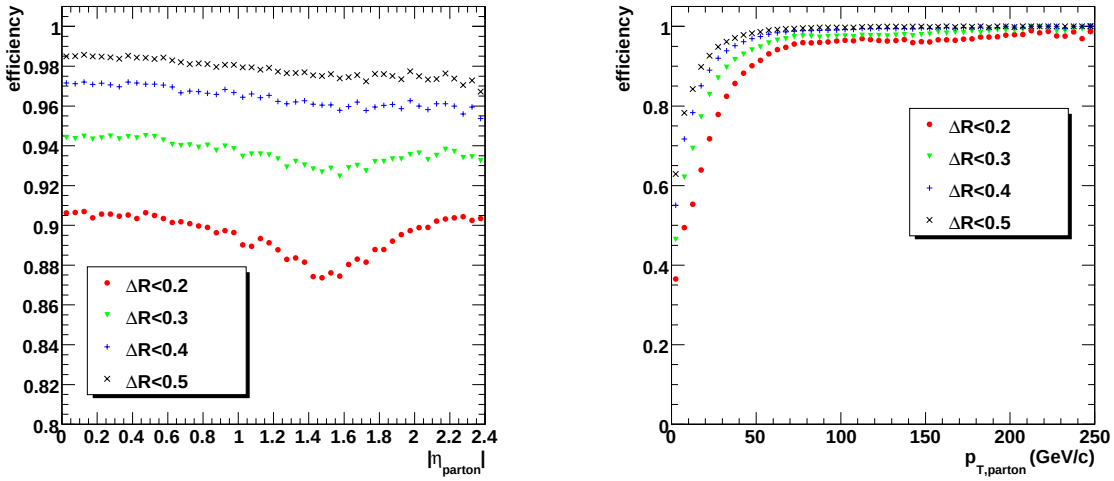


Figure 4.16: Matching efficiency of a GenJet to a CaloJet for different ΔR criteria as a function of the η of the parton (left) and the p_T of the parton (right).

4.1.4 The identification of b jets

While it is nearly impossible to tell from its experimental observation whether a jet is produced by a gluon or a u, d or s quark, the jets coming from a c or b quark can be identified with some precision, especially in the latter case. Due to the large mass of the b quark, B hadrons are produced during the fragmentation. B hadrons have a relatively large mass (for instance for the B^0 meson, $m_{B^0} = 5279.4 \pm 0.5\,\mathrm{MeV}/c^2$), a long lifetime ($\tau_B^{mix}$ =1.568 $\pm$ 0.009 ps [3]), corresponding to a distance maximum distance of about 470 μm when the B hadron would be able to travel at the speed of light, and have a branching ratio of 35.6 $\pm$ 0.5 % to decay into an electron or a muon, including the decays in which a muon or electron originate from a tau lepton or c quark produced during the b quark fragmentation [114]. The properties of the B hadrons provide jets originating from b quarks a distinct signature compared to jets initiated by lighter quarks. Via hypothesis testing techniques this signature can be exploited to assign the flavour of the b quark to a reconstructed jet, a method that is called b-tagging. Up to ten b-tagging discriminators are developed for the CMS experiment and an overview of their performance can be found in [115]. Some of the b-tagging discriminators exploit the presence of a soft lepton in a b-jet, others use information of displaced vertices in the jet. The b-tagging discriminator used in the analysis in the next chapters is based on the tracks associated to the jet.

Vertex reconstruction

Both the reconstruction of the interaction vertex as well as the displaced secondary vertices from the decay of heavy objects produced in the final state play an important role in the reconstruction of each proton collision. The reconstruction of vertices consists of two main steps, vertex finding and vertex fitting.

"Primary" vertex finding is the task of grouping tracks into initial vertex candidates. Tracks, reconstructed as discussed in Section 2.2.2 are used as input for the vertex finding if they pass certain quality cuts. Each track should have a normalized χ^2 of the track fit below 5 and at least 2 hits in the pixel tracker and at least 7 hits in the silicon tracker. In addition, the 2D transverse impact parameter significance ($d_0/\sigma(d_0)$) with respect to the beam spot position should be smaller than 5. These requirements reject fake tracks and tracks originating from secondary vertices. Tracks are then sorted in increasing order of the point of the closest approach to the z axis and grouped together in clusters if the distance between two successive tracks is less then 1 mm. These vertex candidates are then given as input to the vertex fitter if they contain at least two tracks.

The procedure in which (a subset of) the reconstructed tracks are used as an input to calculate the positional parameters of a vertex as well as the covariance matrix of these parameters is known as vertex fitting. The most commonly known method for vertex fitting is the Kalman filter [38]. The Kalman filter is a least-squares estimator that minimizes the sum of the squared standardized distances of all tracks from the vertex position. The Kalman filter is known not to be robust for outlying tracks such as tracks that are wrongly associated to belong to the same vertex or tracks that have been mis-reconstructed. Therefore, several robust vertex algorithms have been developed for the CMS experiment [116], with the Adaptive Vertex Fitter as a

powerful general purpose algorithm. The Adaptive Vertex Fitter [117–119] provides a robust method by downweighting outlying tracks instead of rejecting them. The weight represents the compatibility of the track with the vertex and is recalculated in each iteration until the fit converges.

After the vertex fitting, a cleaning is performed by calculating the transverse distance between the fitted vertices and the beam, using the beamspot position. A primary vertex is required to be closer than $500\,\mu m$ to the beam, otherwise it is rejected. The vertex of the hard interaction is chosen as the one with the highest sum of the squared transverse momenta of the tracks connected to the vertex. The efficiency of finding the correct primary vertex is close to $100\,\%$.

Vertices originating from the decay of heavy objects inside jets are referred to as secondary vertices. In the CMS experiment, also secondary vertices are fitted using the Adaptive Vertex Fitter. The approach is slightly different compared to the case of primary vertex fitting. No initial vertex finding is performed, vertices are instead fitted using all tracks that are given to the algorithm in an iterative procedure. Tracks which are not compatible with a vertex fitted in previous iterations are used for the fitting of remaining vertices until no tracks remain. Quality cuts can be applied on secondary vertex candidates to assign a secondary vertex to a jet.

The identification of b jets using track impact parameter significances

To test the hypothesis if a reconstructed jet is originating from a b quark, a test variable or discriminator must be calculated reflecting the relevant properties of the jet. Two observables that are often used to construct a b-tagging discriminator are the 2D and 3D signed impact parameter significances. As illustrated in Figure 4.17 the impact parameter (IP) is defined as the distance between the primary vertex and the linearised track in the point of minimal distance between the track and the jet axis. The impact parameters of all tracks closer than $\Delta R = 0.5$ to the jet axis are calculated.

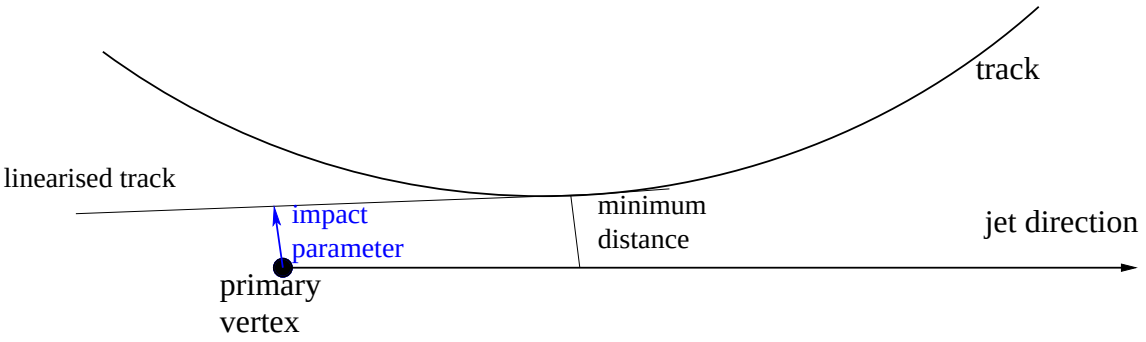


Figure 4.17: Geometrical interpretation of the impact parameter between the track and the primary vertex.

As the uncertainty on the impact parameter can be of the same order of magnitude as the impact parameter itself, a better observable is obtained by dividing the impact

parameter by its uncertainty σ_{IP}. This observable is called the impact parameter significance IP_S:

$$IP_S = \frac{IP}{\sigma_{IP}} \tag{4.2}$$

The impact parameter and its uncertainty can be calculated in the plane transverse to the beam, in which case the observable is called the 2D impact parameter significance IP_S^{2D}, or in three dimensions and referred to as the 3D impact parameter significance IP_S^{3D}. According to the "lifetime signing" convention, a positive sign is given to the impact parameter if the linearised track crosses the jet axis before the primary vertex with respect to the jet direction, a negative sign if it crosses the jet axis after the primary vertex.

To construct the b-tagging discriminators that make use of the 2D and 3D signed impact parameter significances, the tracks are ordered in decreasing 3D significance. The "track counting high efficiency" and the "track counting high purity" discriminators [115] are defined as the signed impact parameter significance of the second and the third track respectively. A cut on the discriminator is often made to increase the number of jets originating from the fragmentation of a b quark with respect to the number of selected jets.

Another possible discriminating method consists in considering the signed impact parameter significances of all the tracks associated with a jet. The aim is to combine the information from all the observables into one discriminator by means of a likelihood or a neural network. This is done for the construction of the "impact parameter MVA" discriminator [120]. Eight observables are fed into an artificial neural network, in this case a multilayer perceptron (MLP) [121]. Six of the eight observables are the 2D and 3D signed impact parameter significances of the first three tracks ordered in decreasing 3D significance, while the other two observables contain the combined information of respectively the 2D or the 3D signed impact parameter significances of the remaining tracks. The 2D signed impact parameter significances of the remaining tracks are combined into one observable by a likelihood ratio as follows:

$$IP_{S,LR}^{2D} = \frac{\prod_{i=4}^{n} IP_{S,i}^{2D,b}}{\prod_{i=4}^{n} IP_{S,i}^{2D,b} + \prod_{i=4}^{n} IP_{S,i}^{2D,non-b}} \tag{4.3}$$

where n is the number of tracks associated to the jet, and $IP_{S,i}^{2D,b}$ and $IP_{S,i}^{2D,non-b}$ are the probability distribution functions of the 2D signed impact parameter significance of track i of a jet obtained from the simulated b quark and non-b quark jets respectively. Similarly the 3D signed impact parameter significances of the remaining tracks are combined into one observable. The neural network combines the eight variables into a single discriminator, shown in Figure 4.18 for true b jets and non-b jets in $t\bar{t}$ separately. There is a clear separation between the two distributions, illustrating the discriminating power of the b-tagging algorithm.

In order to compare the performance of b-tagging discriminators, one needs to determine which parton flavour is assigned to a jet. Therefore, all the partons that are closer then $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.3$ to the jet-axis are listed. If one parton is found, the jet is assigned the flavour of this parton. If the list contains more then one parton, there are two possibilities to assign a parton flavour to the jet, a "physics" definition

and an "algorithmic" definition. Using the physics definition a jet obtains the flavour of the initial parton, i.e. before radiation, while for the algorithmic definition, the flavour of the heaviest parton in the list is assigned to the jet. The physics definition is interesting if one wants to study the partons of the hard process, but the algorithmic definition avoids for instance the situation that, as a consequence of gluon splitting to $b\bar{b}$, a gluon jet contains B hadrons. The algorithmic definition is thus used when the performance of b-tagging algorithms is studied. By applying cuts on the value of the discriminator an efficiency to select true b quark jets and an efficiency to select non-b quark jets is obtained. The first number is usually noted as the b-tag efficiency, while the second reflects the mistag efficiency. These numbers reflect the performance of the b-tagging algorithm. The b versus non-b jet efficiency for the "impact parameter MVA", "track counting high efficiency" and "track counting high purity" algorithms is shown in Figure 4.19. The performance of the "impact parameter MVA" discriminator

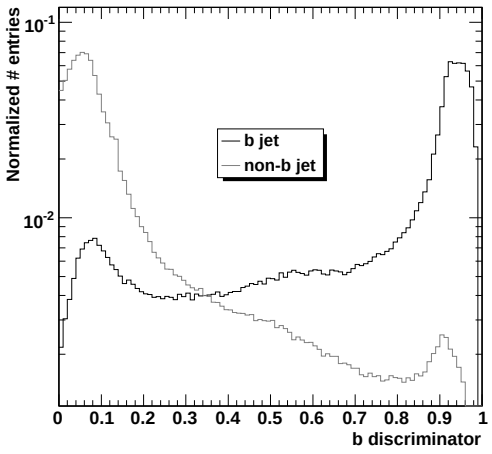


Figure 4.18: The "impact parameter MVA" discriminator constructed from the 2D and 3D impact parameter significances using all tracks associated to the jet. The distribution for true b jets is peaked at one, while the distribution for non-b jets is peaked at zero.

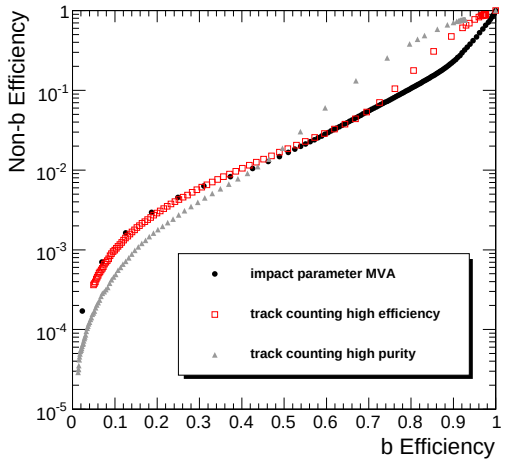


Figure 4.19: The b versus non-b efficiency as an illustration of the performance of the "impact parameter MVA" b-tagging discriminator. Also the b versus non-b efficiency for the "track counting high efficiency" and "track counting high purity" discriminators are shown.

is similar to the performance of the "track counting high efficiency" discriminator for b-tag efficiencies up to 0.7. For higher b-tag efficiencies the "impact parameter MVA" algorithm performs better, because a lower non-b efficiency is obtained then for the two track counting algorithms for an equal b jet efficiency.

4.1.5 Jet energy corrections

Up to here the momenta of the jets were taken as the raw reconstructed energy. The interaction of the particles produced in the hard scattering with the material of the detector and the bending of the charged particles in the magnetic field will result in

reconstructed jet energies that can be very different from the original parton energies. Therefore, jet energy corrections need to be derived that are able to relate the reconstructed jet energy with the energy of the primary parton or GenJet. In the CMS collaboration a factorized approach has been chosen to correct the jet energy, such that the total correction consists of subsequent factors correcting for different detector and physics effects which should be applied in a fixed order [122, 123]. The different levels of corrections that can be applied on the reconstructed jet energy in order to obtain a calibrated jet are listed below. The methods to obtain these corrections either from simulated proton collisions or using data driven methods [1] are described afterwards.

- **Offset correction** (L1)
  The goal of the first correction is to correct for the effect of electronic noise and pile-up of multiple proton collisions. After this correction the jet energy should be independent of the luminosity.

- **Relative η dependent correction** (L2)
  Due to detector effects the jet response varies with the pseudo-rapidity, therefore a jet energy correction relative to a control region needs to be applied.

- **Absolute p_T dependent correction** (L3)
  The primary goal of the absolute correction is to correct for variations in the jet response with the transverse momentum of the jet.

- **Correction for dependency on electromagnetic fraction** (L4)
  A different jet energy response can be expected depending on the electromagnetic fraction of the jet. The L4 correction aims to correct for that variation.

- **Correction for dependency on the jet flavour** (L5)
  The particle content for the various jet flavours is different, thus a specific correction is needed to correct with respect to the flavour of the jet.

- **Correction for the underlying event** (L6)
  This is a correction for the underlying event energy that contributes to the jet energy.

- **Correction to the parton level** (L7)
  The last correction aims to correct the jet energy to the original parton energy, in order to correct for the energy loss due to radiation.

The first three levels of corrections are required for any physics analysis, while the last four are optional corrections. The corrections are derived in such a way that the CaloJet energy up to each correction level can be related to the uncorrected CaloJet energy. For instance when the required corrections are applied, the new jet energy $E_{CaloJet,L3}$ is related to the uncorrected or raw jet energy $E_{CaloJet,RAW}$ as follows

$$E_{CaloJet,L3} = (E_{CaloJet,RAW} - C(L1)) \times C(L2) \times C(L3) \tag{4.4}$$

[1] The terminology "data driven" refers to methods that make no use of the information from the simulation.

and in case the optional corrections are also included this becomes

$$E_{CaloJet,L7} = E_{CaloJet,L3} \times C(L4) \times C(L5) \times C(L6) \times C(L7) \tag{4.5}$$

where $C(LX)$ is the correction factor for level X, with X between 1 and 7.

L1: offset correction

The offset correction aims to subtract from the jet energy, the energy from multiple proton collisions (in-time plus out-of-time pile-up) deposited in the calorimeter towers of the jet and the energy from the calorimeter towers for which the electronic noise passes the Scheme B thresholds described in Section 4.1.1. The offset energy is determined by looking at the energy deposited in a random cone with radius $\Delta R = 0.5$ in events collected by a random trigger with the single requirement that there has been a beam crossing (zero bias events). It is important to determine the offset energy as a function of the pseudo-rapidity as the contribution of pile-up to the offset energy increases in the forward directions. The contribution to the offset energy from electronic noise has been found to be neglible, while the contribution to the transverse energy offset from a single minimum-bias event (event that requires only an interaction to trigger) is below 0.3 GeV per jet [122]. The offset energy is expected to scale linearly with the average number of minimum-bias events.

Recently a study has been performed [124] using different average number of minimum-bias events. The average energy offset due to full pile-up with an average of 1(5) minimum-bias events is below 1(7.5) GeV per jet for $|\eta| < 2$ and below 2(17) GeV per jet for $2 < |\eta| < 3$. Because of the strong dependence of the magnitude of the offset on the values of the calorimeter deposition thresholds and the loss of real energy from jets, the calorimeter cell and tower thresholds have been re-optimized recently and are referred to as Scheme 6 thresholds [125].

In the simulated samples on which the analysis has been based, the old thresholds are used. Furthermore, no offset correction for these samples has been applied because the offset energy for interactions without pile-up is very small as was discussed above.

L2: relative η dependent correction

A data driven method has been developed within the CMS experiment to derive the η dependent corrections from QCD di-jet events based on the principle of transverse momentum conservation [126]. For events with a final state consisting of two partons, the two jets originating from these partons should have equal transverse momenta. However, in the transition regions of the pseudo-rapidity between the different calorimeter subsystems, the jet energy response is too low. For higher pseudo-rapidities, jets with fixed transverse momentum contain more energy compared to lower pseudo-rapidities, due to $E = p_T cosh\eta$. The jet energy response with respect to the pseudo-rapidity is flattened by considering events where one jet is observed in the barrel (control) region ($|\eta| < 1.3$) and the other jet (probe jet) can have an arbitrary value of the pseudo-rapidity η^{probe}. The jets are chosen to be back-to-back by requiring that the difference in azimuthal angle between the two jets exceeds 2.5 rad ($\Delta\phi > 2.5$ rad). A

possible third jet in the events should have $p_T < 0.25\, p_T^{di-jet}$, where the average transverse momentum of the di-jet system is defined as $p_T^{di-jet} = (p_T^{probe} + p_T^{barrel})/2$. The observable $B = (p_T^{probe} - p_T^{barrel})/p_T^{di-jet}$ is then studied as a function of the pseudo-rapidity of the probe jet η^{probe} and in bins of the p_T^{di-jet} variable. The relative response $R(\eta^{probe}, p_T^{di-jet}) = (2+B)/(2-B)$ is obtained from the most probable value of B (i.e. expectation value of a Gaussian fit) in each η^{probe} and p_T^{di-jet} bin. This response is then inverted and used as a correction for the average p_T^{probe} in the corresponding η^{probe} and p_T^{di-jet} bin.

Corrections based on the generated particles from simulated samples will be used for early data taking and have been used for the analysis. To obtain these corrections CaloJets were matched with GenJets for QCD di-jet events by requiring that they are closer in $(\eta \times \phi)$ space then $\Delta R = 0.25$. For different p_T^{GenJet} regions of the matched jets, the most probable value of a Gaussian fit on the distribution $\Delta p_T = p_T^{CaloJet} - p_T^{GenJet}$ in a given η bin is used to calculate the jet response

$$R(\eta, p_T^{GenJet}) = 1 + \Delta p_T / p_T^{GenJet}. \tag{4.6}$$

where p_T^{GenJet} of the bin is taken as the average transverse momentum in the bin. The relative jet response is obtained by dividing the jet response in a certain η bin by the jet response in the barrel region ($|\eta| < 1.3$). The relative correction is found by inverting the relative jet response. In Figure 4.20 the jet energy response in $t\bar{t}$ events as a function of the pseudo-rapidity of the corresponding GenJet ($\Delta R < 0.25$) is shown for two different p_T bins of the GenJet before and after applying the L2 correction. The observed structure before the L2 correction is applied originates from the detector geometry.

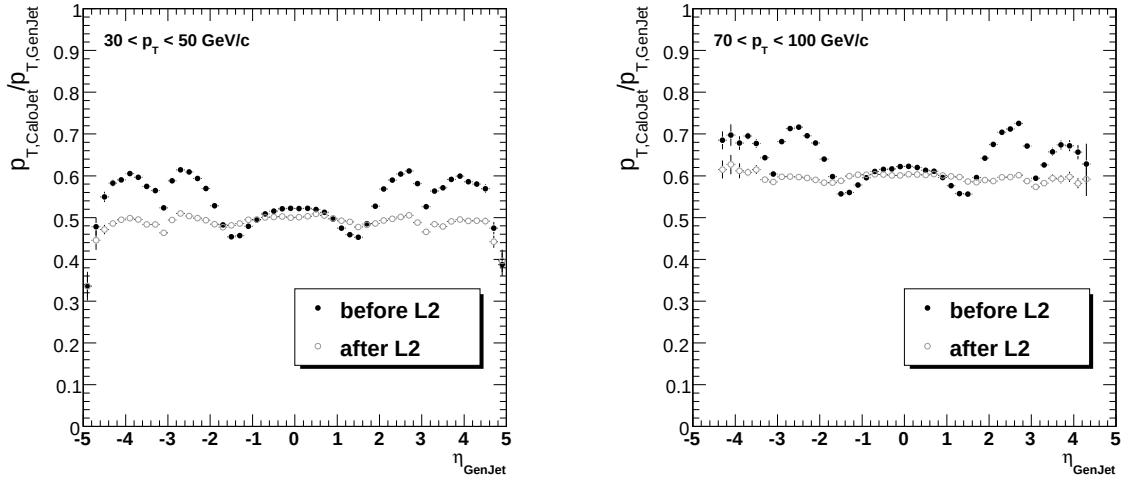


Figure 4.20: Jet energy response as a function of the pseudo-rapidity of the GenJet before and after L2 correction for two different p_T-bins of the GenJet.

L3: absolute p_T dependent correction

Due to the non-linear response of the calorimeter, the jet response needs to be corrected as a function of the p_T of the jet. Data driven methods have been developed based

on either γ+jet [127] or Z+jet events [128, 129]. In the case the electromagnetic calorimeter has been calibrated (e.g. using the mass of the π^0 as a constraint in $\pi^0 \to \gamma\gamma$) the jet energy corrections can be derived from γ+jet or $Z(\to e^+e^-)$+jet events from the p_T balance in the observed final state of the event. The transverse momentum of the parton should be equal to the transverse momentum of the photon or Z boson, $p_T^{parton} = p_T^{\gamma/Z}$. To reduce the background in the γ+jet events, the photon is required to be isolated. The photon or Z boson are required to be back-to-back with the jet by asking $|\Delta\phi - \pi| < 0.2\,\text{rad}$. In addition a possible second jet in the event should have a transverse momentum that is small compared to the transverse momentum of the photon or Z boson ($p_T^{jet2} < 0.2\ p_T^Z$ for Z+jet events and $p_T^{jet2} < 0.1\ p_T^\gamma$ for γ+jet events). The response $p_T^{jet}/p_T^{\gamma/Z}$ is inverted and mapped from a function of $p_T^{\gamma/Z}$ to a function of $p_T^{CaloJet}$. Similarly $Z(\to \mu^+\mu^-)$+jet events can be used, which has the advantage that the excellent muon system of the CMS experiment can be exploited.

Currently the corrections have been derived from simulated QCD di-jet events. GenJets are matched to CaloJets if they are closer then $\Delta R = 0.25$ in $(\eta \times \phi)$ space. Both jets are required to be in the barrel region of the calorimeter ($|\eta_{jet}| < 1.3$). The most probable value Δp_T obtained from the expectation value of a Gaussian fit on the distribution $\Delta p_T = p_T^{CaloJet} - p_T^{GenJet}$ in the different p_T^{GenJet} bins is used for the calculation of the jet response, $R(p_T^{GenJet}) = 1 + \Delta p_T / p_T^{GenJet}$. The jet energy corrections are then obtained by inverting the response. In Figure 4.21 the jet response is shown for $t\bar{t}$ events as a function of the p_T of the GenJet for two different bins of the pseudo-rapidity of the jet. After the L3 jet energy correction is applied on top of the previous L2 correction, the jet response is almost flat around unity except for small p_T. The dependency that is still visible is due to the different flavour mixture in $t\bar{t}$ events compared to the QCD di-jet events from which the corrections were derived.

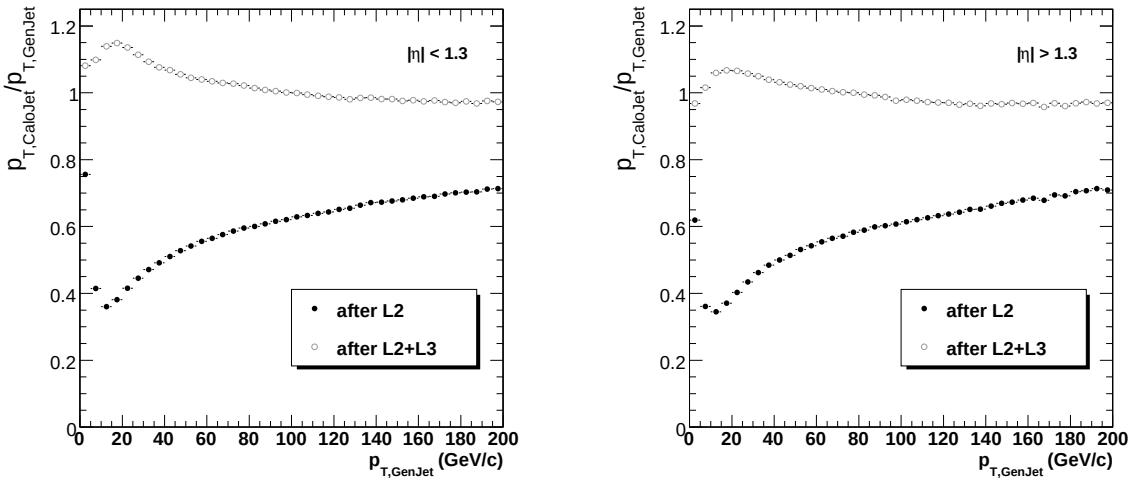


Figure 4.21: Jet energy response in $t\bar{t}$ events as a function of the transverse momentum of the GenJet before and after L3 correction for two different η-bins.

L4: correction for dependency on the electromagnetic fraction

The jet energy resolution can be improved by taking into account the dependency of the jet response on the fraction of jet energy deposited in the electromagnetic calorimeter [122]. The response of a jet shows significant variations for both low and high electromagnetic fractions. This optional correction can be derived either from simulated collision events or by measuring the jet response as a function of the electromagnetic fraction of the jet using for example γ+jet events.

L5: correction for dependency on the jet flavour

The jet response depends on the flavour of the parton from which the jet originates due to the different particle content in the jet. For instance light (u,d,s) quarks fragment in particles with a higher momentum than gluons and jets originating from a b or c quark have a lower response because they can contain neutrinos in the decay chain of heavy B or C hadrons. In addition, the flavour composition of the samples used to determine the L2 and L3 jet energy corrections can be very different from the flavour composition of the samples on which they are applied, as is shown for instance in Figure 4.21. For instance in QCD di-jet events the flavour mixture is dominated by gluons, leading to residual jet energy corrections that are different from unity in $\mathrm{t\bar{t}}$ events. Flavour corrections are usually developed from simulated samples using the parton information in the Monte Carlo. They are determined as a function of the p_T of the GenJet and inverted to be applied as a function of the p_T of the reconstructed CaloJet. Data-driven methods can be developed to obtain flavour dependent jet energy corrections from $\gamma + b$, $b\bar{b}Z$, $Z \rightarrow b\bar{b}$ or $\mathrm{t\bar{t}}$ events [123].

In Figure 4.22 the jet energy response is shown for the different quark flavours in $\mathrm{t\bar{t}}$ events as a function of the transverse momentum of the GenJet before and after correcting for the jet flavour, but after applying the previous L2 and L3 corrections. The jet flavour was determined by requiring jets that are matched both with a GenJet

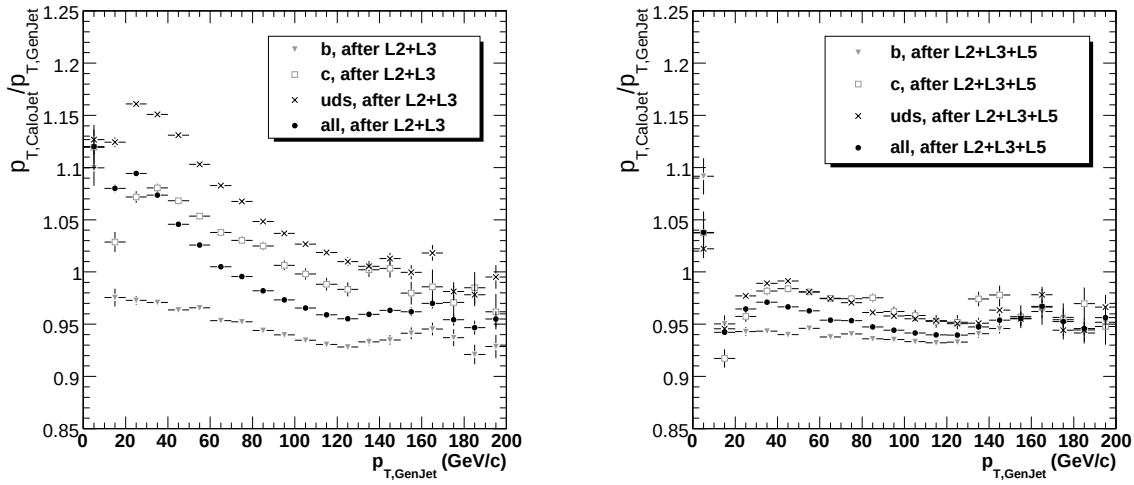


Figure 4.22: Jet energy response as a function of the transverse momentum of the GenJet for different flavours before and after L5 correction.

in $(\eta \times \phi)$ space by requiring $\Delta R < 0.25$ and with a parton by requiring $\Delta R < 0.3$. The flavour of the parton is then assigned to the jet.

After applying the L5 corrections, the response still differs for the different flavours. Biases can occur because di-jet events have been used to derive the jet corrections, while the corrections are applied on $t\bar{t}$ events. In this case not only the flavour composition of di-jet and $t\bar{t}$ events is different but top quark pairs lead to a denser environment. Additionally, (the modelling of) the color flow in $t\bar{t}$ and other processes can be different. As mentioned in Section 4.1.4, different algorithms exist to assign the flavour of a parton to a jet, which can also lead to biases as there is no unambiguous method to assign parton flavours to jets.

L6: correction for the underlying event

As the underlying event is dependent on the hard scattering in the proton collision, the energy contribution of the underlying event to the total energy of the jets in the event will be different for various physics processes. Therefore, the correction for the underlying event energy is treated as an optional correction. The correction for the underlying event is often part of the correction to the parton level. A possibility to obtain a correction for the underlying event is to study for each physics process various Monte Carlo event generators which use different models for the underlying event. For early data taking, the energy from the underlying event contributing to the jet energy will be estimated as the energy in a random jet area from minimum bias events at low luminosity after subtracting the offset energy. For the simulated samples used in this thesis, no separate corrections for the underlying event were determined or applied.

L7: correction to the parton level

While the previous corrections correct the jet energy back to the energy of the corresponding GenJet, the parton correction is intended to correct the jets back to the parton level. In the factorized approach, the parton correction should therefore correct the energy of the GenJet back to the energy of the parton. The $p_T^{GenJet}/p_T^{parton}$ response is calculated as a function of p_T^{parton} to correct for effects originating from the parton showering or hadronization process [130]. The parton correction will be different depending on the flavour since for instance for gluon jets more final state radiation will fall outside the reconstructed GenJet area compared to quark jets. The quark energies in simulated proton collisions are used to determine the correction to the parton level and it should be kept in mind that the correction is strongly model and process dependent. When the Monte Carlo generators are tuned to the data, these corrections can be more precisely obtained. With this strategy also the contribution of the underlying event is part of the parton correction.

Figure 4.23 shows the GenJet to parton response for the different parton flavours in $t\bar{t}$ events. In Figure 4.24 the parton corrections have been applied on top of the L2, L3 and L5 corrections, showing a consistency with unity within 3 % when looking at the different quark flavours together. A difference from unity is still observed, because as in the case of the L5 correction, also the L7 corrections were obtained from di-jet events.

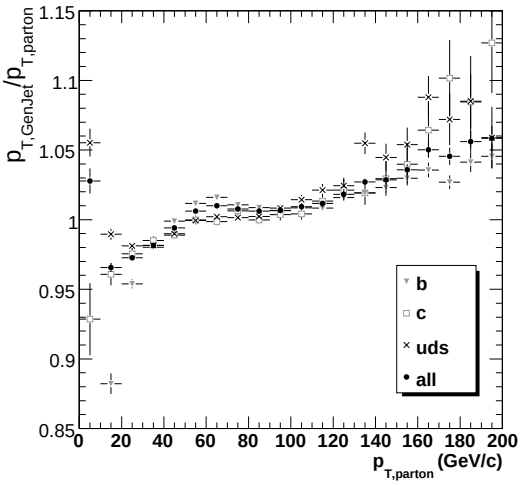


Figure 4.23: GenJet to parton response ($p_T^{GenJet}/p_T^{parton}$) as a function of the transverse momentum of the parton.

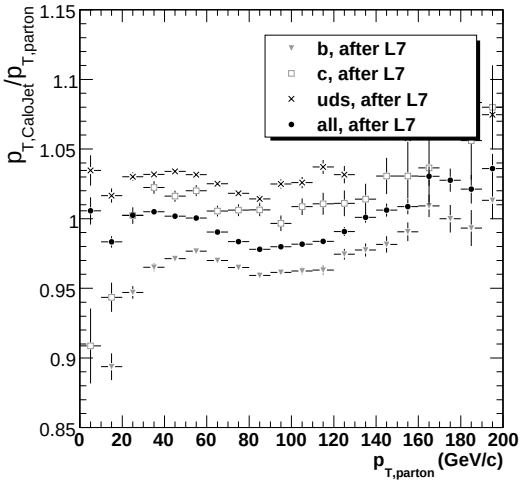


Figure 4.24: Jet energy response as a function of the transverse momentum of the parton for different flavours after L7 correction.

In the following sections and chapters, only the required L2 and L3 jet energy corrections have been applied. These required jet energy corrections have been derived from simulated QCD di-jet samples as described above.

4.1.6 Resolution on the jet direction and transverse energy

To determine the uncertainty on the direction and the transverse energy of the SC5 jets, so-called jet resolutions are calculated. After the application of the required L2 and L3 corrections, the jets in top quark pair events are matched to the partons from which they originate using a ΔR matching requirement of $\Delta R < 0.3$. The resolution on a variable of interest is then calculated as the width of a Gaussian fit on the distribution of the difference between the jet variable and the variable of the matched parton. The resolutions are determined in bins of the transverse momentum p_T and pseudo-rapidity η of the jet. A differentiation according to the jet flavour has been performed, by considering b jets as the jets matched to a b quark and non-b jets as the jets matched to any other quark. In Figure 4.25 the resolutions on the transverse energy, the polar angle and the azimuthal angle are shown as a function of the transverse momentum for b and non-b jets in different bins of the pseudo-rapidity. Ten bins have been considered for the pseudo-rapidity as defined in the legend. To guide the eye over the different pseudora-pidity bins in the figures, the resolution curves with high and low absolute values for the pseudo-rapidity have been indicated. From the upper plots, showing the absolute resolution on the transverse energy, we see that resolution is worse for jets with high transverse momenta, but the opposite is true if we consider relative resolutions. This is related to the magnetic field, which strongly bends particles with low transverse momenta with a result that these particles are not clustered by the jet reconstruction algorithms into the same jet. If we consider the different pseudo-rapidity

bins, we see that jets in the forward direction have a better resolution compared to jets lying in the transverse plane ($|\eta| \simeq 0$). This dependency on the pseudo-rapidity can be explained by the fact that in a certain bin of the transverse energy or transverse momentum, relatively higher energies are needed for jets with high pseudo-rapidities then for jets with lower pseudo-rapidities, hence the jets have a larger boost for high pseudo-rapidity bins. If the transverse energy resolutions for b and non-b jets are compared, we see that b jets tend to have a worse resolution. The origin of this effect lies in the possible presence of muons and neutrinos in the b jets, leading to reduced reconstructed energies and non-Gaussian effects in the fitted distributions. The plots in the middle and on the bottom of Figure 4.25 show the resolutions on the reconstructed polar and the azimuthal angle. A first observation for both the resolutions on θ and ϕ is the improvement of the resolution for higher transverse momenta. What is also visible from the plots is a better resolution for both angles with increasing pseudo-rapidities. This is clearly more pronounced for θ then for ϕ and is related to the requirement of the matching criterion. The ΔR variable is reconstructed in $(\eta \times \phi)$ space, resulting in a more stringent criterion for θ in the forward direction, as reflected by the better resolutions for high pseudo-rapidities. The resolution on θ is better than the resolution on ϕ because particles are bent by the magnetic field in the transverse plane, which leads to worse resolution on the reconstructed azimuthal angle ϕ. If the resolutions for b and non-b jets, are compared, no real differences are observed.

In Figures 4.26 and 4.27, the effect of the matching requirement is studied for b and non-b jets respectively. For two extreme bins of the pseudo-rapidity the matching criterion is altered between a ΔR of 0.1 and 0.4 in steps of 0.1. For high transverse momenta the resolution does not change, while effects are clearly visible for low transverse momenta, where the (relative) resolutions are worse due to the effects discussed above. The resolution on all the variables improves while tightening the matching requirement. Between $\Delta R = 0.1$ and $\Delta R = 0.2$ a bigger discrepancy is observed, indicating that we start selecting a subset of very well reconstructed jets and thus improve the resolution artificially. The observed effect is similar in the low and high pseudo-rapidity range.

4.2 Muon reconstruction

Due to their mass, which is about 200 times higher than the mass of electrons, muons lose less energy by bremsstrahlung than electrons. Therefore, they penetrate matter deeper compared to any other Standard Model particle, with the exception of neutrinos. Muons are easier identified and reconstructed compared to other particles because they are the only particles interacting with all subdetectors, including the muon system surrounding all other subdetectors.

The reconstruction of muons using only the information from the muon subdetector systems is discussed in Section 4.2.1, while the reconstruction of muons using the additional information from the tracker is specified in Section 4.2.2. In Section 4.2.3 several variables are introduced which are defined to identify muons. The resolutions on the reconstructed muon transverse momentum and direction are shown in Section 4.2.4.

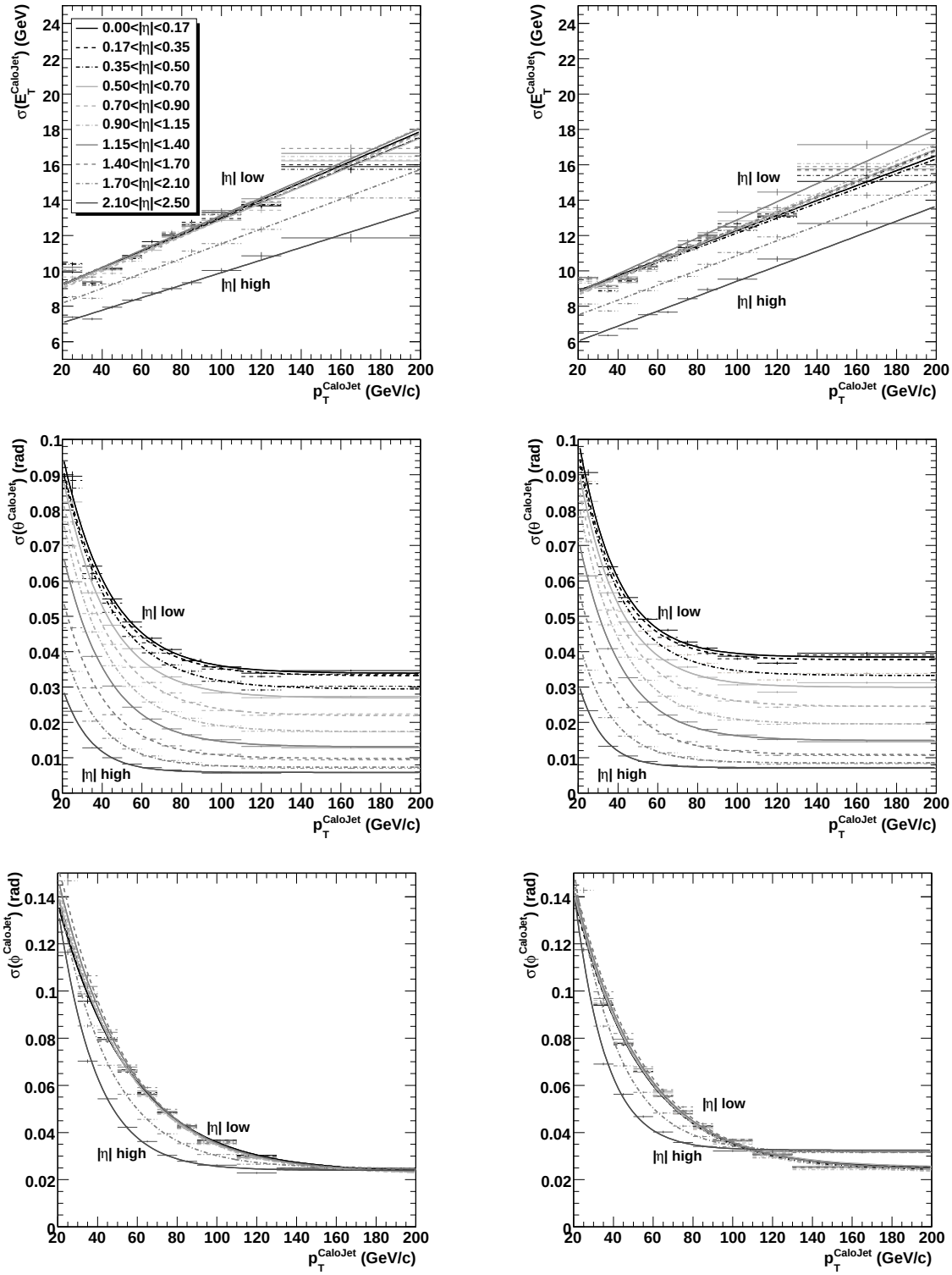


Figure 4.25: Absolute resolution on the transverse energy (upper), polar angle (middle) an azimuthal angle (lower) for b jets (left) and non-b jets (right).

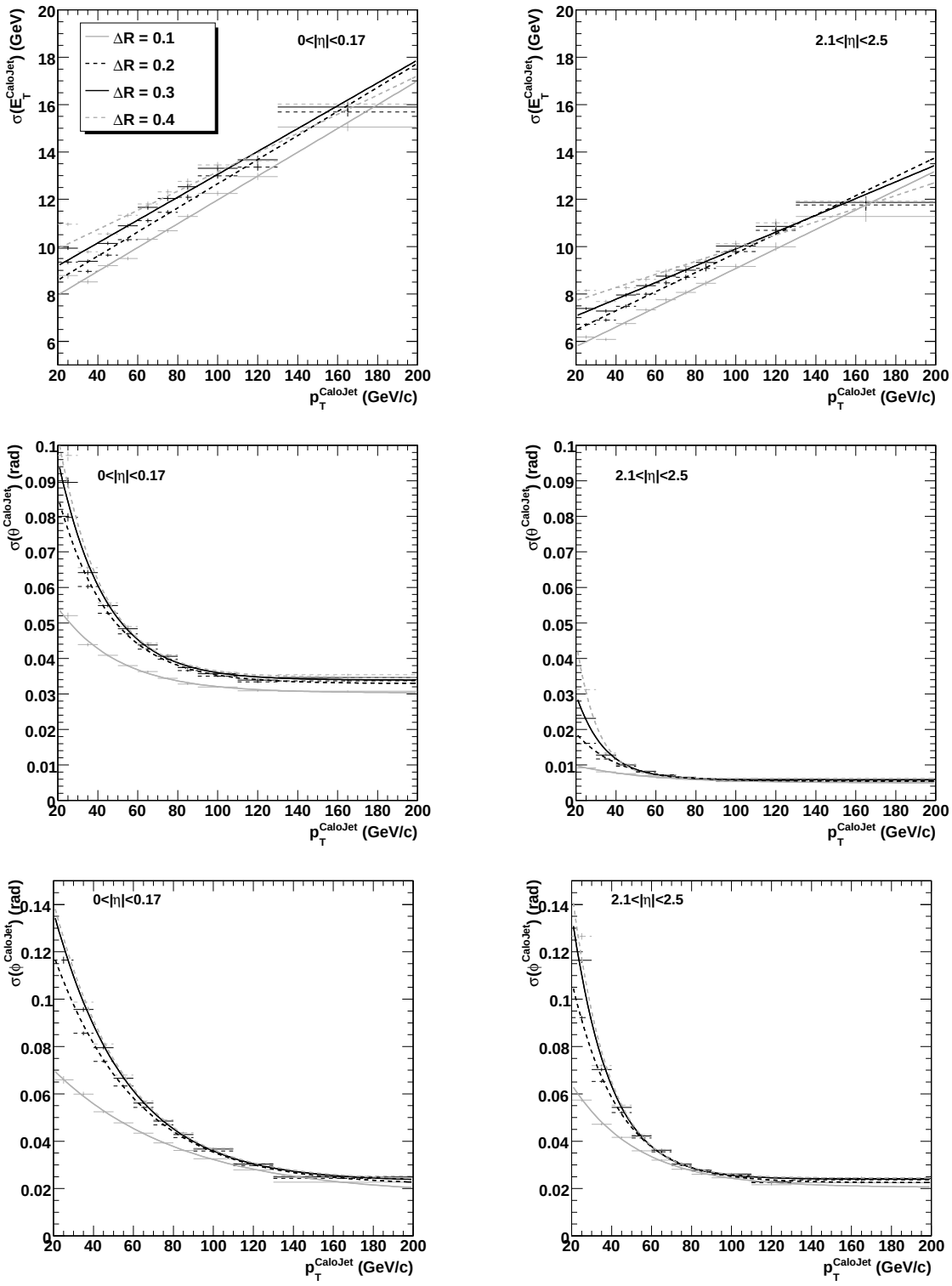


Figure 4.26: Effect of the matching criterion on the absolute resolutions for b jets in a central (left) and forward (right) pseudo-rapidity range.

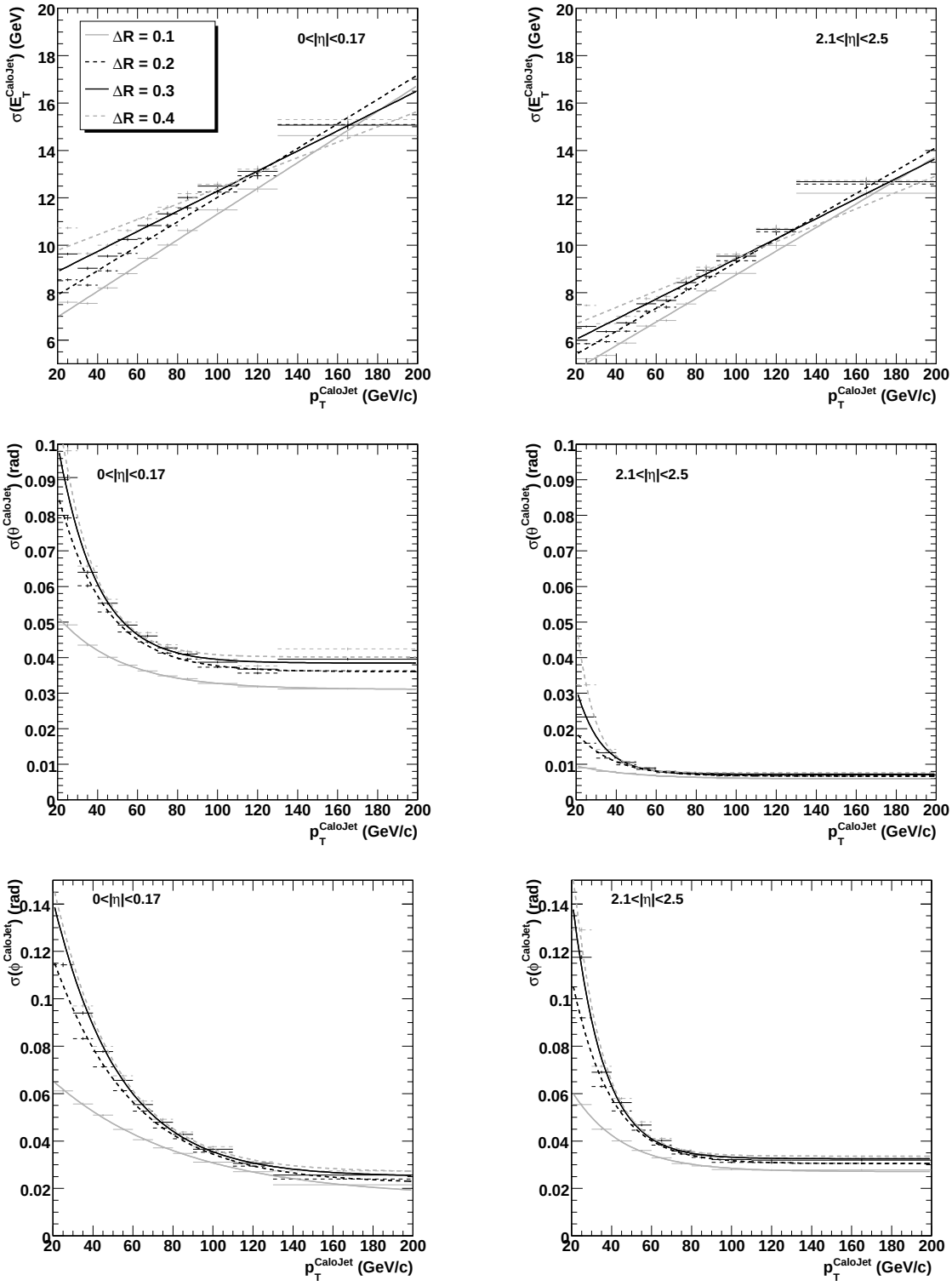


Figure 4.27: Effect of the matching criterion on the absolute resolutions for non-b jets in a central (left) and forward (right) pseudo-rapidity range.

4.2.1 Stand-alone muon reconstruction

The reconstruction of the muon path or track starts with the search for patterns of segments or so-called seeds in the DT and CSC chambers (Section 2.2.4). In the overlap region between the barrel and the endcap, the information of the RPC chambers is also used. Using the trajectory parameters of the seeds in the innermost chambers, the track parameters are propagated to the outermost layers using the Kalman filter technique [38]. The prediction of the track parameters and their uncertainties are compared and updated with each measurement that is available in the next chamber. The effect of multiple scattering, the non-uniformity of the magnetic field and muon energy loss in the material are taken into account when propagating the parameters of the initial track. After the outermost muon system is reached, a backward Kalman filter is applied, fitting the track from the outermost to the innermost chamber to define the track parameters in the latter. In the last stage, the track momentum resolution is improved by an extrapolation of the track using a fit that constrains it to the vertex of the hard interaction.

4.2.2 Global muon reconstruction

Global muons are obtained by matching the stand-alone muon tracks to the tracks reconstructed in the silicon tracker. Instead of matching the stand-alone muon tracks to all tracks that are reconstructed in the silicon tracker, a region of interest in $(\eta \times \phi)$ space in the silicon tracker is defined that contains the muon candidate tracks. The input for the definition of such a region is the location of the primary vertex, the direction and the minimum transverse momentum of the muon as given by the stand-alone muon track. The latter two variables are used to determine the region around the direction of the muon and to take into account the curvature of the tracking region. The subset of tracks lying in the region of interest are one by one matched with the stand-alone muon track. Therefore the track parameters of the stand-alone muon track and the candidate muon tracks in the silicon tracker are propagated to a common surface. Several tracker tracks can match the stand-alone muon track and a global refit is performed for each pair of tracker track and stand-alone track. The global fit uses both the hits from the tracker track and the stand-alone track to construct the global muon track. In case several global muon tracks are reconstructed for the same stand-alone muon track, the one with the best χ^2 of the fit is chosen. For muons with a high transverse momentum, the global fit is performed using only a subset of the tracker hits and the hits in the innermost muon system. The χ^2 of the global fit is then compared to the χ^2 of the track fitted in the tracker to detect important energy loss of the muon before it reaches the muon system.

4.2.3 Variables for muon identification

Apart from the muons produced in the decay of heavy bosons, also muons coming from the decay of kaons or pions or from the leptonic decay of heavy hadrons inside jets are reconstructed by the global muon reconstruction procedure. In Figure 4.28 the number of reconstructed global muons in the fully hadronic and semi-muonic $t\bar{t}$ decay is shown.

To reduce the contamination of muons not from the heavy boson decay in the collection of reconstructed global muons, muon identification criteria have been developed in the CMS collaboration [131]. A first variable used for muon identification is the normalized χ^2 of the global muon fit, shown in Figure 4.29. In the case of the semi-muonic $t\bar{t}$ decay,

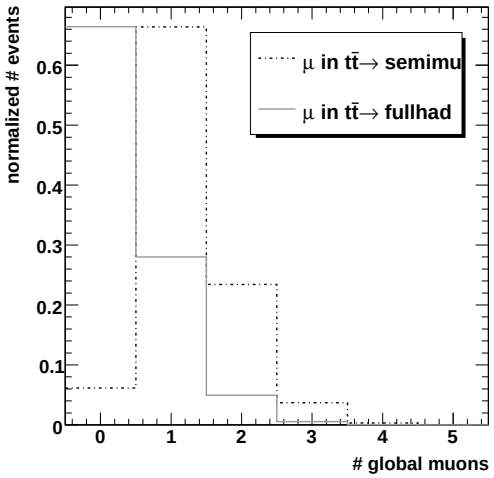


Figure 4.28: Number of global muons reconstructed in $t\bar{t}$ events.

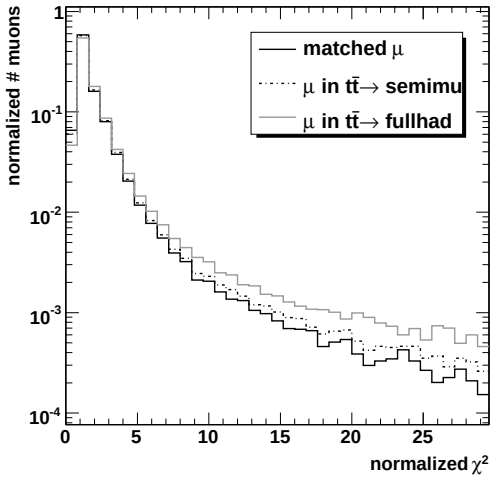


Figure 4.29: Normalized χ^2 of the global muon fit for global muons reconstructed in $t\bar{t}$ events.

the muon originating from the W boson has been identified by matching it with the generated muon by requiring $\Delta R < 0.1$. The value of the normalized χ^2 from the global fit is smaller for matched muons compared to muons originating from the decay of kaons, pions or the leptonic decay of heavy hadrons. A requirement for this variable has been proposed in the collaboration of normalized $\chi^2 < 10$. Two other requirements have been introduced for variables related to the fit in the silicon tracker, the number of hits in the silicon tracker fit and the transverse impact parameter of the fitted silicon track. The distributions of these variables are shown in Figure 4.30. The number of hits in the silicon fit is clearly higher for muons that have been matched compared to the muons in e.g. the fully hadronic $t\bar{t}$ decay. A loose cut on the number of hits in the silicon tracker fit is made by requiring that the number of hits is at least 11. Also for the transverse impact parameter a clear difference is visible between muons that are matched compared to all the muons in the semi-muonic or hadronic $t\bar{t}$ decay. A cut on the absolute value of the transverse impact parameter $|d_0| < 2\,\text{mm}$ was introduced. The three requirements introduced in this section are applied in the next section and chapters.

4.2.4 Resolution on the muon direction and transverse momentum

Resolutions on the muon direction and the transverse momentum have been calculated for muons reconstructed by both the global muon reconstruction and by the silicon

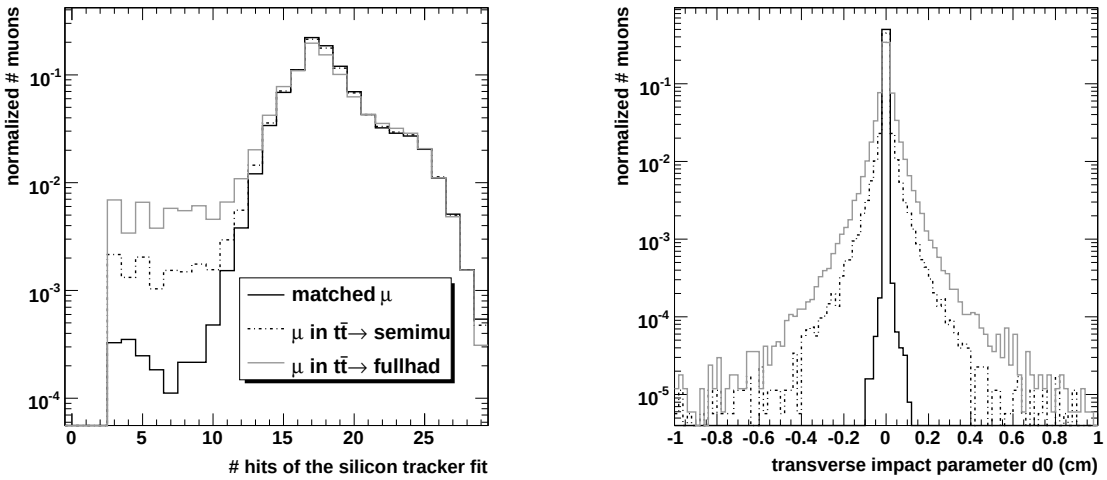


Figure 4.30: Number of valid hits in the silicon tracker fit (left) and transverse impact parameter for the silicon fit (right) for global muons reconstructed in $t\bar{t}$ events.

tracker. The cuts on the variables for muon identification have been applied as discussed in the previous section. Reconstructed muons are then matched with the muons in the generated $t\bar{t}$ final state by applying a $\Delta R < 0.1$ requirement between reconstructed and generated muons. In Figure 4.31 the absolute resolution on the polar angle and azimuthal angle are shown. Similarly as in the case of jets, the absolute resolution on the direction of muons becomes better for muons with high transverse momenta. But if we compare the absolute values of the resolutions of muons and jets, it is clear that the uncertainty on the muon direction is two orders of magnitude smaller than in the case of jets. The resolution on the polar angle is better for high pseudo-rapidities, while the opposite is true for the azimuthal angle. This is related to the magnetic field, as the muons with higher pseudo-rapidities have a longer path in the magnetic field.

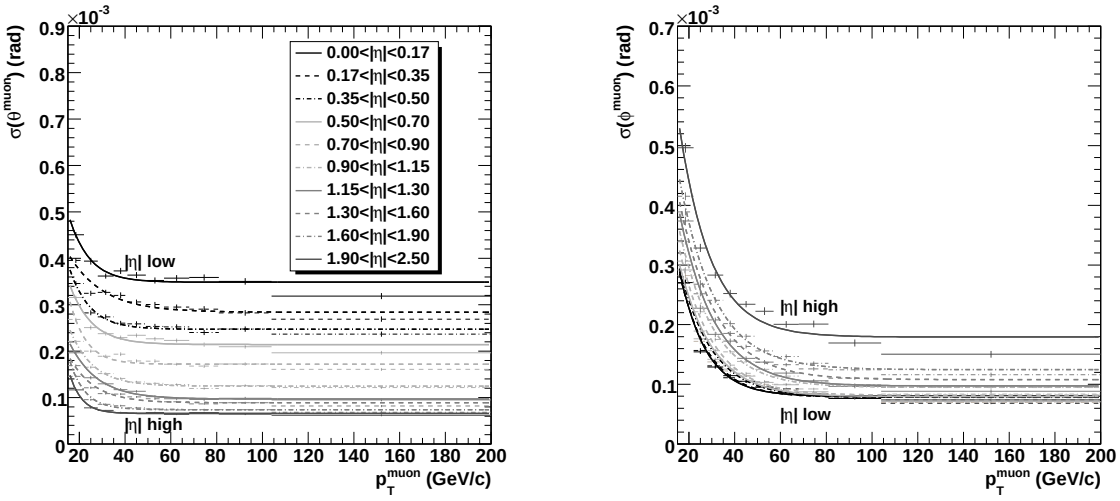


Figure 4.31: Absolute resolution on the polar angle (left) and the azimuthal angle (right) for muons.

The absolute resolution on the transverse momentum is shown in Figure 4.32. As for the ϕ resolution, the absolute resolution on the transverse momentum becomes worse with increasing pseudo-rapidity. The relative resolution on the transverse momentum improves from 5 % for 20 GeV/c muons to about 2 % for 200 GeV/c muons depending on the pseudo-rapidity. The effect of the matching requirement has been studied by altering ΔR between 0.1 and 0.3 in steps of 0.1 and as expected no change was seen in the resolutions.

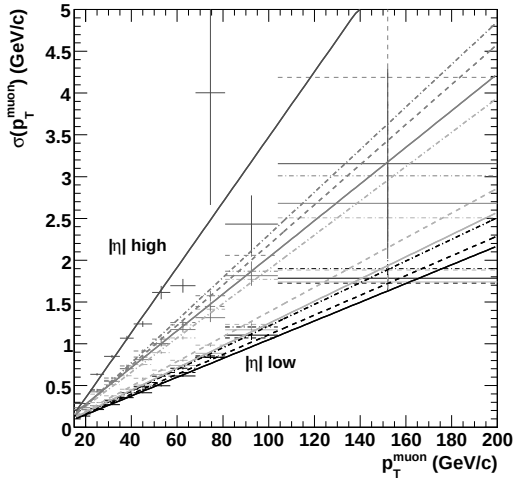


Figure 4.32: Absolute resolution on the transverse momentum of muons.

4.3 Missing transverse energy

Apart from the neutrino, the reconstruction of all objects produced in the final state of the $pp \to t\bar{t} \to bW\bar{b}W \to bq\bar{q}\bar{b}\mu\bar{\nu}_\mu$ process, have been discussed. Due to their very weak interaction with other particles, the neutrino escapes the CMS experiment undetected. However, by making the detector as hermetic as possible in order to register precisely the interaction of all the other particles, a physics object can be reconstructed which is related to the particles escaping detection. Full coverage in the forward direction is impossible, hence quite some energy is lost from particles escaping in the beampipe, but the transverse energy can be reconstructed. From the transverse energy balance it is possible to reconstruct the transverse energy of particles which escaped detection such as the neutrino. The missing transverse energy is calculated from the vectorial sum of all energy deposits in the calorimeters towers. A further improvement to the calculation of the missing transverse energy can be made by taking into account the required jet energy correction factors applied to the jets. The shifts in momentum on the reconstructed jets are therefore subtracted. Furthermore, a correction can be applied to the missing transverse energy by taking into account the transverse momenta of the reconstructed muons. Figure 4.33 shows the uncorrected missing transverse energy obtained in $t\bar{t}$ events by using the energy deposits as well as the missing transverse

energy corrected for the jet energy correction factors and the muon transverse momenta. The resolution on the reconstructed missing transverse energy in $t\bar{t}$ events can be

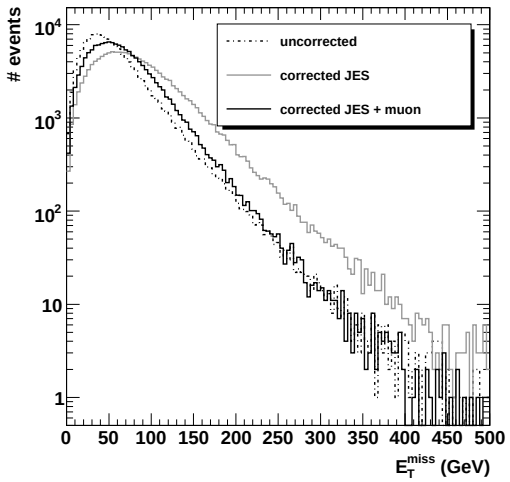


Figure 4.33: Missing transverse energy before and after jet energy corrections and correction for the muon deposits have been applied.

calculated by comparing the missing transverse energy with the transverse energy of the generated neutrino. Also the ϕ angle of the missing transverse energy can be compared with the direction of the neutrino. The resolutions are shown in Figure 4.34. The

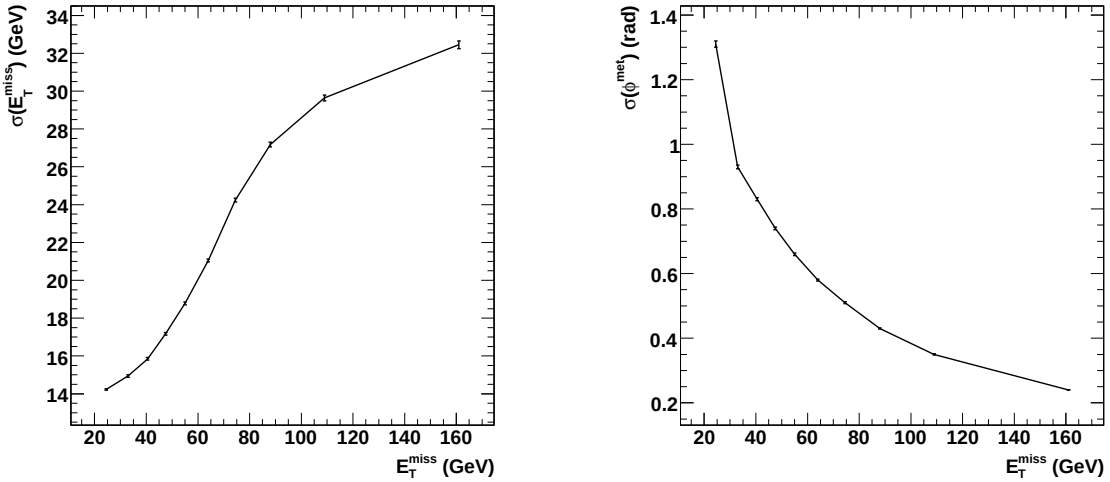


Figure 4.34: Resolution on E_T^{miss} and its ϕ direction as a function of E_T^{miss}.

uncertainty on the value of the missing transverse energy is between 20 and 50 % and decreases with increasing missing transverse energy. The resolution on the azimuthal direction varies between $\pi/3$ and $\pi/9$ radians for a missing transverse energy between 30 and 200 GeV.

Although missing transverse energy is an important handle for the discovery of many new physics phenomena, it should be handled with care due to the significant

uncertainties on both the energy itself as the direction. The missing transverse energy is therefore not used in the analysis.

The topology will drive the architecture.

Bill Tolpegin

Chapter 5

Selection and reconstruction of the event topology

A large variety of physics processes will be produced in the proton collisions at the LHC. After the reconstruction of the individual physics objects in the final state, the second step consists in selecting the physics process of interest among all the collision events that are recorded. Therefore, selection criteria are developed based on the topology of the final state of the physics process under study. These criteria aim to balance between an efficient rejection of background processes and keeping as many signal events as possible. In order to reconstruct the full topology of a cascade decay into hadronic jets, methods are required to assign these jets to the partons from which they originate. Additional ambiguities arise for the jet-parton assignment when additional jets are present in the event, resulting from the fragmentation of partons radiated in the initial or final state.

The physics process considered for the analysis in the current and the following chapters is the $pp \rightarrow t\bar{t} \rightarrow bW\bar{b}W \rightarrow bq\bar{q}\bar{b}\mu\bar{\nu_\mu}$ process. Therefore, at least four jets are expected in the final state, together with an isolated muon and missing transverse energy. In Section 5.1, various event selection criteria applied on the level of the reconstructed physics objects are studied and the applied cut values are listed. A dedicated study is performed on the isolation of the muon in the final state, because it is very important to reject the overwhelming multi-jets background. After applying the selection criteria, the jets coming from the four partons of the hard interaction need to be identified in the reconstructed event and thereafter assigned to the correct partons in order to reconstruct the full event topology. In Section 5.2 the influence of initial and final state radiation on the selection performance of the four correct jets is studied and different methods to assign the jets to the partons are discussed.

5.1 Selection of the event topology

Different event selection criteria are developed based on the expectation to observe at least four jets and an isolated muon in the final state of interest. These criteria are summarized in Section 5.1.1. The criterion that was originally constructed to label a muon as isolated still allowed for a significant efficiency to select events with muons

that are not originating from the decay of a heavy resonance. Therefore, the isolation of muons is studied in detail as described in Section 5.1.2. The cut values on the observables for the event selection criteria are discussed in Section 5.1.3.

5.1.1 Event selection criteria

Event selection criteria are applied on the level of the reconstructed physics objects in the final state. The optimal selection would be obtained by calculating with the observable quantities, the matrix element for the event. This results in a probability for the event to be $t\bar{t}$. However, this event selection method is CPU intensive. In addition, one needs to understand all reconstructed objects of the event to calculate the matrix element. Therefore, simpler event selection criteria are developed. These event selection criteria are sufficient for the measurement techniques proposed in the next chapters. While the event selection criteria are discussed in this section, some of the cut values and the table with the selection efficiencies for the different simulated processes are given in the next sections.

For the proton collisions produced at the LHC and detected by the CMS detector, the event first needs to be accepted by the trigger before it is written to storage. However, for simulated event samples, the trigger efficiency is often studied after all the event selection criteria are applied to obtain the offline trigger efficiency. This is the strategy applied for the presented physics study. The event selection criteria are based on the individual physics objects. Jets are reconstructed using the SC5 (opening angle $\Delta R = 0.5$) reconstruction algorithm and are L2+L3 corrected. For the event selection criteria that are based on muons, the reconstructed global muons are used. The following criteria are studied:

- **The event samples are skimmed (pre-selection) by requiring at least 4 jets after L2+L3 corrections with a $p_T^{\mathrm{L2L3}} > 15\,\mathrm{GeV/c}$ and "pointing" within the tracker acceptance ($|\eta| < 2.4$). In addition at least one muon is required with $p_T > 15\,\mathrm{GeV/c}$ and reconstructed within the muon system acceptance ($|\eta| < 2.5$).**
  The amount of events for the different simulated event samples is drastically reduced by these pre-selection requirements in the skimming. The skimmed event samples consist of reconstructed high-level physics objects or PAT candidates and are used as the starting point for different physics studies. The total computing time to repeat a physics study is significantly reduced by starting from the skimmed samples relative to the initial samples. In the construction of the PAT candidates the collections of physics objects are ordered in decreasing transverse momentum.

  After the pre-selection 72 % of the semi-muonic $t\bar{t}$ events are left. At this stage more specific event selection criteria are studied:

- **At least four jets with $p_T^{\mathrm{L2L3}} > 30\,\mathrm{GeV/c}$ are required.**
  The criterion is applied on the jets reconstructed with the SC5 jet clustering algorithm and the L2 and L3 corrections are applied. For many of the process

background samples, softer jets are expected. In addition, as discussed in Section 4.1.5, the response for L3 corrected jets is not flat for jets with low transverse momenta. In Figure 5.1 the transverse momentum of the fourth leading jet is shown for different processes. The amount of multi-jets ($pp \rightarrow \mu + X\,(p_T^{\mu} > 15)\,GeV/c$) events is overwhelming. The pseudo-data points are generated by selecting randomly events according to an integrated luminosity of $50\,pb^{-1}$ [1].

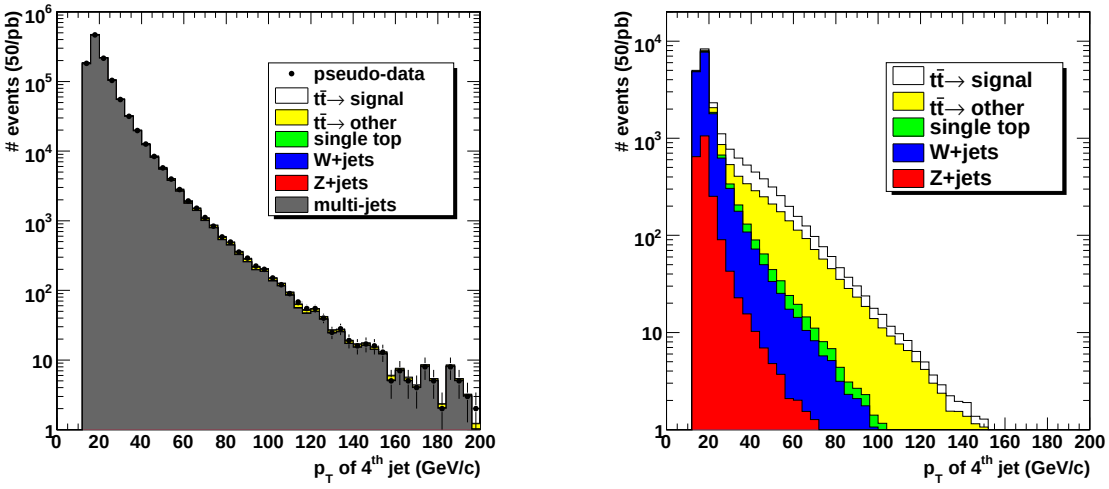


Figure 5.1: The transverse momentum of the fourth leading jet in the event with (left) and without (right) the inclusion of the multi-jets sample.

- **Muons are required to be in the L1 trigger acceptance ($|\eta| < 2.1$), to have a $p_T > 20\,GeV/c$ and to pass the muon identification criteria (quality cuts) to be accepted. At least one accepted muon is requested in the event to be selected.**
  The cuts discussed on the variables for muon identification in Section 4.2.3 are applied. These so-called quality cuts reduce the amount of events in which there are muons present originating from the leptonic decay of heavy hadrons. As an illustration, in Figure 5.2, the transverse momentum of the leading muon passing the muon identification criteria is shown for the different simulated processes. The muons produced in the decay of heavy objects, such as the W and Z boson decays have clearly higher transverse momenta compared to the muons produced in for example multi-jets events. The depletion of the transverse momentum distribution of the Z+jet events at low transverse momentum in the right figure is due to the intrinsic presence of two muons with a high transverse momentum from the Z boson decay compare to only one from the decay of the W boson in other processes.

- **At least one accepted muon is required to pass the isolation requirements.**

[1] The expected contribution of multi-jets events is added rather than randomized because the simulated sample has an integrated luminosity of only $50\,pb^{-1}$.

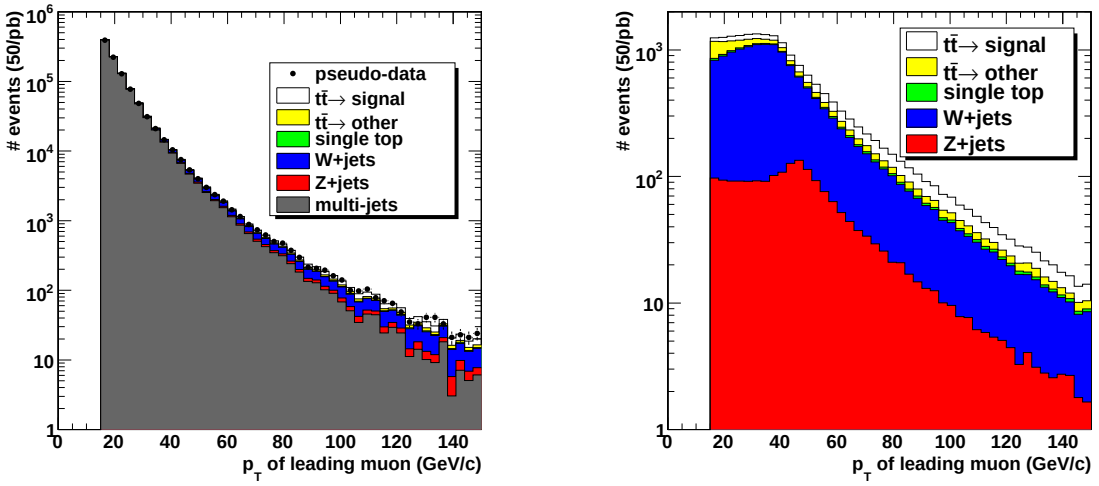


Figure 5.2: The transverse momentum of the leading muon passing the muon identification criteria with (left) and without (right) the inclusion of the multi-jets sample.

The requirement on the isolation of the muon is the most powerful criterion applied on the reconstructed physics objects in the final state. A large fraction of events in which muons were reconstructed that were not produced in the decay of heavy resonances will be removed. Given the importance of the criterion, the next section is dedicated to it.

- **A veto is applied for a second isolated muon. The isolated muon with the highest transverse momentum is identified as the muon of interest.** In some processes, events with two isolated muons are expected, like for instance in the $Z(\rightarrow \mu\mu)$+jets or di-leptonic decay of the $t\bar{t}$ process, where the two leptons are muons. By vetoing a second isolated muon in the list of accepted muons, we can reduce the number of events of these background processes, while keeping almost all the signal events.

- **The event should be accepted by the isolated single muon trigger.** As already mentioned earlier, the trigger efficiency was studied after the other event selection criteria are applied. The offline trigger efficiency gives an estimation of the amount of potential interesting events that might be lost due to the trigger requirements.

  The trigger stream that was used in this analysis is the softest single muon trigger stream for which no pre-scaling is needed, and it is given in Table 5.1. The HLT trigger bit, `HLT_Mu9`, is based on the seed from the L1 trigger bit, `L1_SingleMu7`. The muon seed from the `L1_SingleMu7` trigger bit is a muon reconstructed by one or more muon subdetector types and it's transverse momentum should exceed 7 GeV/c. The `HLT_Mu9` is based on a muon reconstructed using the combined muon system and silicon tracker (HLT muon). The transverse momentum of the HLT muon is required to exceed 9 GeV/c. Additionally, the transverse impact parameter between the muon and the beamspot is required to be less than 2 cm.

trigger level	trigger bit	single muon p_T-threshold	event rate
L1	`L1_SingleMu7`	7 GeV/c	32.03±0.94 Hz
HLT	`HLT_Mu9`	9 GeV/c	1.82±0.22 Hz

Table 5.1: Low luminosity ($2.9\,10^{30}\mathrm{cm}^{-2}\mathrm{s}^{-1}$) trigger thresholds and expected rates for the single muon stream relevant for the physics study in this analysis.

The criterion on the transverse impact parameter is applied to reduce random triggers due to cosmics rays propagating through the detector.

5.1.2 Muon isolation

In many physics studies in hadron collisions, it is important to determine whether or not a muon was produced isolated in the hard event from heavy boson decays or non-isolated in the fragmentation process. Many variables for this purpose exist that are able to differentiate between isolated and non-isolated muons. Usually the "activity" in a region around the muon is studied and the muon is considered isolated if the "activity" is low. One strategy is explained by considering the illustration in Figure 5.3. An isolation cone with opening angle $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ is constructed around the

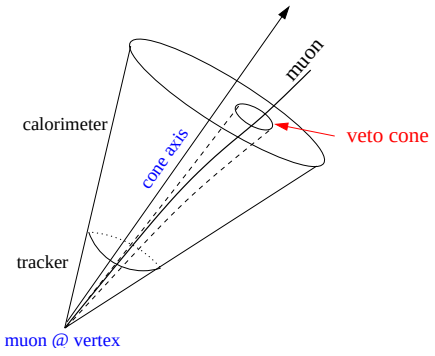


Figure 5.3: Illustration of the isolation cone around a muon and the veto cone used to subtract the contribution of the muon itself.

direction of the muon at the vertex. Depending on the variable under consideration, a veto cone is constructed around the direction of the muon at the surface of the ECAL and/or HCAL. The deposits of the muon in respectively the ECAL or HCAL veto cone are removed from the calculation of the "activity". The size of the veto cone depends on the subdetector for which the isolation variable is calculated.

A variable that is often used in the Top Physics Analysis Group in the CMS collaboration is the so-called relative isolation. The opening angle of the isolation cone is taken to be $\Delta R = 0.3$, while the opening angle of the veto cone is taken as $\Delta R = 0.07$

in the ECAL and $\Delta R = 0.1$ in the HCAL. The relative isolation is then defined as

$$RelIso = \frac{p_T^{\mu}}{p_T^{\mu} + \sum p_T^{tracker} + \sum E_T^{ECAL} + \sum E_T^{HCAL}}, \tag{5.1}$$

where p_T^{μ} is the transverse momentum of the muon, $\sum p_T^{tracker}$ denotes the sum of the transverse momenta of the tracks in the isolation cone with exclusion of the muon track and $\sum E_T^{ECAL}$ ($\sum E_T^{HCAL}$) are the sum of the transverse energy deposits in the ECAL (HCAL) with the exclusion of the crystals (towers) in the veto cone. After applying the pre-selection, the cuts on the jets and the cuts on the transverse momentum, pseudo-rapidity and identification variables of the leading muon, the relative isolation variable is shown in Figure 5.4 for semi-muonic $t\bar{t}$ decays and for the multi-jets ($pp \rightarrow \mu + X\,(p_T^{\mu} > 15)\,GeV/c$) process. For the signal process, the distribution is peaked at one, indicating that the muon is isolated as the denominator in Equation 5.2 is almost equal to the numerator or p_T^{μ}. A cut can be applied on the relative isolation and the efficiency for selecting the signal events and the purity (the ratio of the number of signal events after the cut with respect to the total number of signal and background events after the cut) can be calculated for different cut values. The right plot of Figure 5.4 shows the efficiency versus the purity of a cut on the RelIso variable. The purity starts at 20% because a lower cut value of RelIso = 0.7 was applied. The efficiency decreases rather slowly when cutting harder on the relative isolation and a relatively high purity can be reached without loosing too many signal events.

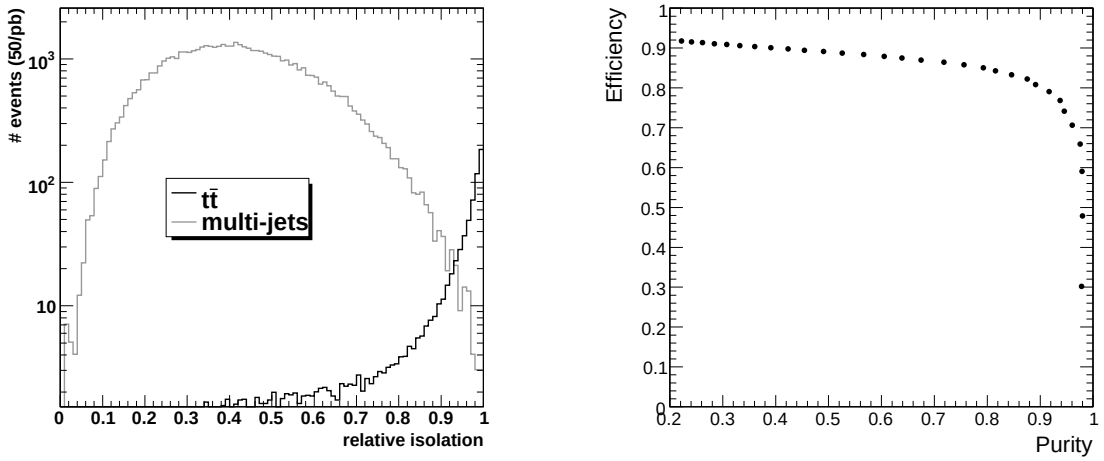


Figure 5.4: The relative isolation variable (left) and the efficiency versus the purity of a cut on this variable (right), starting from a lower cut value of 0.70.

In the Top Physics Analysis Group in the CMS collaboration usually a requirement on the relative isolation of RelIso > 0.95 is applied [132], which is also the case for the physics analysis presented in this thesis. Using the requirement RelIso > 0.95 yields a signal selection efficiency of 74 %, while a purity of 0.95 % with respect to the amount of multi-jets events, is reached.

After applying the requirement on the relative isolation, there are still multi-jets events selected. To understand why these events reach relative isolation values close to one, additional studies are performed. Because the amount of simulated multi-jets

events is rather low after the relative isolation criterion, the criterion was relaxed to 0.9 for these additional studies. Figure 5.5 shows the smallest ΔR angle between the muon and the closest jet, where the jet was required to have a transverse momentum exceeding 15 GeV/c. For about 50±6 % of the multi-jets events the ΔR angle is smaller than 0.1, which indicates that the muon is reconstructed inside the jet and is therefore not isolated. For a cut on the relative isolation of 0.95, the fraction of events with a ΔR angle smaller than 0.1 decreases to about 35±8 %, which is still relatively high. The reason why these events are not removed by the cut on the relative isolation is because the jet in which the muon is contained is rather narrow and the contribution of the jet activity to the denominator in the relative isolation expression is falling inside the veto cone around the muon. The hypothesis that the jet is small is confirmed by looking at the number of constituents carrying 90 % of the energy of the jet closest to the muon, as shown in Figure 5.6. For jets closer than $\Delta R = 0.3$ to the muon, the number of constituents is lower than for jets that are well separated from the muon. In Section 4.1.4

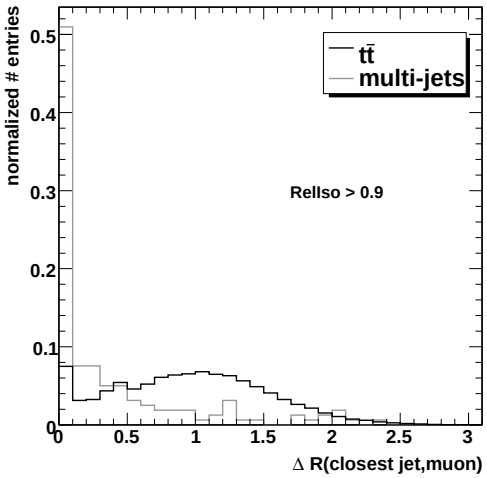


Figure 5.5: ΔR between the muon and the closest jet for the muon passing the relative isolation cut of 0.9.

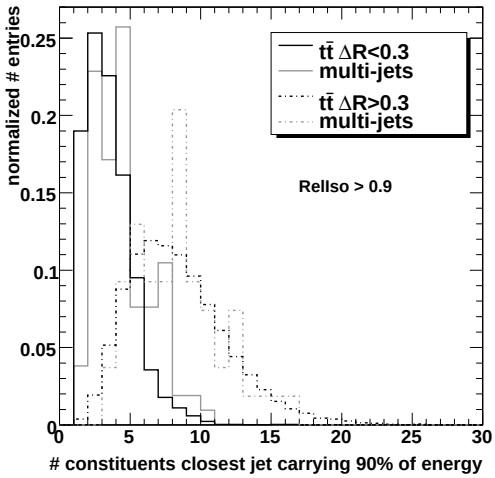


Figure 5.6: The number of constituents carrying 90% of the energy of the jet closest to the muon.

it was discussed that jets originating from b or c quarks contain heavy hadrons that may decay into muons. Such muons have lower transverse momenta compared to the muons produced in the hard interaction. Figure 5.7 shows the transverse momentum of the leading muon for events where there is a jet closer than $\Delta R = 0.3$ to the muon and for events where the jets are further away. In contrast with the expectation, the muons in the vicinity of the jet have a higher transverse momentum. A muon passes the RelIso > 0.9 criterion either because there are not too much energy deposits around the muon or because the muon has a very high transverse momentum such that the possible energy deposits do not contribute too much to the denominator in the RelIso variable. Therefore, by requiring RelIso > 0.9 and $\Delta R < 0.3$, muons with a higher transverse momentum are artificially selected. A cut on the transverse momentum, p_T, of the muons is not expected to help in the reduction of these non-isolated high p_T muons.

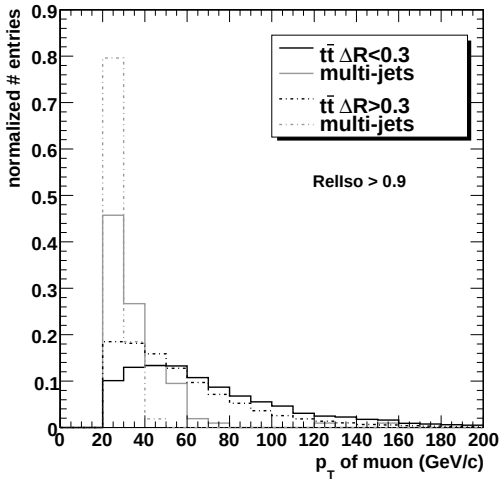


Figure 5.7: Transverse momentum of the muon.

We can also calculate the energy deposited in the ECAL in a (3×3) crystal shape around the crystal crossed by the muon and similarly the energy deposited in the HCAL in a (3×3) tower shape around the crossed tower, as shown in Figure 5.8. Also for these two variables a clear difference is visible between the events in which the muon is inside a jet compared to when the muon is separated from the jets by more than $\Delta R = 0.3$ is visible. The energy deposited is shown to be lower if no jet is in the vicinity of the muon.

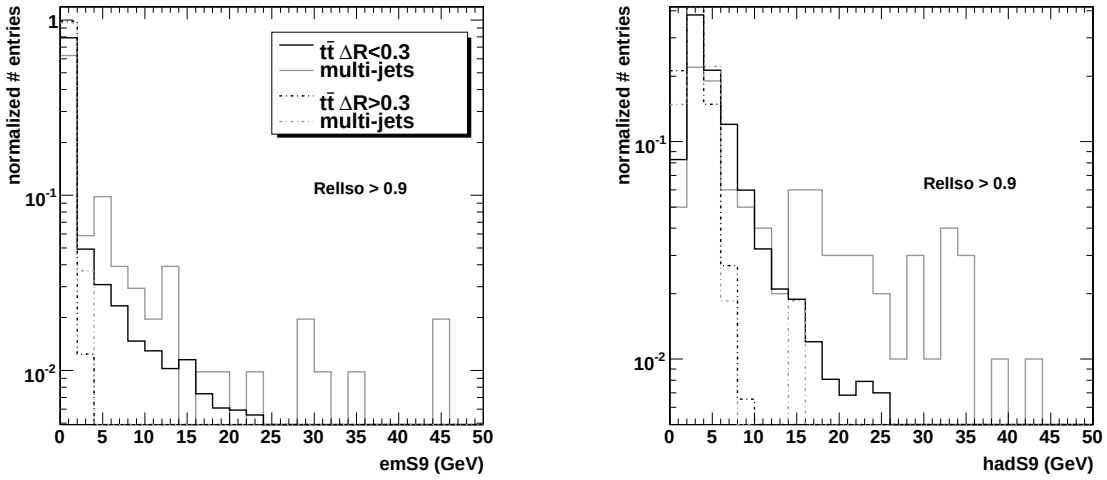


Figure 5.8: Energy deposited in a (3×3) ECAL crystal shape around the crossed crystal (left) and in a (3×3) HCAL tower shape around the crossed tower (right).

Two variables that confirm the hypothesis that the muon is part of a narrow jet of which the activity is removed by applying the veto cone around the muon, are the sum of the energy deposits in the veto cones in the ECAL and the HCAL. These variables are shown in Figure 5.9 and have smaller values in case the closest jet is further than $\Delta R = 0.3$ from the muon.

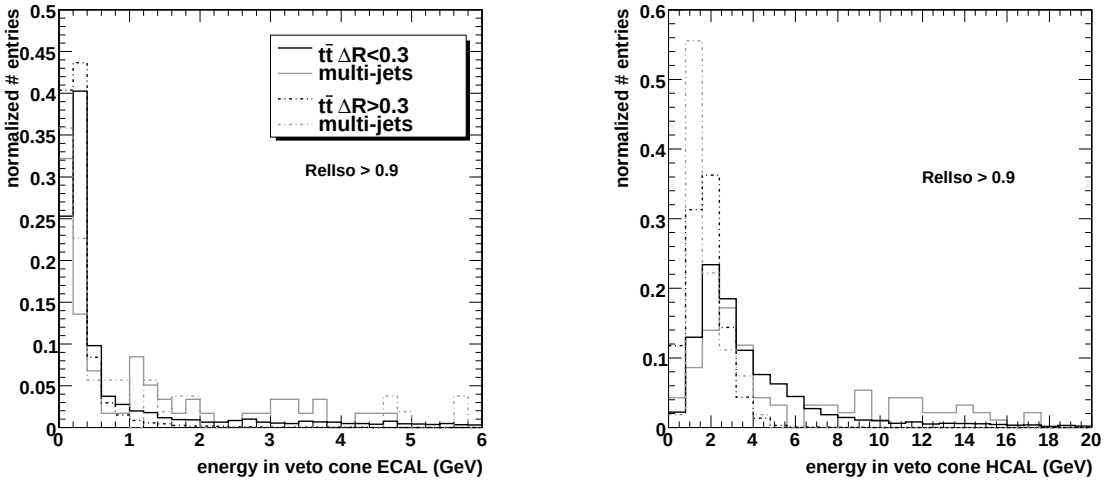


Figure 5.9: Energy deposited in the veto cone around the muon for the ECAL (left) and for the HCAL (right).

Thus, by applying additional selection criteria on the variables that were shown in the plots, the events for which the muon is inside a jet can be removed. The Top Physics Analysis Group agreed on requiring the energy in the veto cone of the ECAL to be below 4 GeV and for the HCAL below 6 GeV. Another possibility that was widely used, is to apply a cut on the angle between the closest jet and the muon of e.g. $\Delta R > 0.3$. There are many possibilities depending on the need for a high efficiency or a high purity in the physics measurements. In Figure 5.10 the efficiency and purity for the two additional cuts on variables discussed above are shown. The solid line with the black dots shows the event selection efficiency versus purity relative to the multi-jets background for a cut on the relative isolation of the leading muon. However, the procedure that is applied in most physics studies and also in this thesis, is to look for the muon with the highest transverse momentum that is passing the relative isolation criterion RelIso > 0.95. This means that in case the leading muon is not passing the relative isolation criterion, the next muon is considered. The open cross in the figure corresponds to the efficiency and purity for a cut on the relative isolation of RelIso > 0.95 following this procedure. If we require in addition the energies in the veto cones to be below the cuts mentioned, we end up with a slightly higher purity (star). The line with the open dots corresponds to the default relative isolation criterion with in addition the ΔR requirement between the muon and the closest jet which was varied between $\Delta R > 0.05$ and $\Delta R > 0.5$. A relative isolation requirement of RelIso > 0.95 together with for example the $\Delta R > 0.3$ requirement yields a higher purity than the requirement on the energies of the vetocones. However, similar purity and efficiency can be obtained by requiring a relative isolation of RelIso > 0.97 on the leading muon. For the measurement method presented in the next chapters, the need for a high efficiency is more important than a high purity, therefore a relative isolation of RelIso > 0.95 was applied with the additional requirement on the energies deposited in the veto cones of the ECAL ($E_{\text{veto}}^{\text{ECAL}} < 4\,\text{GeV}$) and the HCAL ($E_{\text{veto}}^{\text{ECAL}} < 6\,\text{GeV}$).

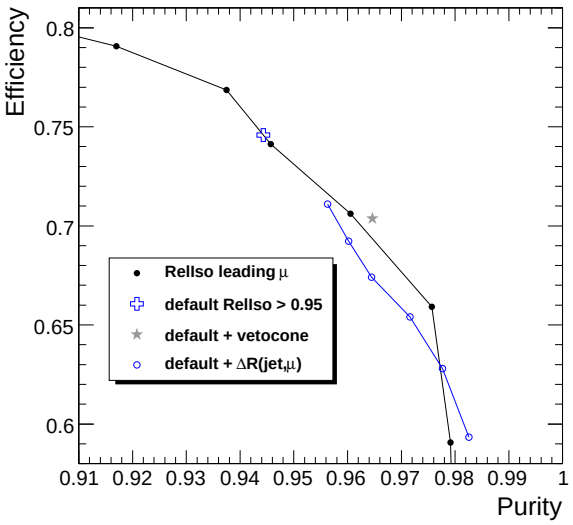


Figure 5.10: Zoom of the efficiency versus the purity of a cut on the relative isolation variable with some reference points.

5.1.3 Efficiency of the event selection criteria

The event selection criteria described in Section 5.1.1 are applied with the values for the muon isolation variables as specified in Section 5.1.2. Figure 5.11 shows the isolation variables and the number of isolated muons passing the isolation requirements for the various simulated signal and background processes. The distributions were obtained after applying the pre-selection criteria and the cut on the transverse momenta and η acceptance of the four leading jets and muon. In addition the requirements on the muon identification variables and the pseudo-rapidity of the muon are applied. Therefore, the three distributions of the isolation variables are made for the not yet isolated muon. After the muon isolation criteria were applied and the muons are thus accepted as isolated, the distribution of the number of isolated muons is made. In this distribution, the contribution to the bin with 2 isolated muons comes from $Z(\rightarrow \mu\mu)$+jets or the di-leptonic decay into muons of the $t\bar{t}$ process.

Table 5.2 summarizes the remaining number of events after the selection together with the efficiencies of the applied event selection criteria for the samples that were used in the analysis. The pre-selection has the largest reduction factor for the W+jets and Z+jets sample, because at least four jets with transverse momenta exceeding 15 GeV/c were required. Requiring at least four jets with transverse momenta exceeding 30 GeV/c reduces the multi-jets background with a factor of about 10 and the W/Z+jets background with a factor of about 25, compared to a factor of about 1.5 for $t\bar{t}$. The quality cuts on the muon, the cut on the transverse momentum of 20 GeV/c and on the pseudo-rapidity reduces by about a factor of two the contribution of the $t\bar{t}$ →other decays and the amount of $pp \rightarrow \mu + X$ multi-jets events. These two background processes are once more drastically reduced by the requirements on the muon isolation, a factor of about 5 for $t\bar{t}$ →other and over three orders of magnitude for $pp \rightarrow \mu + X$ events. The requirements on the muon isolation also reduces

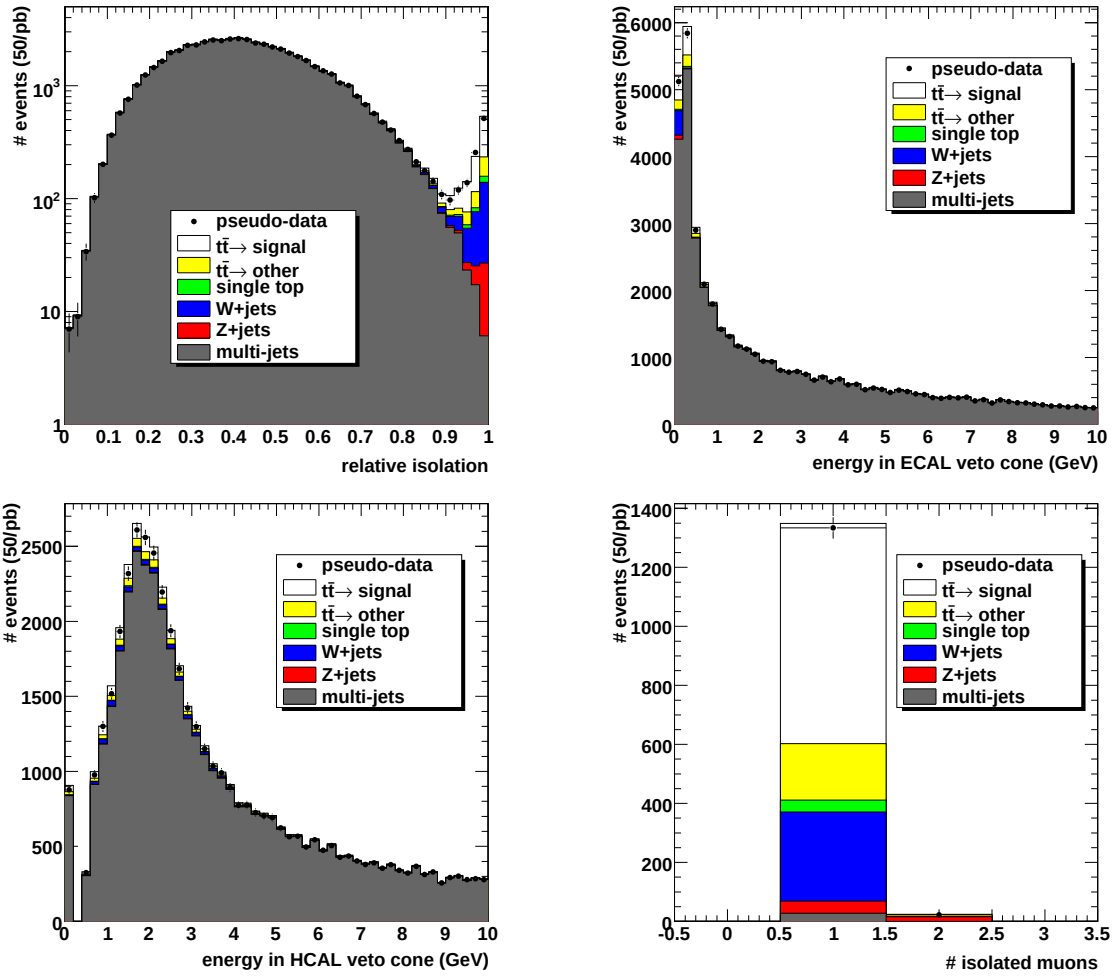


Figure 5.11: Relative isolation (upper left), energy deposited in the ECAL veto cone (upper right) and the HCAL veto cone (lower left) for the leading muon, and the number of isolated muons (lower right).

	$t\bar{t} \to \mu$	$t\bar{t} \to$other	$t(\bar{t})$	W+jets	Z+jets	$pp \to \mu + X$
# events	3.07 k	17.6 k	3.61 k	2.28 M	211 k	6.41 M
pre-selection	$71.3\,10^{-2}$	$14.7\,10^{-2}$	$11.4\,10^{-2}$	$59.8\,10^{-4}$	$10.2\,10^{-3}$	$17.3\,10^{-2}$
4 jets, p_T	$40.7\,10^{-2}$	$10.2\,10^{-2}$	$4.05\,10^{-2}$	$2.53\,10^{-4}$	$4.43\,10^{-4}$	$1.79\,10^{-2}$
p_T,$\|\eta\|$,quality μ	$34.6\,10^{-2}$	$5.18\,10^{-2}$	$2.47\,10^{-2}$	$1.94\,10^{-4}$	$3.54\,10^{-4}$	$8.98\,10^{-3}$
RelIso	$25.8\,10^{-2}$	$1.19\,10^{-2}$	$1.30\,10^{-2}$	$1.41\,10^{-4}$	$2.82\,10^{-4}$	$7.28\,10^{-6}$
vetocone energy	$24.3\,10^{-2}$	$1.13\,10^{-2}$	$1.12\,10^{-2}$	$1.33\,10^{-4}$	$2.72\,10^{-4}$	$4.27\,10^{-6}$
exactly 1 μ	$24.3\,10^{-2}$	$1.08\,10^{-2}$	$1.12\,10^{-2}$	$1.33\,10^{-4}$	$1.97\,10^{-4}$	$4.27\,10^{-6}$
offline HLT	$22.5\,10^{-2}$	$1.00\,10^{-2}$	$1.02\,10^{-2}$	$1.24\,10^{-4}$	$1.88\,10^{-4}$	$3.80\,10^{-6}$
# events	689.6±3.9	176.0±2.0	36.82±0.55	283.50±2.62	39.66±1.04	24.34±4.97

Table 5.2: Overview of efficiencies obtained after the different object selection cuts applied to the signal and background event samples. The indicated number of events before and after the cuts are rescaled to an integrated luminosity of $\mathcal{L}$ =50 pb^{-1} using NLO cross-sections when available, otherwise LO cross sections were used (cfr. Tables 3.3, 3.4,3.5).

the amount of events of processes in which isolated muons are produced, because the reconstructed muon is not within the required pseudo-rapidity and no isolated muon is thus found. Requiring exactly one isolated muon reduces, as expected, only the amount of $t\bar{t} \to$other and Z+jets events. The offline trigger efficiency is 92.4 ± 0.1 % for $t\bar{t} \to \mu$ and 92.3 ± 0.3 % for $t\bar{t} \to$other, compared to 90.6 ± 0.3 % for single top, 93.8 ± 0.2 % for W+jets, 95.4 ± 0.5 % for Z+jets and 89 ± 6 % for the multi-jets events [2]. When all selection cuts are applied, the signal over background ratio is 1.23 and about 690 signal events are expected for an integrated luminosity of 50 pb^{-1} compared to 560 background events.

In Table 5.3 the efficiencies are listed for the background processes that were not used in the analysis to avoid double counting with other simulated samples. For instance some of the VQQ and the Wc processes are also contained in the W+jets sample, while events simulated in the different QCD $\hat{p_T}$-bins can also appear in the

[2] The difference in the offline trigger efficiency between the W/Z+jets samples and the $t\bar{t}$ sample is related to the use of fast simulation for the former samples and full simulation for the latter.

	VQQ	Wc	QCD $\hat{p_T}$ 100-250	QCD $\hat{p_T}$ 250-500	QCD $\hat{p_T}$ 500-1000	QCD $\hat{p_T}$ 1000-∞
# events	14.5 k	74.2 k	749 M	19.9 M	702 k	18.2 k
pre-selection	$1.25\,10^{-2}$	$1.16\,10^{-2}$	$17.9\,10^{-4}$	$22.6\,10^{-3}$	$97.6\,10^{-3}$	$25.1\,10^{-2}$
4 jets, p_T	$4.39\,10^{-4}$	$3.38\,10^{-4}$	$1.17\,10^{-4}$	$8.23\,10^{-3}$	$57.8\,10^{-3}$	$20.0\,10^{-2}$
p_T,$\lvert\eta\rvert$,quality μ	$2.86\,10^{-4}$	$2.43\,10^{-4}$	$1.34\,10^{-5}$	$1.21\,10^{-3}$	$7.00\,10^{-3}$	$1.77\,10^{-2}$
relative isolation	$1.79\,10^{-4}$	$1.79\,10^{-4}$	0.0	$7.90\,10^{-7}$	$1.96\,10^{-6}$	$1.93\,10^{-6}$
veto cone energy	$1.68\,10^{-4}$	$1.68\,10^{-4}$	0.0	$5.92\,10^{-7}$	$1.09\,10^{-6}$	0.0
exactly 1 μ	$1.45\,10^{-4}$	$1.45\,10^{-4}$	0.0	$5.92\,10^{-7}$	$1.09\,10^{-6}$	0.0
offline HLT	$1.38\,10^{-4}$	$1.38\,10^{-4}$	0.0	$1.97\,10^{-7}$	$0.65\,10^{-6}$	0.0
# events ($50\,\mathrm{pb}^{-1}$)	2.0±0.2	10.4±0.5	0±61	4±4	0.5±0.3	0.00±0.02

Table 5.3: Overview of efficiencies obtained after the different object selection cuts applied to background event samples that were not used in the analysis to avoid double counting. The indicated number of events before and after the cuts are rescaled to an integrated luminosity of $\mathcal{L}$ =50 pb^{-1} using LO cross-sections.

$pp \rightarrow \mu + X$ or vice versa. The number of VQQ and Wc events that remains when all selection criteria are applied is about 12 for an integrated luminosity of $50\,\mathrm{pb}^{-1}$. These events are neglible with respect to the 284 selected events remaining in the W+jets sample. Within its large uncertainty, the total number of selected events over all QCD $\hat{p_T}$-bins is compatible with the selected number of events from the $pp \rightarrow \mu + X$ sample. Therefore, only the samples in Table 5.2 are used in the analysis.

5.2 Reconstruction of the event topology

In the previous section, selection cuts on the level of the properties of individual objects were applied to enhance the fraction of $t\bar{t} \rightarrow bW\bar{b}W \rightarrow bq\bar{q}\bar{b}\mu\bar{\nu_\mu}$ events in the selected sample. To reconstruct the event topology, i.e. the reconstruction of the properties of the top quarks and W bosons from their decay products in the hadronic $t \rightarrow bW \rightarrow bq\bar{q}$

and leptonic $\bar{t} \to \bar{b}W \to \bar{b}\mu\bar{\nu_\mu}$ branch [3], four jets need to be assigned correctly to the four initial quarks in the underlying top quark and W boson branching. The matching between jets and quarks and the influence of radiated partons is studied in Section 5.2.1. Methods to associate jets to the quarks in the final state based on variables that have discriminating power between correct and wrong jet-quark assignments are discussed in Section 5.2.2.

5.2.1 Radiation

To study the jet-parton matching efficiency, only the semi-muonic $t\bar{t}$ decays are considered. The final event selection as given in Section 5.1.3 is applied, with the exception of the criterion on the transverse momentum of the four leading jets. For the jets in the event, only the pre-selection cuts are applied, which requires the transverse momentum of the jets to exceed 15 GeV/c and the direction of the jet to be within the tracker acceptance ($|\eta| < 2.4$).

To match the partons of the $bq\bar{q}\bar{b}\mu\bar{\nu_\mu}$ topology to the jets, the partons are first sorted according to decreasing transverse momentum. The parton with the highest transverse momentum is considered first and is matched to its closest jet in $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$. The jet is removed from the list and the next parton is considered for matching. The procedure is repeated until all four partons are matched. The matching between a parton and a jet is successful if they are closer than $\Delta R = 0.3$ to each other.

The fraction of semi-muonic $t\bar{t}$ events for which there are at least four jets passing a threshold on the transverse momentum, relative to the number of pre-selected events, is shown in Figure 5.12. The selection efficiency rapidly drops as a function of the cut on the transverse momenta. For instance for $p_T > 30$ GeV/c as considered in the event selection in Section 5.1.3, the efficiency is about 55 %. Figure 5.13 shows the fraction of events for which the four initial quarks in the decay topology are matched to four of the jets that are part of the collection of jets in the event exceeding the indicated transverse momentum threshold. The fraction is calculated relative to the number of events for which there are at least four jets passing the p_T threshold. The considered fraction is maximal with a value of about 40 % for a transverse momentum of 20-30 GeV/c, which motivates to apply a cut around this value.

To determine if the jets can be matched to the quarks in the event, we consider for the matching all jets with a transverse momentum above 15 GeV/c. If we consider only those events where the matching exists, we can calculate the efficiency of selecting the events after applying a p_T cut on the four leading jets. This efficiency is shown in Figure 5.14. If we compare it with Figure 5.12, we see that the efficiency loss as a function of the cut on the transverse momentum is less steep in case we look at the events for which the matching exist. For instance for $p_T > 30$ GeV/c, the efficiency is around 70 % for events where the matching exist, compared to 55 % in case we consider all events. For transverse momenta higher than $p_T > 60$ GeV/c, the efficiencies are comparable. In Figure 5.15 the purity is shown, i.e. the fraction of events for which the four leading jets are matched to the four quarks in the final state, relative to the number of events that have at least four jets above the indicated p_T threshold. This

[3] The charge conjugation is implicitly considered.

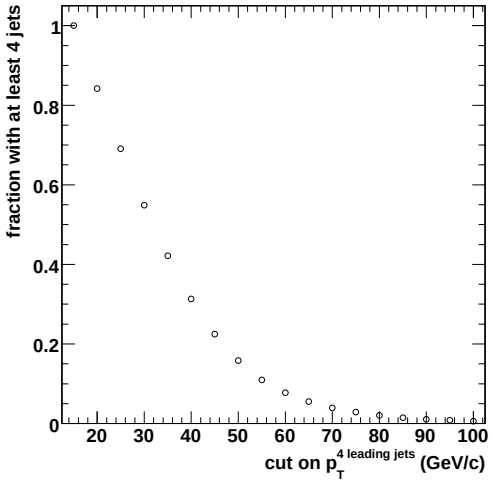


Figure 5.12: Fraction of events with at least 4 jets above a p_T threshold with respect to the number of pre-selected semi-muonic $t\bar{t}$ events.

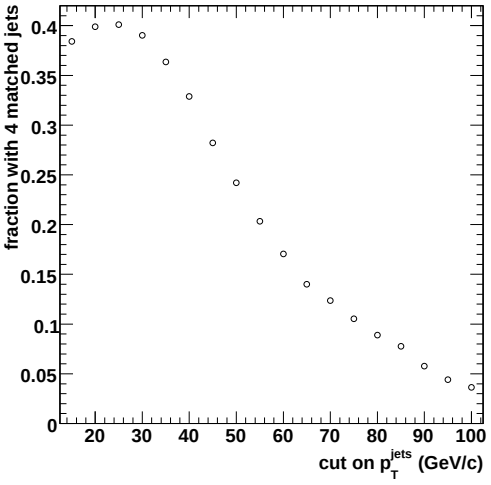


Figure 5.13: Fraction of events where the 4 partons are matched with the jets exceeding a p_T threshold, relative to the number of events with at least 4 jets above the p_T threshold.

purity is maximal for a p_T cut of 30 GeV/c on the four leading jets and reaches a value of about 28 %. If we multiply for each p_T cut value the purity of Figure 5.15 with the efficiency of selecting semi-muonic $t\bar{t}$ events with at least four jets in the final state (cfr. Figure 5.12), we obtain that the maximum fraction of events in which the jets are found and can be matched to the partons beyond the pre-selection is about 15 % when requiring at least four jets with $p_T > 30$ GeV/c. This relatively low number is due to radiation in the initial and final state. While hard radiation in the final state ($q \to q'g$) yields two jets from the q' and g reconstructed generally in a different direction compared to the original quark direction, the jets produced by initial state radiation have a relatively high transverse momentum and thus a high possibility to end up among the four leading jets. Over 55 % of the selected semi-muonic $t\bar{t}$ events contain initial state radiation. In Figure 5.16 the transverse momentum of the quarks in the semi-muonic $t\bar{t}$ final state are shown, together with the transverse momentum of the partons radiated in the initial state. This figure also shows the lowest p_T quark in the $t\bar{t}$ decay topology together with the transverse momentum of the partons radiated in the initial state. The p_T spectrum of the softest quark is clearly lower than the p_T spectrum of the partons radiated in the initial state. This illustrates that the partons radiated in the initial state yield jets that possibly end up among the four jets with the highest transverse momentum in the jet collection of the event.

In Figure 5.17, the fraction of events that pass the p_T requirement on the four leading jets is shown for which at least one of the four leading jets is due to radiation in the initial state. This fraction increases from about 17 % for a cut on the transverse momentum of 20 GeV/c on the four leading jets to 70 % for a cut of $p_T > 100$ GeV/c. This increase with a more stringent requirement on the p_T of the four leading jets explains the decreasing purity as a function of the p_T requirement in Figure 5.15 for

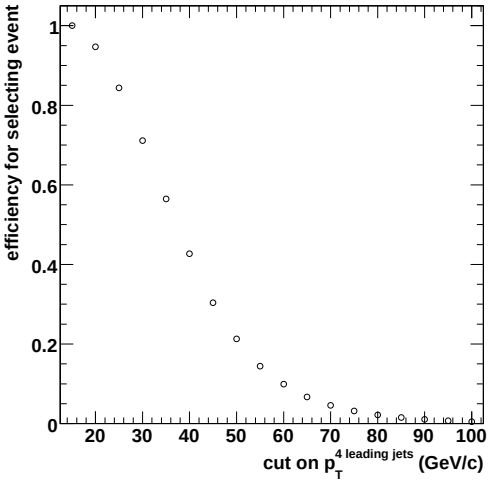


Figure 5.14: For events where the matching exists, the efficiency of selecting them after applying the p_T cut on the 4 leading jets.

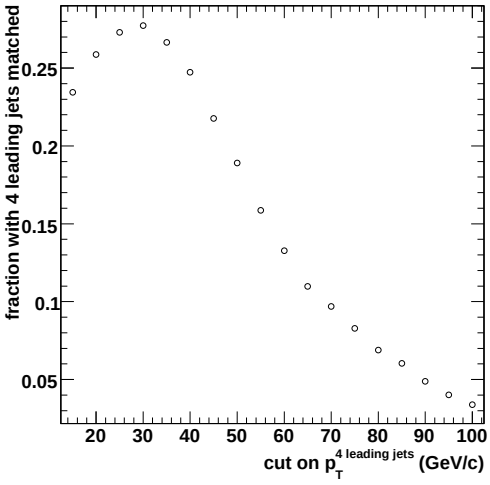
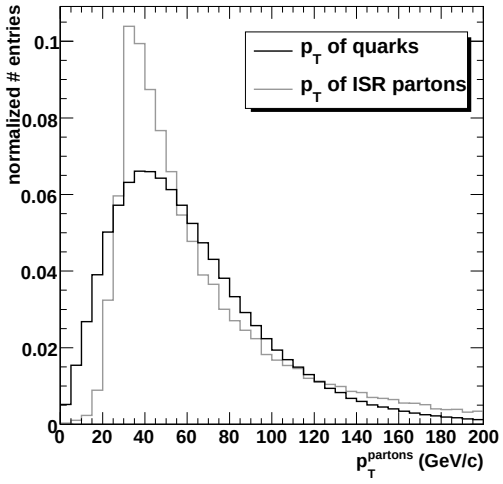


Figure 5.15: Purity, the fraction of events for which the 4 leading jets are matched to the 4 partons relative to the number of events with at least four jets above the p_T threshold.



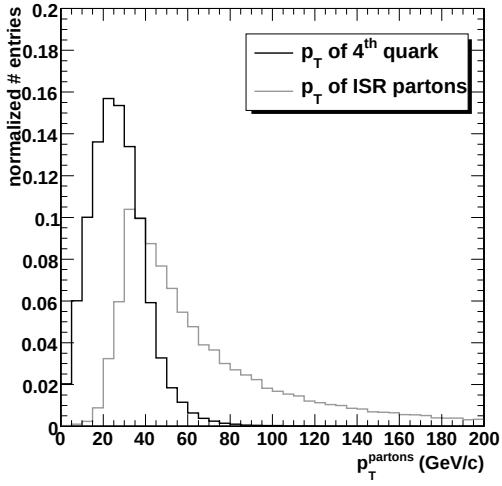


Figure 5.16: Transverse momentum of the four quarks in the semi-muonic $t\bar{t}$ final state and of the partons radiated in the initial state (left) and transverse momentum of the softest quark (right).

transverse momenta above 30 GeV/c. Other reasons for which the correct jets are not found among the four leading jets are related to final state radiation that ends up among the four leading jets or the correct jets could be lost due to acceptance criteria on η or because the jets are not correctly reconstructed. Relative to the events passing the p_T requirement on the four leading jets, Figure 5.18 shows the fraction of events for which a reason other than radiation in the initial state is the cause of the four leading jets not to be matched. This fraction is between 67 % for a cut of $p_T = 15$ GeV/c on

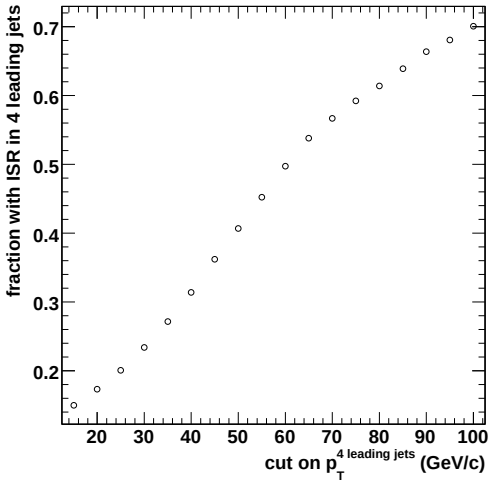


Figure 5.17: Fraction of events where at least one of the four leading jets is due to radiation in the initial state.

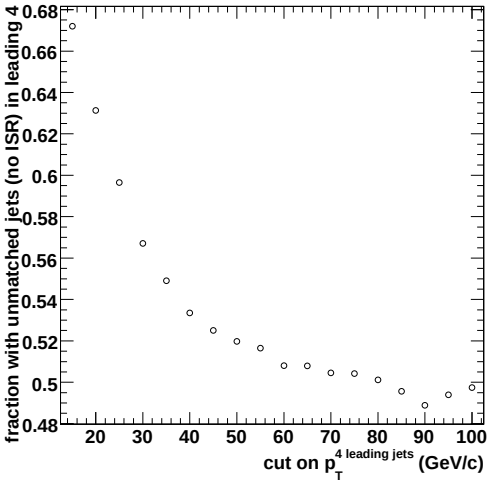


Figure 5.18: Fraction of events where not all of the four leading jets are matched, but the reason is different then radiation in the initial state.

the four leading jets and decreases to about 50 % for high transverse momenta. Since the contribution of other sources is most important in the case the four leading jets have a low transverse momenta the increasing purity in Figure 5.15 between $p_T = 15$ and $p_T = 30$ GeV/c can be explained by the decreasing contribution of jets to the four leading jets due to for example final state radiation and the η acceptance.

Both the fraction of events for which the four partons are matched (cfr. Figure 5.13) as well as the fraction of events for which the four leading jets are matched to the partons (cfr. Figure 5.15), reach a maximum when the four leading jets pass a p_T threshold of about 30 GeV/c. This clearly motivates the applied $p_T > 30$ GeV/c criterion specified in Section 5.1.

5.2.2 Jet-parton association

In the previous section, the maximum fraction of correct jet-parton matchings in preselected signal events is found to be about 28 % (cfr. purity in Figure 5.15). This value is obtained when considering the leading four jets after an event selection cut requiring their transverse momentum to exceed 30 GeV/c. A method to assign these four jets to the correct partons needs to be found. The method should be constructed in such a way that it is able to select efficiently the correct jet-parton matching in the event when

considering the four leading jets. Several methods are developed in the Top Physics Analysis Group to assign the reconstructed jets in the semi-muonic $t\bar{t}$ decay to the quarks from which they are originating. When using the four leading jets, there are 24 possible jet combinations, but this reduces to 12 because if the two jets coming from the hadronically decaying W boson are interchanged, the resulting W boson masses are identical. The methods can be simple, like for instance the method that makes use of the construction of a χ^2-value based on the reconstructed W boson and top quark mass and their resolutions,

$$\chi^2_{\text{jetcomb}} = \left(\frac{m_W^{had,reco} - m_{j1j2}}{\sigma_W^{reco}} \right)^2 + \left(\frac{m_{top}^{had,reco} - m_{bj1j2}}{\sigma_{top}^{reco}} \right)^2 , \tag{5.2}$$

where m_{j1j2} and m_{bj1j2} are the masses of the considered two-jet, W boson and three-jet, top quark candidates respectively. The jet combination that minimizes the χ^2-value is taken as the correct jet combination and the fourth jet is assigned to the b quark in the leptonic branch. Also more complex methods exist to choose the correct jet combination. A discriminant for correct and wrong jet combinations can be constructed by combining various variables that are sensitive to differentiate between correct and wrong jet combinations. The choice of a method to assign the jets to the partons depends on the physics analysis that one wants to perform. The physics studies in the next chapters make use of the true W boson and top quark masses in the hadronic decay branch, and therefore a bias can be introduced by using the reconstructed W boson and top quark mass to choose the correct jet combination in the event. Therefore, the method based on the χ^2-value described above can not be used and a different method is developed.

The differentiation power of several variables is quantified, resulting in an ordering of the variables. The most sensitive variables are combined with multi-variate analysis (MVA) techniques. To develop and train these tools, an adequate event sample needs to be used. Therefore, the effect of the choice of different event selection criteria is considered. The sensitivity of the multi-variate method for the jet-parton association to the event selection was tested by checking the performance for two different sets of event selections with looser or tighter cut values. The observables for the event selection are the ones described in Section 5.1, but for the "loose event selection" the isolation criteria were relaxed to RelIso > 0.9 for the relative isolation and respectively 6 GeV and 8 GeV for the energy deposited in the ECAL and HCAL part of the veto cone around the muon. For the "tight event selection", the cut on the transverse momentum of the muon was increased to 30 GeV/c, while for the jets a value of 40 GeV/c was used. In addition, for both sets of event selections, an extra cut was applied on the smallest angle between the four leading jets of $\Delta R > 1.0$, this corresponds to jets that are well separated.

The matching algorithm that is used to decide wether or not the considered jet combination is correct or wrong is the one described in Section 5.2.1. For the physics analyses in the next chapters only the hadronic branch in the semi-muonic $t\bar{t}$ decay is studied, therefore a jet combination is considered as correct when the three quarks from the hadronic branch are matched to three of the four leading jets.

Variables with discriminating power

About 130 variables reflecting diverse properties of the jet combinations are implemented. They make use of the information of the four-momenta of the reconstructed objects in the final state. The variables are grouped into four different categories:

1. Kinematic variables such as the mass, transverse momentum, pseudo-rapidity, azimuthal and polar angle of individual candidate objects of the $t\bar{t} \rightarrow bW\bar{b}W \rightarrow bq\bar{q}\bar{b}\mu\bar{\nu_\mu}$ final state.

2. Kinematic variables constructed from two candidate objects in the semi-muonic $t\bar{t}$ final state, such as for example the difference in mass, azimuthal angle, space angle ($\theta \times \phi$).

3. Some special variables that combine the transverse momenta of the jets, with the transverse momentum of the hadronic top quark candidate of the jet combination relative to the sum of the transverse momentum of all three-jet candidates in the event as an example.

4. Variables based on the identification of b jets using different b-tag discriminators. Either the b-tag discriminator values of the individual jet candidates may be used or the combined information of two or more jets in the final state.

In the list of variables also those related to the missing transverse energy can be found. However, as the missing transverse energy in top quark pair events might not be well understood for the early data taking at the LHC, the choice is made not to use variables related to it. The variables were sorted according to discriminating power with respect to the correct or good jet combination (meaning the jet combination for which the three quarks $t \rightarrow bW \rightarrow bq\bar{q}$ from the hadronic branch of the semi-muonic $t\bar{t}$ decay were matched) and the wrong or bad jet combinations (at least one of the three quarks $t \rightarrow bW \rightarrow bq\bar{q}$ is not matched). The distributions of the variables for correct and wrong jet combinations are first normalized to unity, while the value of the variable is transformed to a value between zero and one. Two methods are developed based on the normalized distributions to sort the variables according to discriminating power.

- **Surface method**
  The surface method calculates the overlapping surface between the normalized distributions of the correct and wrong jet combinations. Variables with a small overlapping surface are expected to have a higher discriminating power.

- **Purity-times-efficiency method**
  By considering a selection cut on the variable, the fraction of correct jet combinations below the cut value and the fraction of wrong jet combinations above the cut value can be determined. The former is the survival probability and the latter the rejection power, hence in this contect they are denoted as the efficiency and the purity respectively. The cut value can then be moved from low to high values of the variable and for each value the purity times efficiency is determined. The highest value of the purity times efficiency is chosen as a value representing

the discriminating power of the variable. Similary one can inverse the direction and calculate the efficiency as the fraction of correct jet combinations above the cut value and the purity as the fraction of wrong jet combinations below the cut value. Again for each cut value the purity times efficiency is calculated. The highest value of the purity times efficiency is again representing the discriminating power of the variable. The maximum of the two highest values obtained in both directions is used to sort the list of variables according to their discriminating power.

The variables are sorted according to the two described methods and using the two sets of event selection criteria described earlier. When combining the different variables into a multi-variate discriminator, it is important that the variables are not strongly correlated. The list is constructed using only those variables that are less then 40 % linearly correlated. The most discriminating variable is listed and all other variables that are more than 40 % correlated with the first variable are removed. In the remaining list of variables the one with the highest discriminating power is taken and the procedure is repeated until all variables are ordered or removed. Additionally, also the linear correlation with the W boson and top quark mass is required to be low as this is important to avoid a bias in the analyses in the next chapters. In the sorting of the variables it was required that the linear correlation of the variables with the W boson and top quark mass is less than 20 %.

The nine variables with the highest discriminating power in the resulting lists from the two ordering methods and the two sets of event selection criteria, contain the same information and are in most of the cases the same. Two variables are said to contain the same information if the linear correlation between them exceeds 70 %. Although the different lists contain the same variables, the order of the variables in the list is not necessarily the same. It was found that the ordering of the variables using the surface method is more robust with respect to the applied event selection compared to the purity-times-efficiency method, because the latter is more sensitive to the feature that the distributions of correct and wrong jet combinations can cross each other several times. No difference in the list of the first nine variables is found with respect to the loose and thight event selection criteria that are applied. The first three variables in the list are all related to b-tag information and although the correlation between the variables is lower than 40 % as required during the ordering procedure, only one such variable is used in the final method. Seven variables with high discriminating power were therefore used to construct a combined discriminator value. The two variables with the highest discriminating power are the product of the b-tag discriminant values of the two jets that are considered as b jets in the jet combination and the transverse momentum of the hadronic top quark candidate relative to the sum of the transverse momentum of all three-jet candidates in the event. The distributions of these two variables are shown in Figure 5.19.

The five other variables in the list are space angles between two of the various final state candidates, $\Delta\Omega(t^{\mathrm{had}}, W^{\mathrm{had}})$, $\Delta\Omega(t^{\mathrm{had}}, b^{\mathrm{had}})$, $\Delta\Omega(t^{\mathrm{had}}, b^{\mathrm{lep}})$, $\Delta\Omega(t^{\mathrm{had}}, \mu)$ and $\Delta\Omega(b^{\mathrm{lep}}, \mu)$. Their distributions for correct and wrong jet combinations are shown in Figure 5.20.

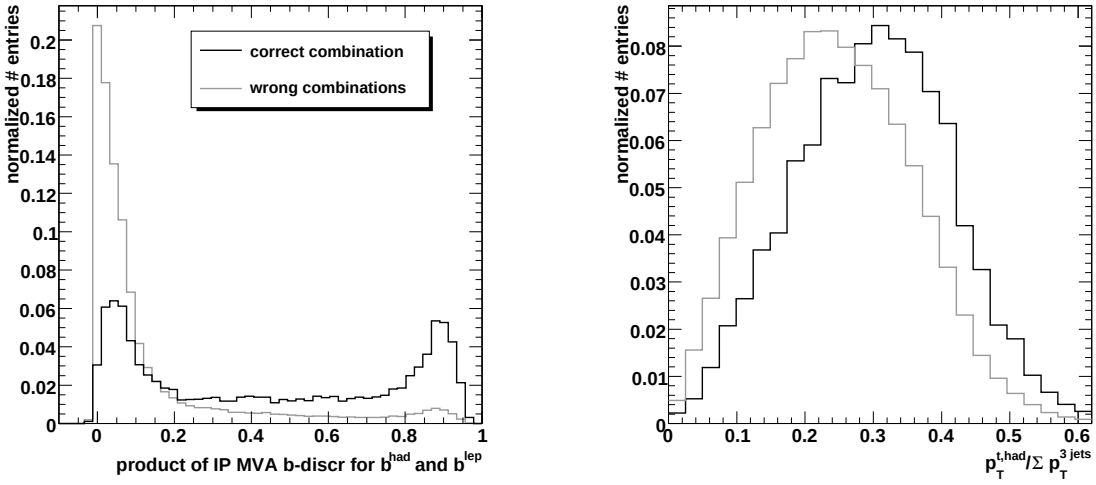


Figure 5.19: The two variables with the highest discriminating power, the product of the b-tag discriminant values of the two jets that are considered as b-jets (left) and the transverse momentum of the hadronic top quark candidate relative to the sum of the transverse momentum of all three-jet candidates (right).

Table 5.4 gives the linear correlations between the various variables. The correlations between the variables are determined for the correct and wrong jet combinations together. By construction the linear correlations between the variables are small. The largest linear correlation is between $p_T^{\mathrm{t,had}} / \sum p_T^{\mathrm{3jets}}$ and $\Delta\Omega(b^{\mathrm{lep}}, \mu)$ and is about -26%.

	b-tag	$\frac{p_T^{\mathrm{t,h}}}{\sum p_T^{\mathrm{3jets}}}$	$\Delta\Omega(t^{\mathrm{h}}, W^{\mathrm{h}})$	$\Delta\Omega(b^{\mathrm{l}}, \mu)$	$\Delta\Omega(t^{\mathrm{h}}, b^{\mathrm{l}})$	$\Delta\Omega(t^{\mathrm{h}}, b^{\mathrm{h}})$	$\Delta\Omega(t^{\mathrm{h}}, \mu)$
b-tag	100	5.0	0.9	-4.2	1.2	0.4	2.5
$\frac{p_T^{\mathrm{t,h}}}{\sum p_T^{\mathrm{3jets}}}$	5.0	100	-16.5	-25.9	-11.1	13.0	23.8
$\Delta\Omega(t^{\mathrm{h}}, W^{\mathrm{h}})$	0.9	-16.5	100	9.2	-16.3	13.8	14.3
$\Delta\Omega(b^{\mathrm{l}}, \mu)$	-4.2	-25.9	9.2	100	6.3	3.2	11.8
$\Delta\Omega(t^{\mathrm{h}}, b^{\mathrm{l}})$	1.2	-11.1	-16.3	6.3	100	8.9	9.6
$\Delta\Omega(t^{\mathrm{h}}, b^{\mathrm{h}})$	0.4	13.0	13.8	3.2	8.9	100	16.0
$\Delta\Omega(t^{\mathrm{h}}, \mu)$	2.5	23.8	14.3	11.8	9.6	16.0	100

Table 5.4: Linear correlations in % between the variables with the highest discriminating power after the "loose" set of event selection criteria are applied. The superscripts "h" and "l" indicate if the reconstructed candidate is belonging to the hadronic or leptonic branch of the event respectively. The maximal change observed in the linear correlations between the variables after applying the "thight" event selection criteria is about 5 %.

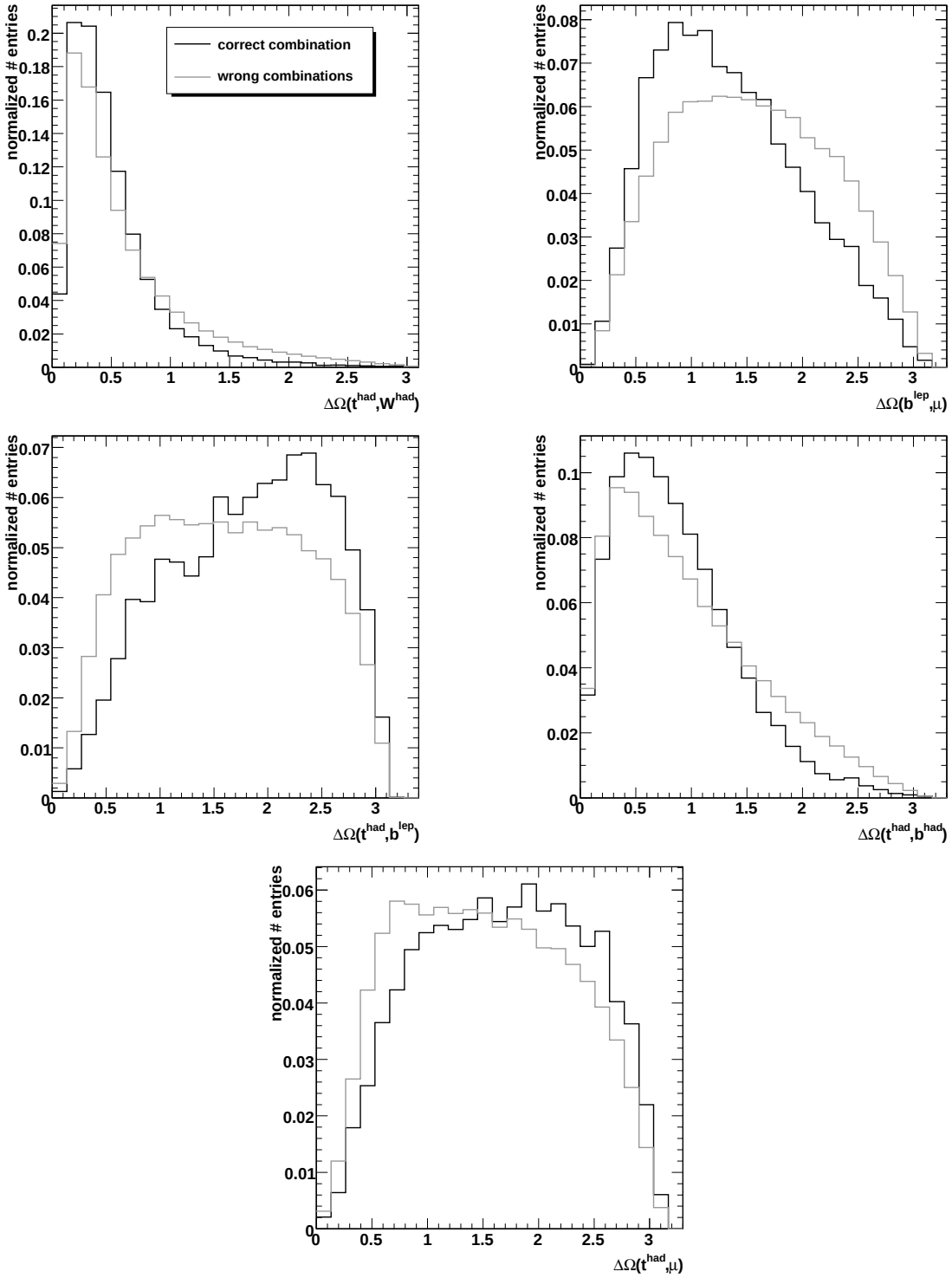


Figure 5.20: The five angular variables with high discriminating power are based on the angle in $(\theta \times \phi)$ space between the reconstructed objects in the final state.

Multi-variate discriminator for correct and wrong jet combinations

Many different methods exist to construct one single variable from the different discriminating variables. Every method needs to be trained on simulated signal, therefore correct, jet combinations and background, therefore wrong, jet combinations before it can be applied on real proton collisions. Two different methods were considered, namely a likelihood ratio method and an artificial neural network [133].

The likelihood ratio of an event is defined as the ratio, R, of the signal to signal-plus-background likelihoods, where the likelihoods are defined as the products of the corresponding probability density functions of the discriminating input variables x_i for signal S and background B, denoted by $P_i^S(x_i)$ and $P_i^B(x_i)$. In the case when seven discriminating input variables are used, the ratio is written as follows

$$R(\vec{x}) = \frac{\prod_{i=1}^{7} P_i^S(x_i)}{\prod_{i=1}^{7} P_i^S(x_i) + \prod_{i=1}^{7} P_i^B(x_i)}. \tag{5.3}$$

The training involves the construction of the functions $P_i^S(x_i)$ and $P_i^B(x_i)$ for all variables x_i. The ratio $R(\vec{x})$ is then determined for each jet combination, and the jet combination with the largest ratio $R(\vec{x})$ in the event is chosen.

The artificial neural network is a multi-layer perceptron [121], which consists of several layers. The first layer is an input layer containing as many neurons as the input variables one wants to combine. The last layer is the output layer, containing a single neuron that provides the signal (or background) weight. Between the input and output layer there is one hidden layer with the same number of neurons as the input layer. All neuron inputs to a certain layer are linear combinations of the neuron output of the previous layer. The transfer from input to output within a neuron is performed by means of a sigmoid activation function. If x_i is the variable in the input layer, then the variable in the hidden layer, x_j, is calculated as

$$x_j = A\left(\sum_{i=1}^{7} W_{ij} x_i\right) \tag{5.4}$$

with the activation function $A(x) = (1 + e^{-x})^{-1}$ and W_{ij} the weight to go from the neuron representing variable i in the input layer to the neuron j in the hidden layer. The training of the neural network uses the simulation to obtain the weights W_{ij}.

Furthermore, for both the likelihood ratio and the neural network approach the effect of a decorrelation of the input variables was tested before combining them into a single variable. The decorrelation is performed by doing a linear transformation on each of the input variables into a non-correlated variable space. The decorrelated transformed variables are then used as input variables for the likelihood ratio or neural network approach.

Using the trained files, the best jet combination is chosen as the jet combination with the highest MVA discriminant value, either the likelihood ratio value or the neural network output. With simulated events, it can be checked for how many of the events the chosen jet combination is correct or wrong. The distributions in case the chosen jet combination is correct or wrong are shown in Figure 5.21 using the likelihood ratio and neural network approach after the "loose" event selection criteria are applied.

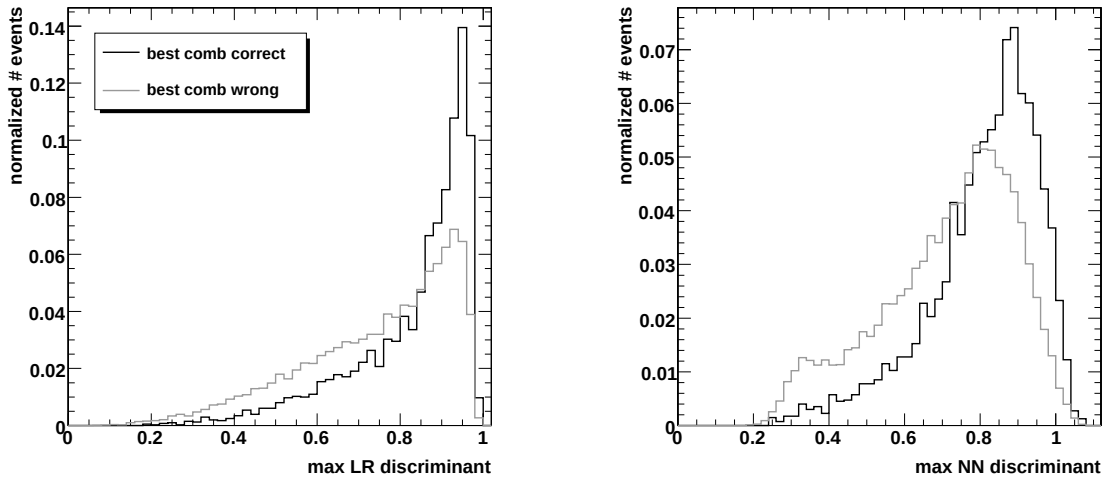


Figure 5.21: The distribution for correct and wrong jet combinations when chosing the jet combination with the highest MVA discriminant value using a likelihood ratio (LR) (left) or neural network (NN) (right) approach.

Figure 5.22 shows the highest MVA discriminant distributions in case the chosen jet combination is correct or wrong for both methods after decorrelating the input variables. The distributions look similar in case the "tight" event selection criteria are

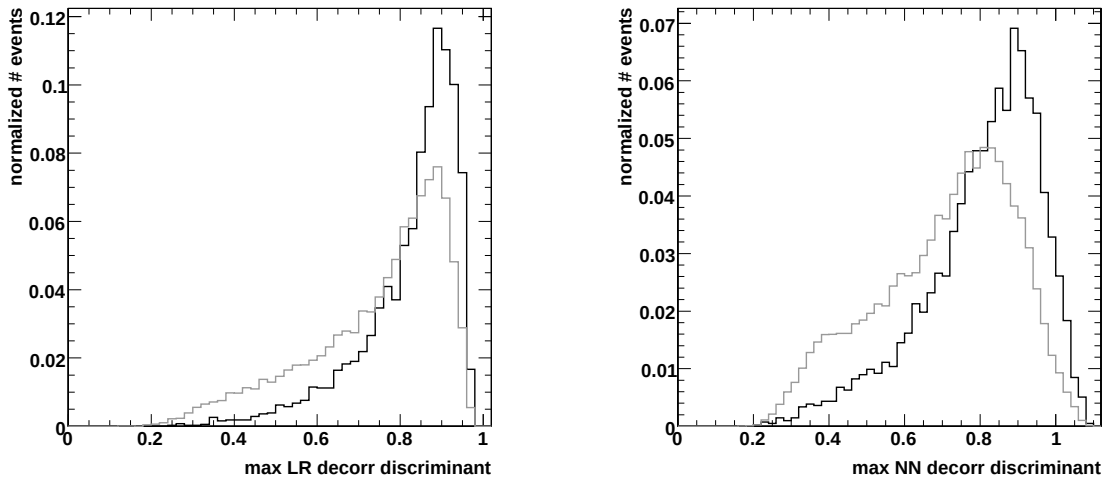


Figure 5.22: The distribution for correct and wrong jet combinations when chosing the jet combination with the highest MVA discriminant value using a likelihood ratio (LR) (left) or neural network (NN) (right) approach with decorrelated input variables.

applied.

To study the effect of the event selection on the training and the subsequent choice of the chosen jet combination, the fraction of events for which the chosen jet combination is the same using the training files with the "loose" or "tight" event selection criteria, is determined. Only the events passing the tight set of event selection criteria are considered. The resulting percentage of events is summarized in Table 5.5 for the different methods and different cut values on the discriminant value of the chosen jet

combination. The fraction of events in which the same jet combination is chosen is

cut	LR	NN	LR decorr	NN decorr
0.0	87.2	84.3	78.2	81.5
0.2	87.3	84.3	78.3	81.6
0.4	87.6	85.2	79.3	81.0
0.6	86.9	83.8	79.8	79.4
0.7	85.7	83.8	78.6	80.4
0.8	85.7	81.4	75.9	78.8

Table 5.5: The percentage for the events passing the "tight" event selection criteria, of events for which the same jet combination is chosen when using the training files for the two different sets of event selection criteria. To check the dependency of the percentage of events for which the same jet combination is chosen, on the applied threshold of the MVA discriminant value, different cuts are applied for the various MVA methods. The dependency on the applied threshold is small.

stable within a few percent with respect to the cut value on the discriminant value of the chosen jet combination. From the table it is clear that the likelihood ratio method has the highest percentage of events in which the same jet combination is selected when different training files are used. Another observation is that the decorrelation of the discriminating input variables yields results that depend more on the event selection used during the training. After decorrelating the input variables, a redefined set of variables is obtained. This new set of variables is not necessarily the same after the loose and tight event selection criteria. Therefore, it is expected that after decorrelation there is less overlap between the collection of the selected jet combinations when using the two sets of selection criteria.

The total selection efficiency after applying the "loose" event selection cuts is about 17 %. In 28 % of these events the four quarks in the semi-muonic $t\bar{t}$ decay can be matched with the four highest p_T jets, while this increases to 36 % of the events if we require that only the three quarks from the hadronically decaying branch are matched. The performance of the different methods is checked by requiring that the three quarks of the hadronically decaying branch are matched for a good jet combination. In Figure 5.23 the efficiency of the cut on the highest discriminant value in the event versus the purity is shown for the different methods and after the "loose" and "tight" event selection is applied. The different methods show a similar performance within an absolute difference of 2 % in purity at high efficiencies of the cut on the discriminant and 10 % at low efficiencies. The performance of the methods to choose the correct jet combination is the same within a few percent when applying the "loose" or "tight" event selection criteria. Since the likelihood ratio has a good performance and is robust for the choice of the jet combination with respect to the applied event selection, the likelihood ratio was used for the choice of the jet combination in the analyses in the next chapters. The

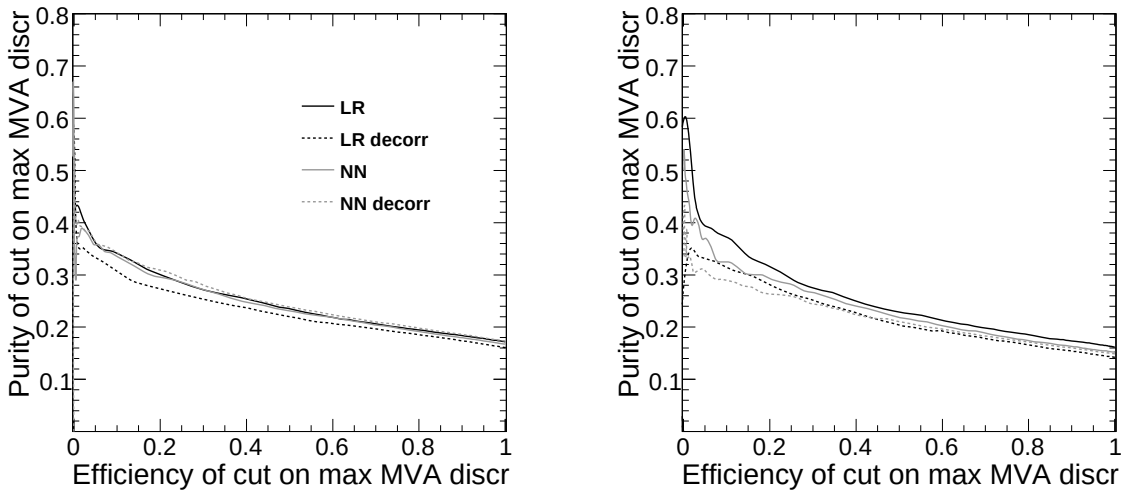


Figure 5.23: The efficiency versus the purity of the cut on the maximum MVA discriminant value for the different methods after applying the "loose" (left) and "tight" (right) event selection criteria.

chosen jet combination is the correct jet combination for 17.2 % of the events that are selected after the loose event selection, if the three quarks $t \to bW \to bq\bar{q}$ are required to match. When only considering the events where the correct matching for the three quarks of the hadronic branch exists, the chosen jet combination is correct for about 47.7 % of the events. If the four quarks are required to match, the chosen jet combination is in 15.1 % of the events correct and this fraction increases to about 53.6 % when considering only the events for which the best matching exists.

Dependency of multi-variate discriminator on the event generator

For the nominal analysis, the training is performed with the `MadGraph/MadEvent` samples. To test the dependency of the choice of the jet combination on the event generator, the training of the likelihood ratio discriminator was performed with events generated by `PYTHIA` using fast simulation. The training results are then applied on a dataset generated with `MadGraph/MadEvent`. When comparing the set of chosen jet combinations between both approaches, it is observed that for 80 % of the events the same jet combination was chosen. Using the training performed with the `PYTHIA` event sample, the obtained purity for the correct jet combination is 16.4 % when requiring that the three quarks of the hadronically decaying top quark are matched compared to 17.2 %. When requiring four matched quarks in the semi-muonic $t\bar{t}$ decay, the purity decreases to 14.4 % compared to 15.1 %. This illustrates that when the training sample and the data sample are not exactly equal, the performance of the method to choose the correct jet combination will change relatively by only 5 %. It would correspond to relatively 5 % less correct jet combinations in the selected event sample. Therefore it is concluded that the choice of the best jet combination and the performance of the likelihood ratio method does not depend strongly on the event generator that was used for the training of the discriminator.

A technique succeeds in mathematical physics, not by a clever trick, or a happy accident, but because it expresses some aspect of a physical truth.

O.G. Sutton

Chapter 6

Inclusive estimation of the jet energy corrections

Physics studies in particle collisions and precision measurements based on reconstructed jets require jet energies which are correctly calibrated. Jet energy correction factors are either derived from simulated proton collisions or using the available data. While the correction factors obtained from simulated proton collisions cope with possible large systematic uncertainties due to the hypotheses in the simulation, the correction factors obtained with the early data can have large statistical uncertainties. Methods were developed to determine corrections for the pseudo-rapidity and transverse momentum dependence of the jet energy using the early proton collisions. In addition to those factors, many physics analyses require jet energy correction factors that correct for the dependence on the flavour of the parton from which the jet originates. Both the validation of the applied jet energy correction factors as well as the estimation of flavour dependent correction factors are possible using the W boson and top quark masses, which were precisely measured in previous experiments. The $pp \to t\bar{t} \to bW\bar{b}\bar{W} \to bq\bar{q}\bar{b}\mu\bar{\nu_\mu}$ process is used for that purpose. In Chapter 5, event selection criteria were discussed and methods were studied to assign the reconstructed jets in the event to the partonic decay topology of the top quark pair event from which they originate. At this stage, estimators are developed to determine the correction factors for the jet energies.

The jets that are used as input for the analysis presented in this thesis are jets reconstructed with the Seedless Infrared Safe Cone algorithm with an opening angle of 0.5 and corrected for the pseudo-rapidity and transverse momentum dependence, i.e. the L2 and L3 jet energy corrections are applied. The event selection criteria summarized in Section 5.1.3 are applied, with the exception of the trigger criterion. The likelihood ratio method is used to choose the jet combination as the one with the highest MVA discriminant value. In Section 6.1 the chosen jet combination is used to reconstruct the W boson and top quark mass spectra. The effect of applying a cut on the MVA discriminant value of the chosen jet combination on the reconstructed W boson and top quark masses is shown. Furthermore, the need for measuring residual jet energy correction factors beyond the initial L2 and L3 corrections is illustrated. A technique to incorporate hypothesized constraints in the observed event topology

by means of a kinematic fit using Lagrange Multipliers is introduced in Section 6.2. Additionally, the performance of the fitting technique is illustrated with an example. Using the kinematic fitting technique with Lagrange Multipliers to incorporate the W boson and top quark mass constraints, an event likelihood as a function of the applied jet energy correction factors is constructed as explained in Section 6.3. Section 6.4 elaborates on the estimator developed for the measurement of the inclusive jet energy correction factors for light quark jets from the W boson decay and b quark jets, while the statistical properties of the estimator are reviewed in Section 6.5. Robustness checks and contributions to possible systematic uncertainties on the measurement are discussed in Section 6.6. Finally, in Section 6.7 the method to estimate the inclusive jet energy corrections was applied on jets reconstructed using another jet reconstruction algorithm.

6.1 Reconstructed W boson and top quark mass

Using the chosen jet combination, the four-momenta of the jets are combined to reconstruct the four-momenta of the W boson and the top quark in the hadronic branch of the semi-muonic $t\bar{t}$ decay. The reconstructed mass spectra of the W boson and top quark are shown in Figure 6.1. In case the two jets in the chosen jet combination assigned to the W boson are correctly matched to the quarks of the W boson, the W boson mass spectrum is more peaked. Similarly, when the three jets in the chosen jet combination assigned to the top quark are correctly matched to the quarks in the $t \to bW \to bq\bar{q}$ decay, the reconstructed top quark mass spectrum is more peaked.

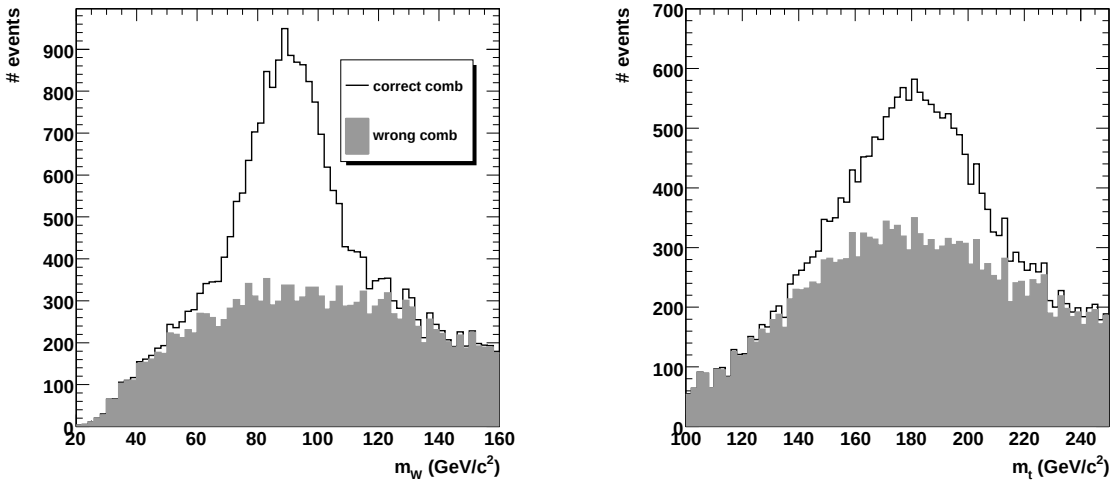


Figure 6.1: The reconstructed W boson mass (left) and the top quark mass (right) spectra in the case the chosen jet combination is correct or wrong.

Figure 6.2 shows the reconstructed top quark mass versus the reconstructed W boson mass for the events for which the chosen jet combination is correct. The distribution is centered around a W boson mass of about 90 GeV/c^2 and a top quark mass of about 185 GeV/c^2. The events were generated using a W boson and top quark mass of respectively 80.4 and 170.9 GeV/c^2. The discrepancy between the values of the recon-

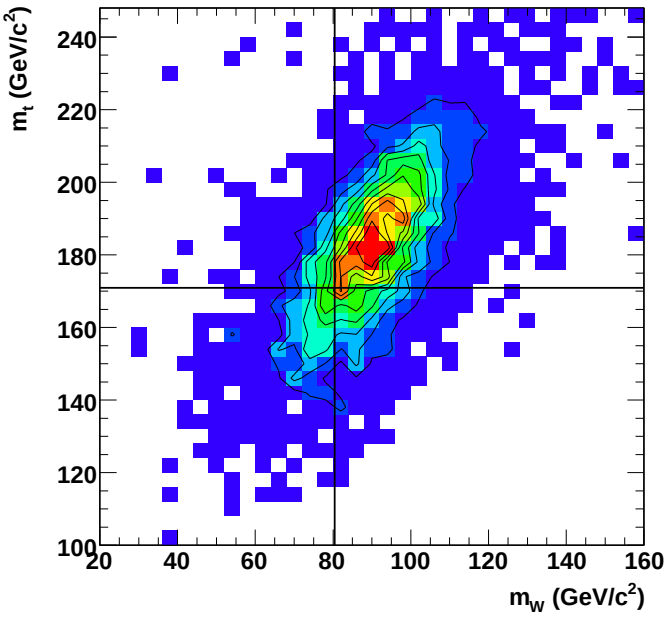


Figure 6.2: Reconstructed W boson mass versus the reconstructed top quark mass using the jets that were matched to the quarks in the hadronic branch of the semi-muonic $t\bar{t}$ decay.

structed W boson and top quark masses and the values that were used to generate the events is related to the presence of non-zero residual jet energy corrections beyond the initial L2 and L3 corrections. While it is possible to improve slightly the resolution of the reconstructed W boson and top quark masses by applying a cut on the value of the MVA discriminant of the chosen jet combination, the position of the peak in the mass spectra will not shift to the correct value by applying this event selection cut. This is shown in Figure 6.3 for two different cuts on the value of the MVA discriminant. The difference between the reconstructed and generated W boson and top quark masses illustrates the need for residual jet energy correction factors. In the next section, a kinematic fit technique is summarized on which the construction of an estimator for the residual jet energy correction factors is based.

6.2 Kinematic fit of the event topology with mass constraints

A kinematic hypothesis for an event, like for instance a mass constraint, can be imposed on the measured four-momenta of the objects in the final state by making use of a non-linear least square fit using Lagrange Multipliers. During the procedure, or so-called kinematic fit, the measured four-momenta are slightly altered, taking into account the uncertainties on the measured four-momenta, to fulfill the imposed kinematic hypothesis or constraint. The kinematic fit returns a probability for each event, reflecting the probability that the imposed hypothesis is true for the event. Section 6.2.1 elaborates

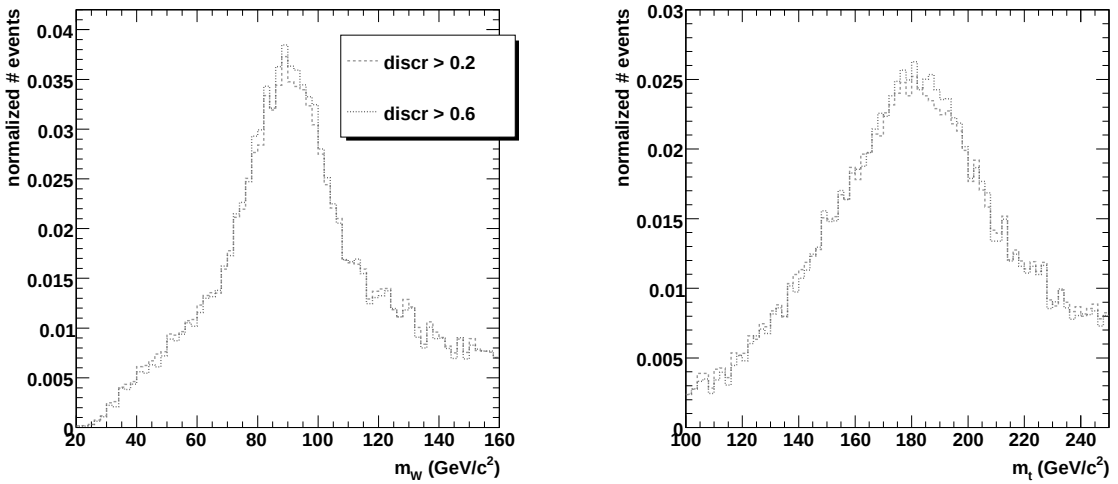


Figure 6.3: The reconstructed W boson mass (left) and top quark mass (right) spectra using the chosen jet combination after different cuts on the MVA discriminant value of the chosen jet combination.

on the basic framework of kinematic fitting needed for this thesis. The performance of the kinematic fit on the resolution of the reconstructed top quark mass is illustrated in Section 6.2.2.

6.2.1 Fit of an event topology with kinematic constraints

A more general framework of the kinematic fit is presented in [134], while in this section only the concepts and formulas needed for the analysis presented in this thesis are introduced. Consider the situation in which one has n measured parameters $\vec{y}$ and one would like to solve the problem in which m constraints $\vec{f}$ are fulfilled for the true parameters $\vec{y}_t$. The constraints $\vec{f}$ are valid for the true parameters and are defined as

$$\begin{aligned} f_1(\vec{y}_{t,1}, ..., \vec{y}_{t,n}) &= 0 \\ &\vdots \\ f_m(\vec{y}_{t,1}, ..., \vec{y}_{t,n}) &= 0 \end{aligned} \tag{6.1}$$

Generally, the constraints are not fulfilled by the measured parameters $\vec{y}$ and corrections to the measured parameters $\Delta\vec{y}$ need to be calculated such that the sum $\vec{y}' = \vec{y} + \Delta\vec{y}$ fulfills the constraints. Additionally the weighted sum

$$S(\vec{y}) = \Delta\vec{y}^T \mathcal{V}^{-1} \Delta\vec{y} \tag{6.2}$$

should be minimal, where $\mathcal{V}$ is the covariance matrix of the measured parameters. A likelihood can be constructed using the Lagrange Multipliers $\vec{\lambda}$ as follows

$$L(\vec{y}, \vec{\lambda}) = S(\vec{y}) + 2 \sum_{k=1}^{m} \lambda_k f_k(\vec{y}). \tag{6.3}$$

Using the definition of the likelihood, a local minimum is found for a minimum of $S(\vec{y})$ under the constraint $f_k(\vec{y}) = 0$. When non-linear constraints are incorporated, as is

the case for mass constraints, the solution is found in an iterative way. The constraints are linearized in every iteration by means of a Taylor expansion

$$f_k(\vec{y}') \approx f_k(\vec{y}^*) + \sum_{i=1}^{n} \frac{\partial f_k}{\partial y_i}(\Delta y_i - \Delta y_i^*) \approx 0, \tag{6.4}$$

where $\vec{y}$, $\vec{y}^*$ and $\vec{y}'$ are respectively the start value, the value after the last iteration and the value after the current iteration of the measured parameters. The $\Delta\vec{y}$ and $\Delta\vec{y}^*$ are defined as respectively $\Delta\vec{y} = \vec{y}' - \vec{y}$ and $\Delta\vec{y}^* = \vec{y}^* - \vec{y}$. Thus the partial derivatives of the constraints with respect to the measured parameters need to be calculated at the value of the parameters after the last iteration $\vec{y}^*$. Since the four-momenta ($\vec{P} = (E, \vec{p})$) of the measured objects are used as input for the kinematic fit, the chain rule is applied to obtain the partial derivatives as $\frac{\partial \vec{f}}{\partial \vec{y}} = \frac{\partial \vec{f}}{\partial \vec{P}} . \frac{\partial \vec{P}}{\partial \vec{y}}$. The variances of the fitted parameters $\vec{y}'$ are obtained by propagating the uncertainties using the covariance matrix of the measured parameters. The linearization is repeated after each iteration step until some predefined convergence criteria are fulfilled or until a given maximum number of iterations is reached. The first convergence criterion requires that the change of the value S of the χ^2 expression from one iteration to the next, is smaller than a value ϵ_S. When n denotes the current iteration and $n-1$ the previous iteration, the criterion is the following

$$\frac{S(n-1) - S(n)}{ndf} < \epsilon_S \tag{6.5}$$

where the number of degrees of freedom ndf, is equal to the number of constraints. A second convergence criterion requires that the constraints are fulfilled better than a given value ϵ_F

$$\sum_{k=1}^{m} |f_k^{(n)}(\vec{y})| < \epsilon_F. \tag{6.6}$$

As shown in the next section by means of a simple example, high convergence rates are achieved. The four-momenta of the reconstructed objects can be parametrized in different ways. The convergence rate is studied for several of these parametrizations.

6.2.2 Performance of the kinematic fit

The performance of the kinematic fit is illustrated in this section by showing the improvement in the resolution of the reconstructed top quark mass spectrum after the application of the W mass constraint on the two jets associated with the quarks originating from the hadronic decay of the W boson. Therefore, the four-momenta of these jets are parametrized and the uncertainties on the measured parameters are determined.

Several options are available in the kinematic fit package for the parametrization of the four-momenta of the measured objects in the final state. Four parametrizations are considered in this section of which two are commonly used at hadron colliders, namely (E_T, θ, ϕ) and (E_T, η, ϕ). For the (E_T, θ, ϕ) parametrization, the four-momentum of the object in the final state is parametrized as

$$\vec{p}_f = \begin{pmatrix} E_T \cos\phi \\ E_T \sin\phi \\ E_T \cot\theta \end{pmatrix} \qquad \text{and} \qquad E_f = \frac{E_T}{\sin\theta}, \tag{6.7}$$

while for the (E_T, η, ϕ) parametrization the four-momentum is written as

$$\vec{p}_f = \begin{pmatrix} E_T \cos\phi \\ E_T \sin\phi \\ E_T \sinh\eta \end{pmatrix} \qquad \text{and} \qquad E_f = \frac{E_T}{\cosh\eta}. \tag{6.8}$$

In these two parametrizations the objects are considered as massless. The (a, b, c, d) parametrization is a four-momentum parametrization that describes the deviation from the measured momentum in a local coordinate system

$$\vec{u}_1 = \frac{\vec{p}_m}{|\vec{p}_m|}, \qquad \vec{u}_2 = \frac{\vec{u}_3 \times \vec{u}_1}{|\vec{u}_3 \times \vec{u}_1|}, \qquad \vec{u}_3 = \frac{\vec{u}_z \times \vec{u}_1}{|\vec{u}_z \times \vec{u}_1|}, \tag{6.9}$$

where $\vec{u}_z$ is the unit vector in the direction of the z-axis. In this way the fitted momentum is parametrized with three parameters a, b and c

$$\vec{p}_f = a|\vec{p}_m|\vec{u}_1 + b\vec{u}_2 + c\vec{u}_3. \tag{6.10}$$

while the energy is parametrized with a fourth parameter d as

$$E_f = \sqrt{|\vec{p}_f|^2 + d^2 m^2}, \tag{6.11}$$

where m is the mass of the object. This energy parametrization implies that the fitted object has a free floating mass. A fourth parametrization, based on the (E_T, θ, ϕ) parametrization was implemented with the additional requirement

$$\frac{E_f}{|\vec{p}_f|} = \frac{E_m}{|\vec{p}_m|}, \tag{6.12}$$

which implies that the fitted ratio of the energy to the momentum is kept constant to the measured value.

The uncertainties on (E_T, θ, ϕ) are given by the resolutions in Chapter 4 as a function of the transverse momentum and pseudo-rapidity of the jet. Similarly the resolutions for the parameters in other parametrizations were determined.

The jets associated with the quarks from the hadronic decay of the W boson are used in the event-by-event kinematic fit described in Section 6.2.1 with the measured world average of the W boson mass as a constraint. In the absence of real collision data this is taken to be the value used in the event simulation. The convergence rates using the four different parametrizations described above are summarized in Table 6.1. The convergence rate is nearly 100 % in case the jets were matched to the quarks of the W boson. In case a wrong jet combination was used, the convergence rate varies between 99 % and about 80 % depending on the parametrization that was used. The number of iterations needed for the kinematic fit to convergence is shown in Figure 6.4 for the four considered parametrizations. The default maximum number of iterations in the kinematic fit is conservatively set to 200, but as can be seen from the plot, this is clearly too large because the number of iterations for the considered parametrizations is well below 25. For all parametrizations the number of iterations needed in the kinematic

	(E_T, θ, ϕ), $E/\|p\|$ fix	(E_T, θ, ϕ)	(E_T, η, ϕ)	(a, b, c, d)
correct combination	99.99	100.00	99.93	99.84
wrong combination	98.98	98.95	91.24	79.86

Table 6.1: Convergence rates of the kinematic fit in % for the different parametrizations.

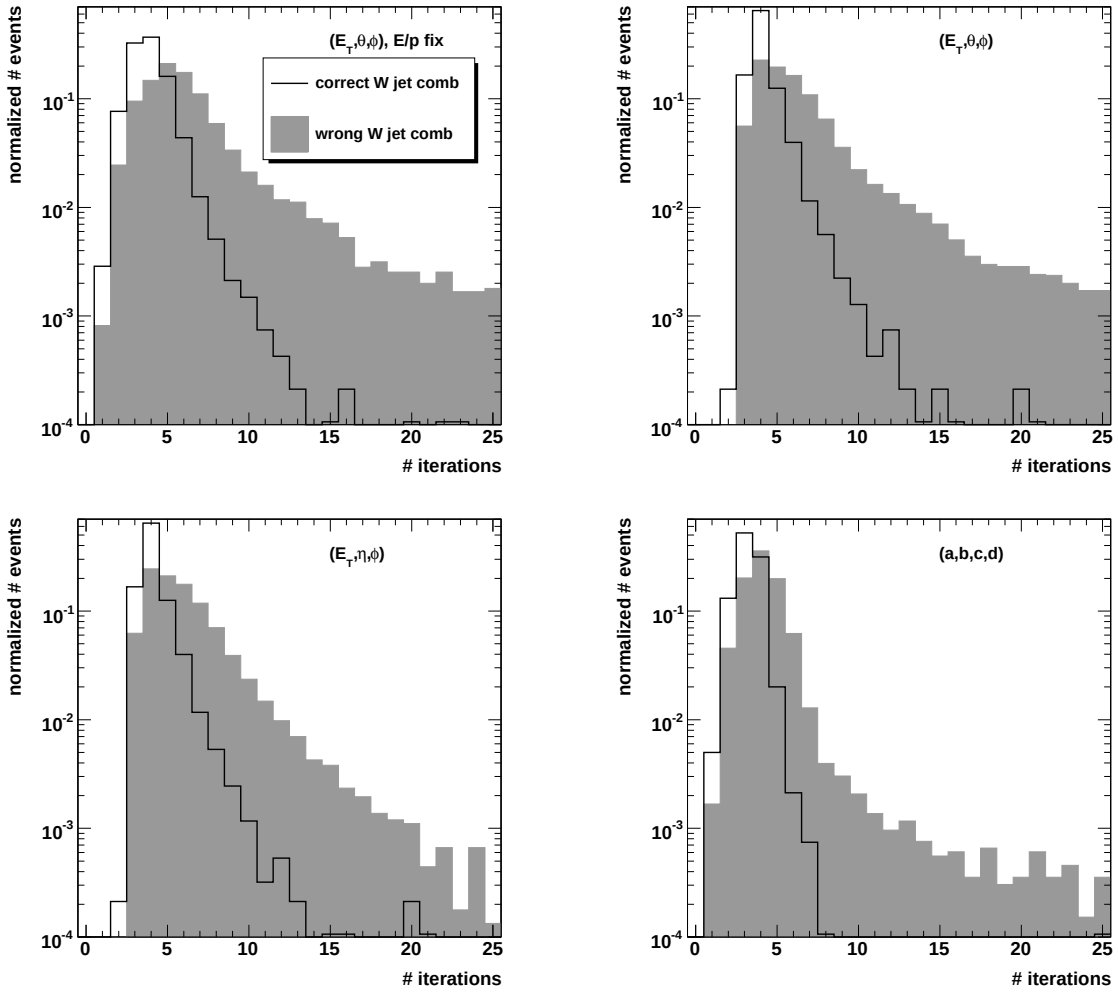


Figure 6.4: The number of iterations needed for the kinematic fit to converge using the chosen jet combination as input for four different parametrizations.

fit to convergence is higher in case the chosen jet combination is wrong. Figure 6.5 shows the number of iterations before convergence is reached in case the chosen jet combination is correct. The (a, b, c, d) parametrization converges slightly faster than the other three. The (E_T, θ, ϕ) and (E_T, η, ϕ) have a similar distribution, while the (E_T, θ, ϕ) parametrization for which $E/|p|$ is fixed to the measured value has a broader distribution.

As an illustration, the fitted four-momenta of the jets associated with the quarks of the W boson were used to reconstruct the top quark mass. The (E_T, θ, ϕ) parametrization for which $E/|p|$ is fixed, was used in the kinematic fit. Figure 6.6 shows the reconstructed top quark mass spectrum before and after the kinematic fit was applied.

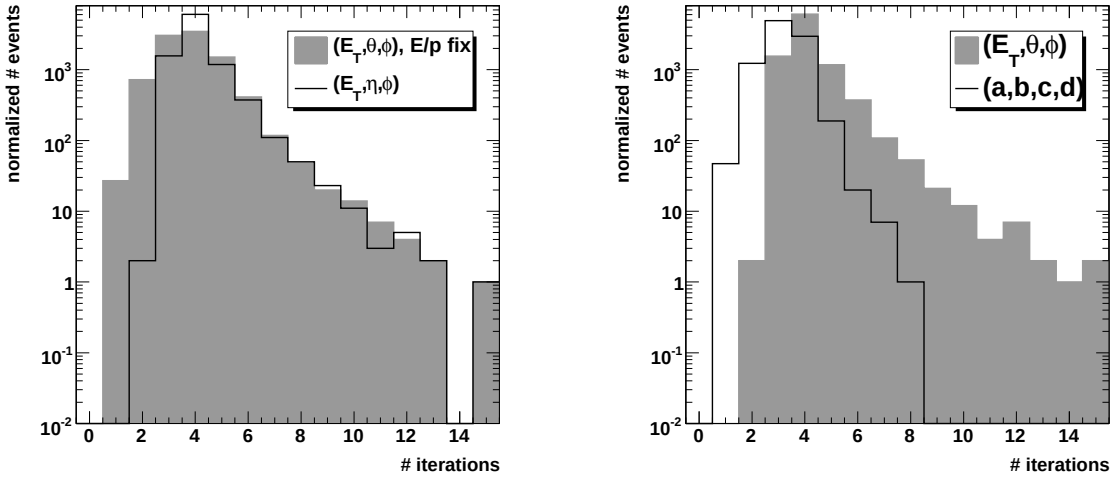


Figure 6.5: Number of iterations needed for the kinematic fit to converge in case the chosen jet combination is correct and using four different parametrizations.

The contribution of the wrong jet combinations is also shown, the chosen jet combina-

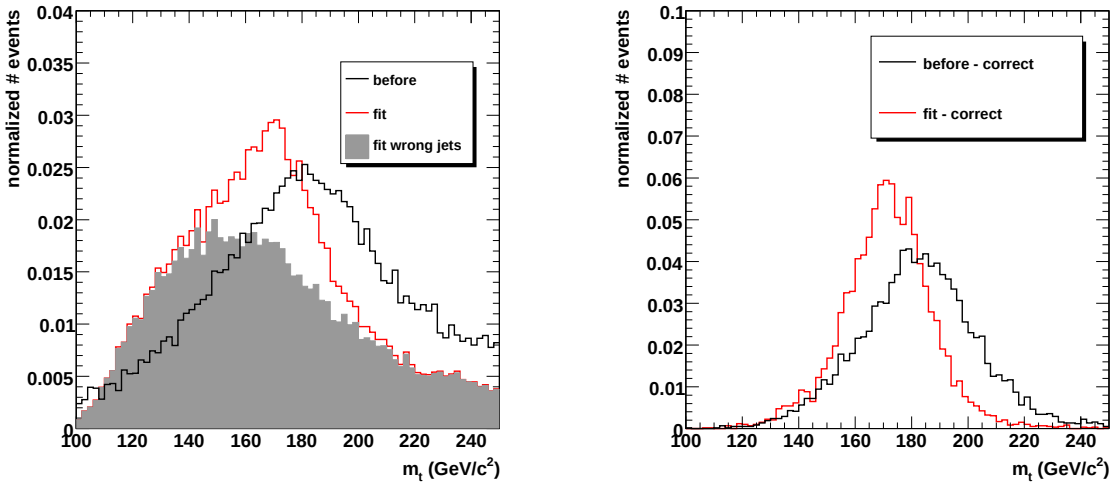


Figure 6.6: The reconstructed top quark mass spectrum before and after the kinematic fit was applied using the chosen jet combination (left) and using only the chosen jet combination if it is correct (right).

tion is considered as wrong if at least one of the three jets in the chosen jet combination is not matched to the quarks in the hadronic branch of the semi-muonic $t\bar{t}$ decay. A clear improvement after the kinematic fit is visible for both the resolution of the distribution as for the peak value, which is closer to the input value used to simulate to $t\bar{t}$ events.

6.3 The event likelihood

In Section 6.1 it was explained that the shift on the reconstructed W boson and top quark masses is due to residual jet energy corrections. The aim of this chapter is to discuss how the residual jet energy corrections for both the light quark jets from the W boson and the bottom quark jets are estimated by applying an event-by-event kinematic fit using the world average of the W boson and top quark masses as constraints on the hadronic branch of the semi-muonic $t\bar{t}$ decay. To demonstrate the method on simulated events, these masses are taken to be equal to the values used in the simulation.

The event selection as discussed in Section 5.1.3 is applied with the exception of the trigger criteria. The transverse momenta of the three jets from the hadronic branch of the semi-muonic top quark pair decay are required to be below 200 GeV/c to avoid aspects specific to high p_T jets. Additionally, to avoid a possible bias on the estimation of the jet energy corrections a cut is applied on the smallest angle between the jets with $p_T > 15\,\text{GeV/c}$ in the event. Figure 6.7 shows the smallest ΔR angle in $(\eta \times \phi)$ space between the jets in the event versus the energy that is in the overlap region. This overlap energy is calculated by summing the energy in the towers of jet i for the towers that are closer than $\Delta R < 0.5$ from the axis of jet j. Even in the case the jets are very

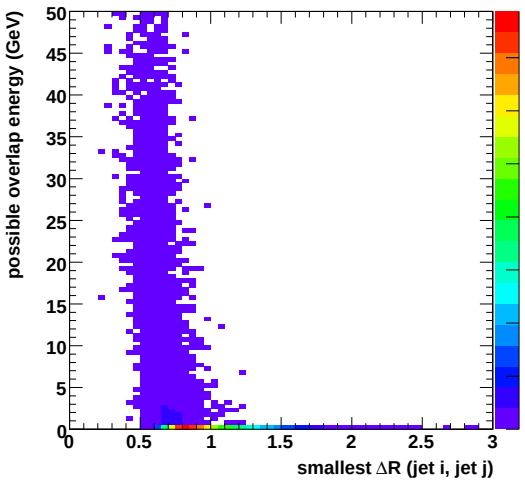


Figure 6.7: Energy in the overlap region between the two jets that are closest in the selected semi-muonic top quark pair decay events versus the smallest ΔR angle between them.

close to each other, the overlap energy is zero for most events. This indicates that there are not so many events for which the jets could have shared an significant amount of energy. The possible energy sharing is important, because if two jets could have shared energy, it would introduce a bias on the jet energy corrections to be estimated. A loose cut is applied on the two variables by requiring that the smallest ΔR angle between the jets in the event is at least 0.8 or, in case the smallest ΔR is smaller than 0.8, that the overlap energy is zero.

After applying these extra event selection cuts, the three jets from the $t \to bW \to bq\bar{q}$ decay, as given by the jet combination with the highest MVA discriminant value, are

used in the kinematic fit that was described in Section 6.2. Only the events for which the chosen jet combination has an MVA discriminant value higher than 0.2 are taken into account. The event selection efficiencies after these cuts are shown in Table 6.2.

	$t\bar{t} \to \mu$	$t\bar{t} \to$other	$t(\bar{t})$	W+jets	Z+jets
# events	3.07 k	17.6 k	3.61 k	2.28 M	211 k
selection Section 5.1.3	$24.3\,10^{-2}$	$1.08\,10^{-2}$	$1.12\,10^{-2}$	$1.33\,10^{-4}$	$1.97\,10^{-4}$
p_T(3jets) $< 200\,\mathrm{GeV/c}$	$22.4\,10^{-2}$	$9.55\,10^{-3}$	$4.10\,10^{-3}$	$4.34\,10^{-5}$	$1.70\,10^{-4}$
$\Delta R > 0.8$ or $E_{\mathrm{overlap}} = 0$	$16.3\,10^{-2}$	$6.96\,10^{-3}$	$3.06\,10^{-3}$	$3.01\,10^{-5}$	$1.21\,10^{-4}$
MVA discr > 0.2	$16.3\,10^{-2}$	$6.94\,10^{-3}$	$3.03\,10^{-3}$	$2.77\,10^{-5}$	$1.12\,10^{-4}$
P^{fit} cuts	$5.66\,10^{-2}$	$1.46\,10^{-3}$	$6.54\,10^{-4}$	$4.28\,10^{-6}$	$1.49\,10^{-5}$
# events ($50\,\mathrm{pb}^{-1}$)	173.7±1.9	25.71±0.75	6.02±0.22	24.27±0.77	3.14±0.29

Table 6.2: Overview of efficiencies obtained after the event selection cuts applied to the signal and background event samples. The indicated number of events before and after the cuts are rescaled to an integrated luminosity of $\mathcal{L}$ =50 pb^{-1}.

For the estimation of the residual jet energy corrections, it is essential to keep the energy over momentum ratio $E/|p|$ fixed to the measured value. As a result, the (E_T, θ, ϕ) parametrization for which $E/|p|$ is fixed, is used to parametrize the measured four-momenta in the kinematic fit. The kinematic fit provides a probability obtained from the χ^2 value of the fit. This probability reflects how likely the reconstructed event in the chosen jet combination agrees with the hypothesis of the world average empirical determinations of the top quark and W boson masses. In the study presented in this thesis, these mass values are taken to be equal to the mass values used during the event simulation as already mentioned before. The jet energy scales of the jets in the $t \to bW \to bq\bar{q}$ decay are then adapted to maximize the probability returned by the kinematic fit. This was obtained by determining this χ^2 probability, P^{fit}, in the dimensions of the three residual jet energy corrections, resulting in $P^{\mathrm{fit}} = P^{\mathrm{fit}}(\Delta E_b, \Delta E_q, \Delta E_{\bar{q}})$ [1]. The energies, and accordingly the momenta of the jets to keep the $E/|p|$ constant, are shifted in steps of 2 % between -40 % and +40 % for both the light quark jets from the W boson decay and the bottom quark jets.

[1] The notation ΔE is a short notation for $\Delta E/E$, indicating the relative jet energy correction, hence quoted in %.

For each event and at each point in this three dimensional space of $(\Delta E_b, \Delta E_q, \Delta E_{\bar{q}})$ the kinematic fit was applied. Hence for each selected event the kinematic fit is repeated $41^3 = 68921$ times, creating a CPU intensive analysis. Therefore, this part of the analysis is only performed after the event selection described in Section 6.1. Because the aim is an inclusive estimation of the residual energy corrections, the energy corrections for the two jets from the W boson decay are required to be equal, thus $\Delta E_q = \Delta E_{\bar{q}} = \Delta E_l$. For each event we thus have a two dimensional probability or likelihood $P^{\text{fit}}(\Delta E_b, \Delta E_q = \Delta E_{\bar{q}}) = P^{\text{fit}}(\Delta E_b, \Delta E_l)$ in the dimensions of the residual b and and light quark jet energy corrections. In Figure 6.8 the likelihood is shown for a signal and a background event.

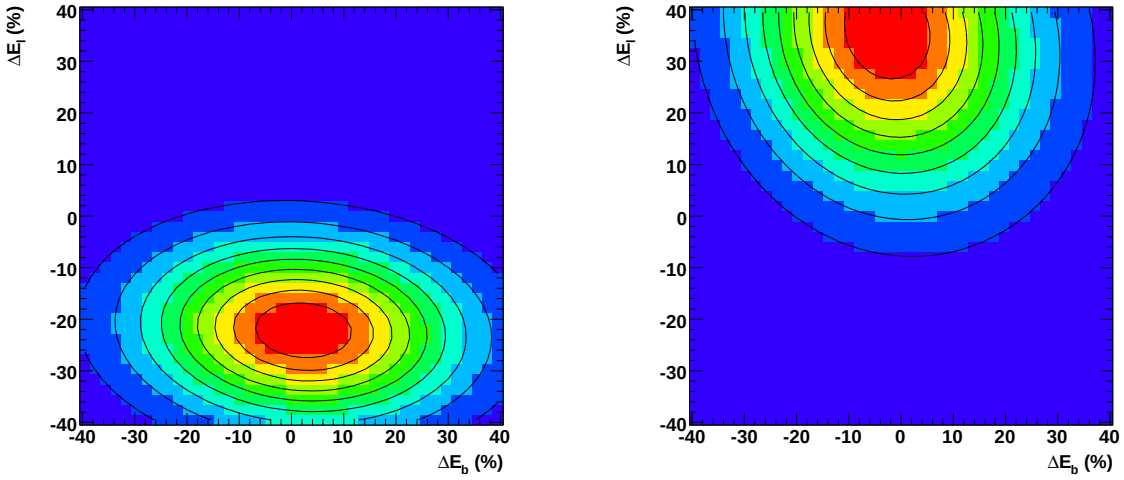


Figure 6.8: Two dimensional likelihood $P^{\text{fit}}(\Delta E_b, \Delta E_q)$, for a signal (left) and a background (right) event.

An extra selection criterion is defined by taking an event only into account when it has in the range $-40\,\% < \Delta E_i < +40\,\%$, where i indicates a light or b quark jet, a maximal fit probability exceeding 0.98. The distribution of the maximal fit probability is shown in Figure 6.9. The criterion corresponds to the requirement that the jet energy corrections to fulfill the mass constraints in the event were found in the region that is scanned. Additionally, events for which the maximal fit probability is reached at -40 % or +40 % are removed because the edges of the scanned range tend to pile up events for which the maximal fit probability is outside the range. Another selection criterion aims to remove the events for which no correct parton matching exists. This is obtained by selecting only events with the fit probability when applying no additional corrections on the jet energy scale is above 0.02 or $P^{\text{fit}} = P^{\text{fit}}(\Delta E_b = 0, \Delta E_l = 0) > 0.02$. If the initial jet energy scales are correct and therefore no bias is observed in the reconstructed top quark and W boson masses, this variable should be uniformly distributed. As is shown in Figure 6.9 this is clearly not the case, the peak at zero reflects events that are wrongly reconstructed or events in which a gluon was radiated from the hard partons in the top quark decay. The cumulative efficiency of these extra selection criteria is quoted in Table 6.2.

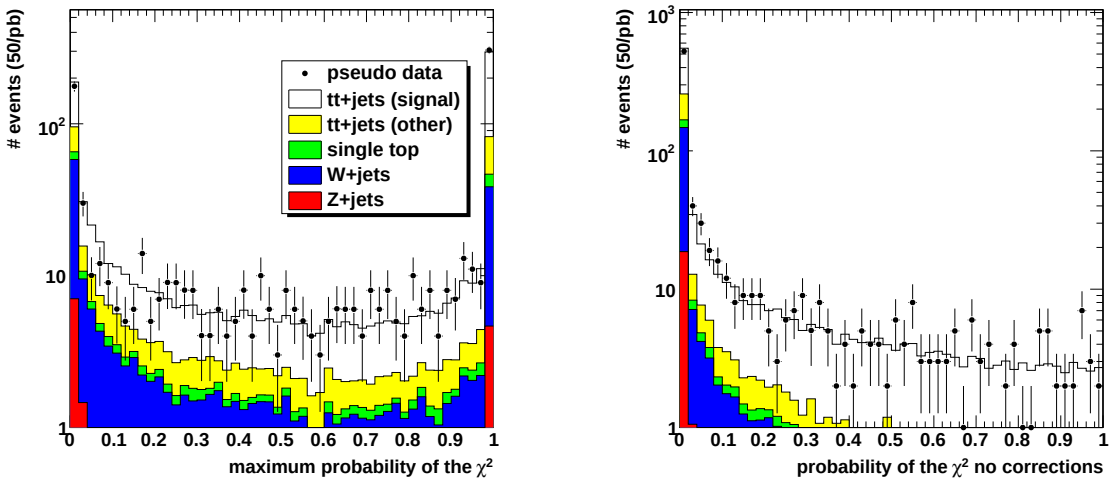


Figure 6.9: Maximum probability $P^{\text{fit}}(\Delta E_b, \Delta E_l)$ returned by the kinematic fit in an event (left) and probability $P^{\text{fit}}(\Delta E_b = 0, \Delta E_l = 0)$ when no energy correction factors are applied (right).

6.4 Estimation of the jet energy corrections

The jet energy corrections can now be estimated by combining the $\chi^2(\Delta E_b, \Delta E_l)$ distributions of the selected events. The $P^{\text{fit}}(\Delta E_b, \Delta E_l)$ probabilities returned by the kinematic fit are thus first transformed into the corresponding $\chi^2(\Delta E_b, \Delta E_l)$ values. Numerically it is more stable to combine the $\chi^2(\Delta E_b, \Delta E_l)$ values than the probabilities, because a sum can be used to combine the $\chi^2(\Delta E_b, \Delta E_l)$ values, while a product would be needed to combine the possibly small values of the probabilities. The estimated jet energy corrections correspond to the values of ΔE_b and ΔE_l where the combined $\chi^2(\Delta E_b, \Delta E_l)$ distribution reaches a minimum.

However, to make the method more robust, an outlier rejection is applied before combining the $\chi^2(\Delta E_b, \Delta E_l)$ distributions of the selected events. Therefore, the Gaussian shape of the probability distribution $P^{\text{fit}}(\Delta E_b, \Delta E_l)$ is used. If a Gaussian distribution is considered and the second derivative is calculated, one would see that in the central 68 % probability interval around the most probable value of the Gaussian, the second derivative is negative and outside this interval it is positive. Thus, for each event and in each point $(\Delta E_b, \Delta E_l)$ of the applied jet energy corrections, the second derivative to the b quark jet energy correction and the second derivative to the light quark jet energy correction are calculated. A new distribution is constructed by combining the information from all events. Every event provides an entry for each point in the two dimensional space $(\Delta E_b, \Delta E_l)$ of the applied jet energy correction shifts where the following relations hold

$$\frac{\partial^2(P^{\text{fit}}(\Delta E_b, \Delta E_l))}{(\partial \Delta E_b)^2} < 0 \qquad \text{and} \qquad \frac{\partial^2(P^{\text{fit}}(\Delta E_b, \Delta E_l))}{(\partial \Delta E_l)^2} < 0 \tag{6.13}$$

and no entry for the points where one or both relations are not fulfilled. The resulting distribution of the number of events with negative second derivatives of $P^{\text{fit}}(\Delta E_b, \Delta E_l)$ in each point $(\Delta E_b, \Delta E_l)$ is shown in Figure 6.10. In the same figure also the contour

plot is shown, where each contour from outside inwards corresponds to the addition of a fraction of 10 % relative to the number of entries in the bin with the maximal number of events. The bin in $(\Delta E_b, \Delta E_l)$ with the maximal number of events is in the

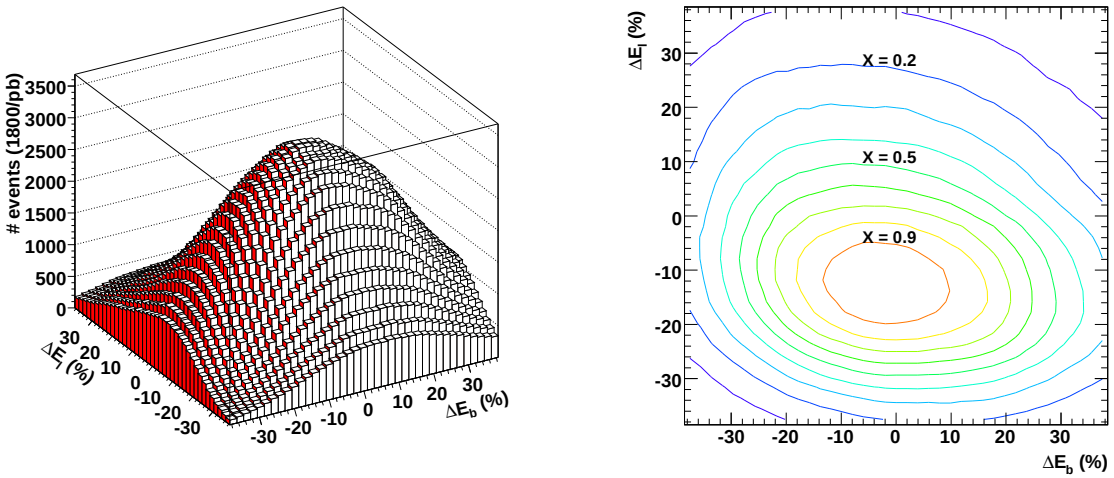


Figure 6.10: The distribution of the number of events with negative second derivatives of $P^{\text{fit}}(\Delta E_b, \Delta E_l)$ in each point $(\Delta E_b, \Delta E_l)$.

following called the "central value". The outlier rejection is now possible by requiring that the maximal probability of the fit $P^{\text{fit}}(\Delta E_b, \Delta E_l)$ in the event is reached at a point $(\Delta E_b, \Delta E_l)$ close to the central value. Therefore, a region is defined around the central value, which corresponds to a contour line in Figure 6.10, i.e by taking into account all $(\Delta E_b, \Delta E_l)$ bins that contain at least a fraction X of the number of entries corresponding to the bin content of the central value. To find an optimal region around the central value, the jet energy corrections are estimated for different values of X and the bias between the expected and estimated corrections is calculated. The methods to determine the expected and estimated residual jet energy corrections are explained in the following.

The expected residual jet energy corrections are determined from the semi-muonic $\mathrm{t\bar{t}}$ events for which the three jets assigned by the MVA method to the hadronic top quark decay are correctly matched to the partons in the $\mathrm{t \to bW \to bq\bar{q}}$ decay. For the matching quark-jet pairs, the difference between the quark energy and the jet energy is calculated relative to the jet energy. Examples of the resulting distributions are shown in Figure 6.11 for b quark and light quark jets. These relative error distributions are then fitted with a Gaussian function in a range of ± 1.5 times the root mean square (rms) of the distribution around the arithmetic mean. The expectation value of the fitted Gaussian and the uncertainty on this number are taken as the expected residual jet energy corrections $\Delta E_{b/l,\text{exp}}$. The expectation value of the fitted Gaussian depends strongly on the range which is taken into account for the fit. In Table 6.3 the expectation values from the Gaussian fit are shown for different fit ranges between ± 1 and ± 3 times the root mean square of the distribution around the arithmetic mean. Especially for the expected jet energy correction of the b quark jets a large absolute variation of about 4 % is observed. For the light quark jets, the effect is less pronounced

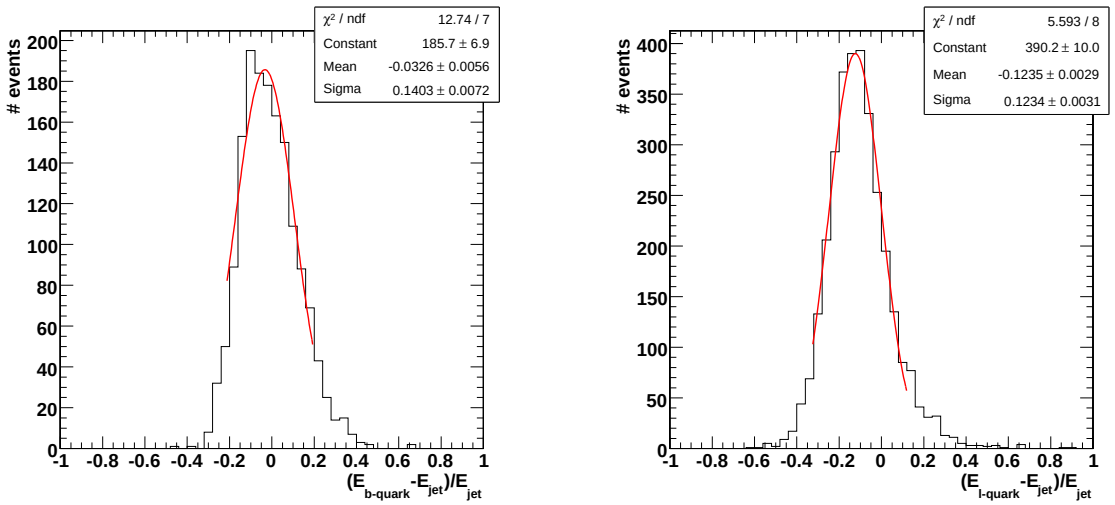


Figure 6.11: The expected jet energy corrections in % for b jets (left) and light (right) quark jets.

with an absolute change of only 1.5 %. The range of ±1.5 times the root mean square is used to fit the Gaussian function, to avoid the non-Gaussian effects in the tails of the distribution which start to contribute for a larger range. The non-Gaussian tails are more pronounced for the residual correction on the b quark jet energy due to the presence of neutrino's and muons from the decay of heavy hadrons. The dependence on the range which is used to perform the Gaussian fit should be kept in mind when comparing the estimated and expected values of the jet energy corrections in the current and following chapters. On the estimated values, which are in the end relevant for the measurement on data, it will be shown that no such ambiguities appear.

	±1 rms	±1.5 rms	±3 rms
$\Delta E_{b,\mathrm{exp}}$ (%)	-5.36±1.13	-3.26±0.56	-1.31±0.37
$\Delta E_{l,\mathrm{exp}}$ (%)	-12.92±0.51	-12.35±0.29	-11.41±0.25

Table 6.3: The expected jet energy corrections obtained from the Gaussian fit using different fit ranges.

The estimated values of the residual jet energy corrections $(\Delta E_b, \Delta E_l)$ are obtained by minimizing the combined $\Delta\chi^2(\Delta E_b, \Delta E_l)$ distribution. The values minimizing the distribution are denoted as $(\Delta E_b)_{\mathrm{min}}$ and $(\Delta E_l)_{\mathrm{min}}$. They are obtained on the two dimensional grid of points. In order to estimate the final residual corrections the two dimensional distribution is projected into one dimension as $\Delta\chi^2(\Delta E_b) = \Delta\chi^2(\Delta E_b, (\Delta E_l)_{\mathrm{min}})$ and $\Delta\chi^2(\Delta E_l) = \Delta\chi^2((\Delta E_b)_{\mathrm{min}}, (\Delta E_l))$. The parabolic tendency of the resulting one dimensional $\Delta\chi^2$ distributions is shown in Figure 6.12. A parabola is analytically calculated using three points around the minimum. From this parabola the best estimates $\Delta E_{b/l,\mathrm{est}}$ and the uncertainty on the residual corrections are obtained.

The 1σ, 2σ and 5σ contours around the minimum of the $\Delta\chi^2(\Delta E_b, \Delta E_l)$ distribution

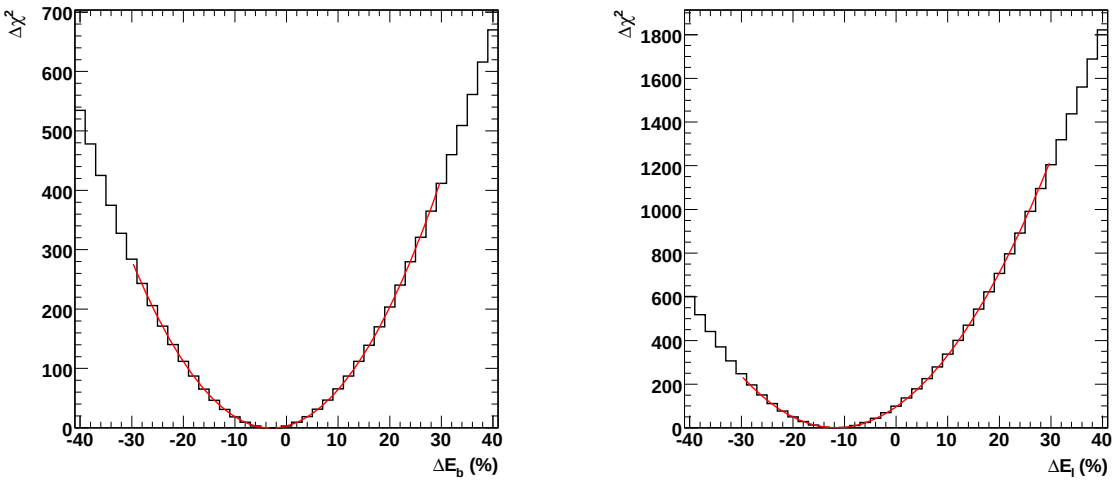


Figure 6.12: The $\Delta\chi^2$ distribution as a function of the jet energy corrections in % for b jets (left) and light (right) quark jets. The estimated jet energy corrections corresp to the minimum of the analytically calculated parabola. Expectation for $50\,\mathrm{pb}^{-1}$.

are shown in Figure 6.13. No correlation between the b quark jet energy corrections and light quark jet energy corrections is observed.

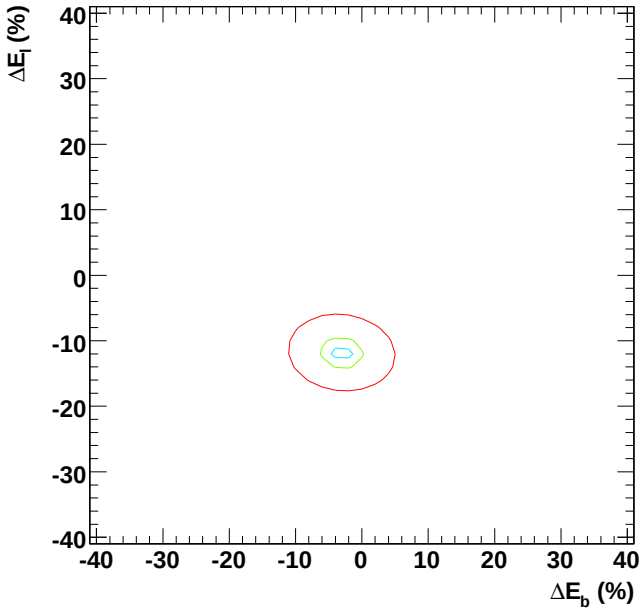


Figure 6.13: The 1σ, 2σ and 5σ contours around the minimum of the $\Delta\chi^2(\Delta E_b, \Delta E_l)$ distribution. Expectation for $50\,\mathrm{pb}^{-1}$.

The estimation of the residual jet energy corrections is repeated for different sizes of the region around the central value. Different fractions X between 0.0 and 0.9 are

considered. The bias $\Delta E_{\text{est}} - \Delta E_{\text{exp}}$ as a function of the value X applied to reduce the outliers is shown in Figure 6.14. For the estimation of the residual light quark jet

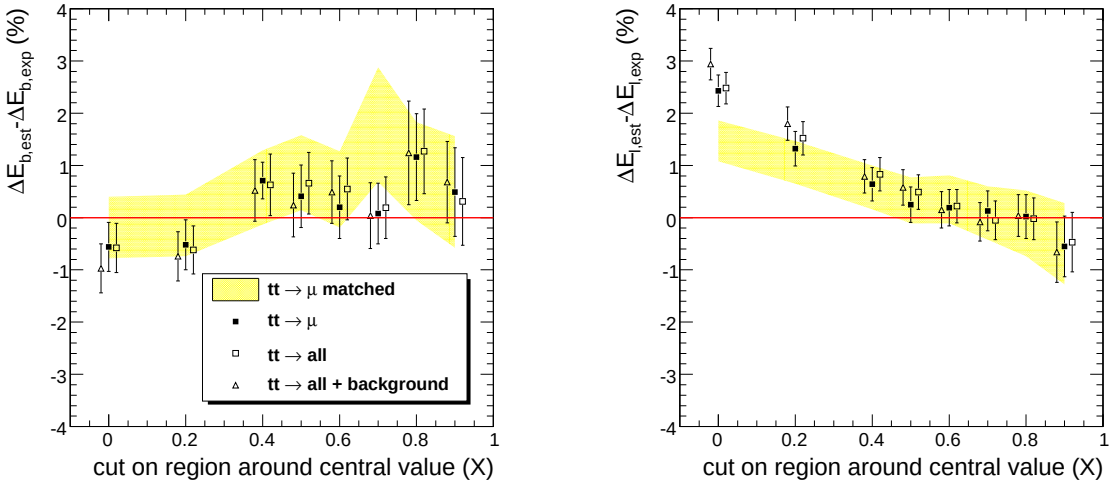


Figure 6.14: Effect of the cut on the area around the central value (parameter X) on the bias of the estimation of the b (left) and light (right) jet energy corrections.

energy corrections a bias is visible if no cut is applied, i.e. if all events are taken into account or $X = 0$. An influence on the estimated jet energy corrections of non-signal events is observed. These effects disappear by taking into account only the events for which the maximal probability of the fit lies in a "smaller region" around the central value. This smaller region is defined by the collection of $(\Delta E_b, \Delta E_l)$ bins that contain at least a fraction $X = 0.4$ of the number of events in the bin of the central value.

The region around the central value is defined by the number of events entering the region. Therefore, if $\mathrm{t\bar{t}}$ and background events are used together to define the region, the region does not necessarily contain the same amount of signal events as in the case in which only semi-muonic $\mathrm{t\bar{t}}$ events are used to define the region. The fraction $X = 0.4$ was argumented on signal events only. Hence, it is interesting to see in which region about the same amount of signal events is contained if we use either only signal events or signal together with background events to define the region. The signal efficiency when using different values of X is shown in Figure 6.15. To obtain the same signal efficiency when using a fraction of $X = 0.4$ to define the region around the central value in case only signal events were used, a fraction of about $X = 0.5$ is needed when background is added. This increase of X to obtain the same signal efficiency reflects the fact that background events contain no intrinsic information of the $\mathrm{t \to bW \to bq\bar{q}}$ decay and are therefore spread much more in the $(\Delta E_b, \Delta E_l)$ space compared to signal events. Therefore, the fraction of $X = 0.5$ is used to remove the outliers in the remainder of the analysis. After applying this outlier rejection, the number of events that remains for the estimation is given in Table 6.4. More than half of the W+jets and Z+jets events are removed by the outlier rejection, while 70 % of the signal events are kept.

Using the fraction of $X = 0.5$ to remove outliers, the estimated residual jet energy

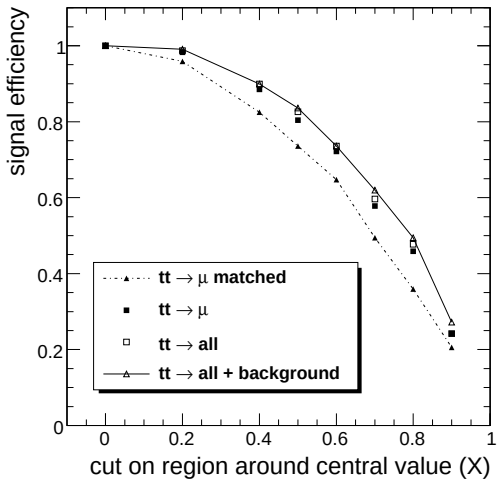


Figure 6.15: Efficiency of the cut on the area around the central value on the selection efficiency of the signal events.

	$t\bar{t} \to \mu$	$t\bar{t}$ →other	$t(\bar{t})$	W+jets	Z+jets
# events before	173.7±1.9	25.71±0.75	6.02±0.22	24.27±0.77	3.14±0.29
# events after	122.0±1.6	15.70±0.59	3.67±0.23	12.00±0.54	1.19±0.18

Table 6.4: The number of events before and after the outlier rejection for signal and background event samples. The indicated number of events before and after the cuts are rescaled to an integrated luminosity of $\mathcal{L}$ =50 pb^{-1}.

corrections are

$$\Delta E_{b,\mathrm{est}} = -3.02 \pm 1.44\,\% \quad \text{and} \quad \Delta E_{l,\mathrm{est}} = -11.77 \pm 1.02\,\%. \tag{6.14}$$

The statistical uncertainty was calculated using a number of events corresponding to an integrated luminosity of 1800 pb^{-1} but were rescaled to an integrated luminosity of 50 pb^{-1}. The expected residual jet energy corrections are

$$\Delta E_{b,\mathrm{exp}} = -3.26 \pm 0.56\,\% \quad \text{and} \quad \Delta E_{l,\mathrm{exp}} = -12.35 \pm 0.29\,\%. \tag{6.15}$$

Using the statistical uncertainty corresponding to the number of events for an integrated luminosity of 1800 pb^{-1}, the estimated jet energy corrections and uncertainty are

$$\Delta E_{b,\mathrm{est}} = -3.02 \pm 0.24\,\% \quad \text{and} \quad \Delta E_{l,\mathrm{est}} = -11.77 \pm 0.17\,\%. \tag{6.16}$$

With these uncertainties the bias is calculated from $\Delta E_{b/l,\mathrm{est}} - \Delta E_{b/l,\mathrm{exp}}$ and results in

$$\Delta E_{b,\mathrm{est}} - \Delta E_{b,\mathrm{exp}} = +0.24 \pm 0.61\,\% \quad \text{and} \quad \Delta E_{l,\mathrm{est}} - \Delta E_{l,\mathrm{exp}} = +0.58 \pm 0.34\,\%. \tag{6.17}$$

Therefore, we can conclude that the estimation and expectation of the residual jet energy corrections are indeed compatible.

6.5 Statistical properties of the estimator

6.5.1 Linearity of the method

A calibration curve is obtained by applying a relative inclusive shift on the energy scales of the reconstructed jets prior to the full analysis. When for each shift the analysis is performed, the resulting estimates of the residual corrections $\Delta E_{b/l,est}$ are compared to the expected corrections $\Delta E_{b/l,exp}$ for the selected events. The resulting curves are shown in Figure 6.16 for the case were only signal events were used and in Figure 6.17 for all $t\bar{t}$ events. The line that is shown corresponds to the case for which the expected and estimated residual jet energy corrections are equal, i.e. $\Delta E_{b/l,est} = \Delta E_{b/l,exp}$. It is observed that the estimator is linear with a unity slope for the calibration results when only signal events are considered. When adding also other $t\bar{t}$ decays the estimator has the tendency to become non-linear and has a slope different from unity in case the estimated corrections deviate strongly from zero. This is solved by performing several iterations.

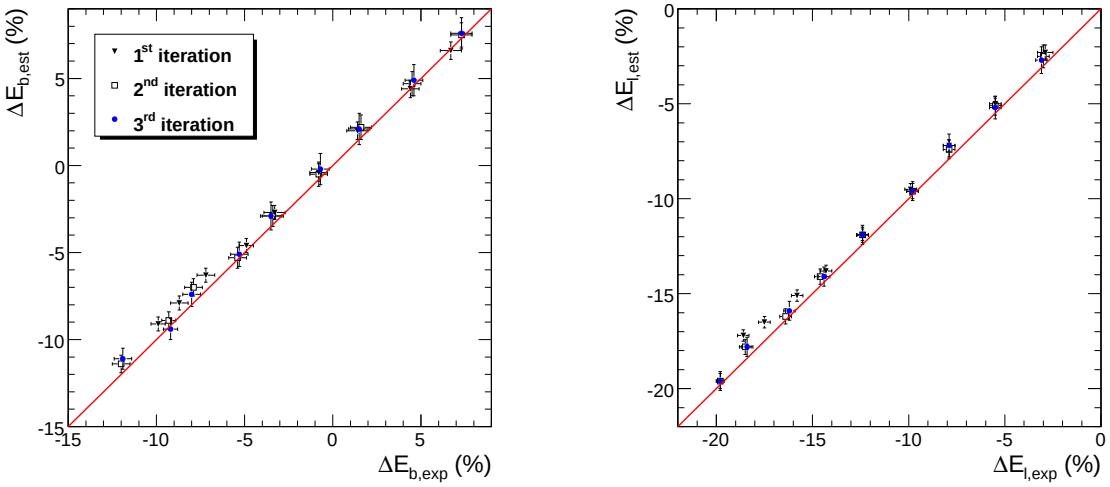


Figure 6.16: Calibration results representing the estimated corrections from successive iterations versus the expected corrections for b (left) and light (right) quark jets. Only signal events were used.

The estimated residual jet energy correction factors from this first iteration can be applied on the initial jet energy scales. In a second iteration the residual jet energy correction factors of the corrected jets of the first iteration can again be estimated. Ideally, if the jet energy corrections estimated in a first iteration are equal to the expected jet energy correction factors, the residual jet energy correction factor in a second iteration should be equal to one (or 0 %) reflecting that no residual correction is needed. If a difference was observed between the expected and estimated jet energy corrections after the first iteration, the difference becomes smaller after the second iteration. Thus, if the residual jet energy corrections to be estimated are far from zero, several iterations can be applied so the estimation converges to the line $\Delta E_{b/l,est} = \Delta E_{b/l,exp}$, hence a linear calibration curve with a slope of unity. It is shown by the plots that three iterations are sufficient, but when applying the calibration method on

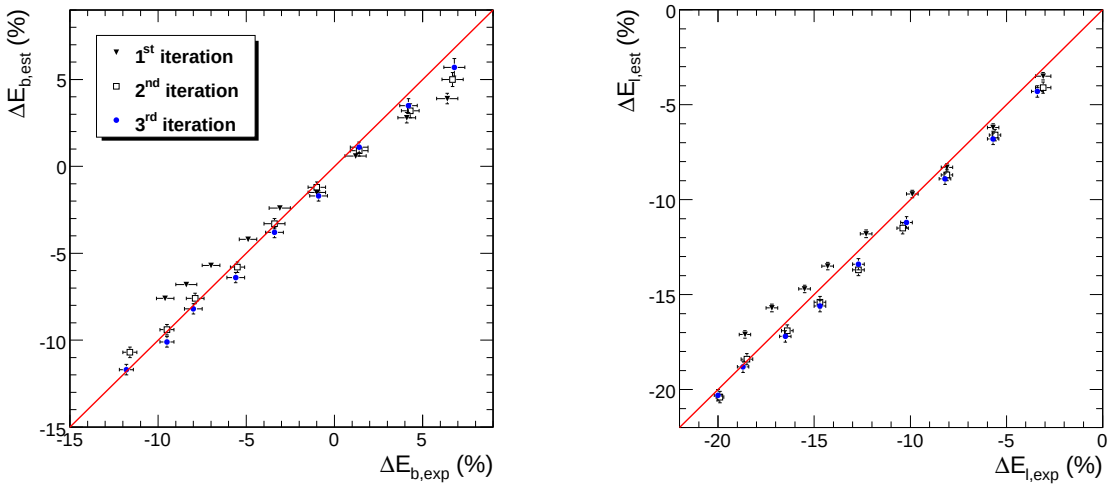


Figure 6.17: Calibration results representing the estimated corrections from successive iterations versus the expected corrections for b (left) and light (right) quark jets. All $t\bar{t}$ events were used.

real proton collisions one can iterate until the estimated jet energy corrections are zero. Additionally, it is interesting to note that the method is linear but if this was not the case, possible non-linearities of the estimator can be resolved by iterating several times.

6.5.2 Pull and statistical uncertainty

Resampling techniques are applied to study the properties of the estimation of the statistical uncertainty. Each data set reflects $50\,\mathrm{pb}^{-1}$ of integrated luminosity. In total about 400 pseudo-experiments are performed containing a Poissonian distributed amount of events for the $t\bar{t}$ events. Background events were not taken into account. The pull distributions are calculated as

$$(\Delta E^i_{b,l} - < \Delta E_{b,l} >)/\delta\Delta E^i_{b,l} \tag{6.18}$$

where $\Delta E^i_{b,l}$ and $\delta\Delta E^i_{b,l}$ are respectively the estimated corrections and the uncertainty on the estimated corrections as obtained from the data set $i \in 1, ..., N$ and $< \Delta E_{b,l} >$ is the average jet energy correction estimated over all N data sets. The resulting pull distributions are shown in Figure 6.18, by construction they are centered around zero. The distributions of the pull variable are fitted with a Gaussian function, resulting in a width equal to 1.4 and 1.3 for the estimators of respectively the b-quark jet and the light quark jet. If the uncertainties are correctly estimated, the pull distribution would have a width of one. The non-unity width can have several origins, for example due to the uncertainties on the jet parameters that are underestimated. The statistical uncertainties of the estimated residual jet energy scale corrections are corrected for the non-unity width of the pull distributions, resulting in

$$\Delta E_{b,\mathrm{est}} = -3.02 \pm 2.02\,\% \quad \text{and} \quad \Delta E_{l,\mathrm{est}} = -11.77 \pm 1.33\,\%. \tag{6.19}$$

Using a data set corresponding to an integrated luminosity of $50\,\mathrm{pb}^{-1}$ at a center of mass energy of 10 TeV, the jet energy corrections can thus be estimated with a

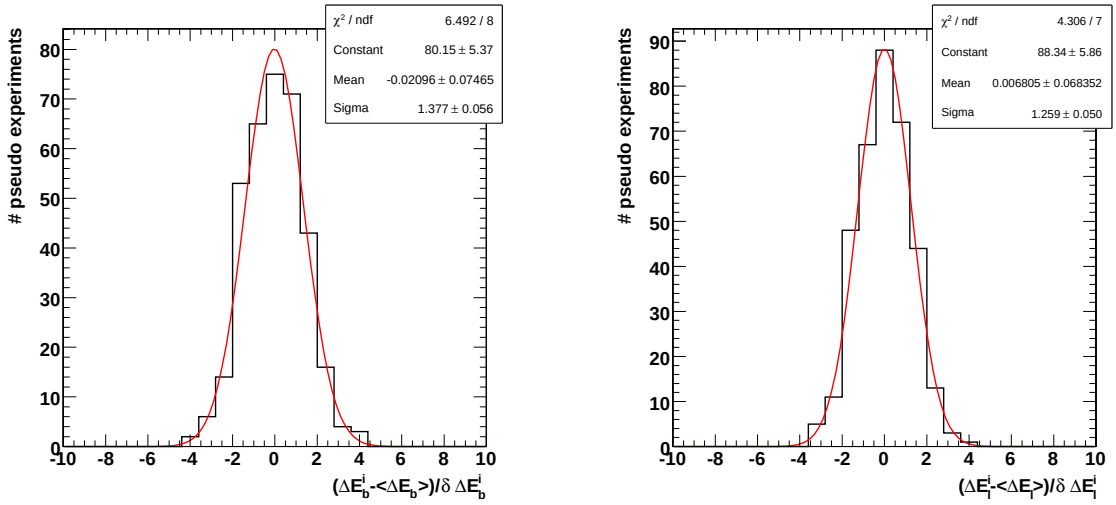


Figure 6.18: Pull distribution constructed from about 400 pseudo-experiments for the b quark jet energy correction estimator (left) and the light quark jet energy correction estimator (right).

statistical precision of respectively 2.0 % for the b quark jets and 1.3 % for the light quark jets. When the full integrated luminosity of 1800 pb^{-1} is used, the uncertainties on the estimation of the jet energy corrections are respectively 0.34 % for b quark jets and 0.22 % for light quark jets. The uncertainty on the bias is also calculated using the pull corrected estimation

$$\Delta E_{b,\mathrm{est}} - \Delta E_{b,\mathrm{exp}} = +0.24 \pm 0.66\,\% \quad \text{and} \quad \Delta E_{l,\mathrm{est}} - \Delta E_{l,\mathrm{exp}} = +0.58 \pm 0.36\,\%. \tag{6.20}$$

The distributions of the uncertainties $\delta\Delta E^i_{b,l}$ are shown in Figure 6.19 and fitted with a Gaussian function resulting in an excellent fit probability. This Gaussian behaviour illustrates the robustness of the estimator on samples with an integrated luminosity of 50 pb^{-1}.

The pull distributions were also calculated using pseudo-experiments containing a Poissonian distributed amount of events for both signal and background and the resulting distributions look similar.

6.5.3 Improvement when including more jet combinations

To improve the statistical performance of the method, more jet combinations can be taken into account. The information of the different jet combinations can be combined according to

$$P^{\mathrm{fit}}_{\mathrm{combined}}(\Delta E_b, \Delta E_l) = \sum_{i=1}^{n} w_i P^{\mathrm{fit}}_i(\Delta E_b, \Delta E_l) \tag{6.21}$$

where w_i is a weight depending on the probability for jet combination i to be correct and n is the number of jet combinations to be taken into account. When using all semi-muonic $\mathrm{t\bar{t}}$ decays after the kinematic event selection, the probability for a given jet combination to be correct can be calculated as a function of the MVA discriminant value of the jet combination. This probability is shown in Figure 6.20 for the three

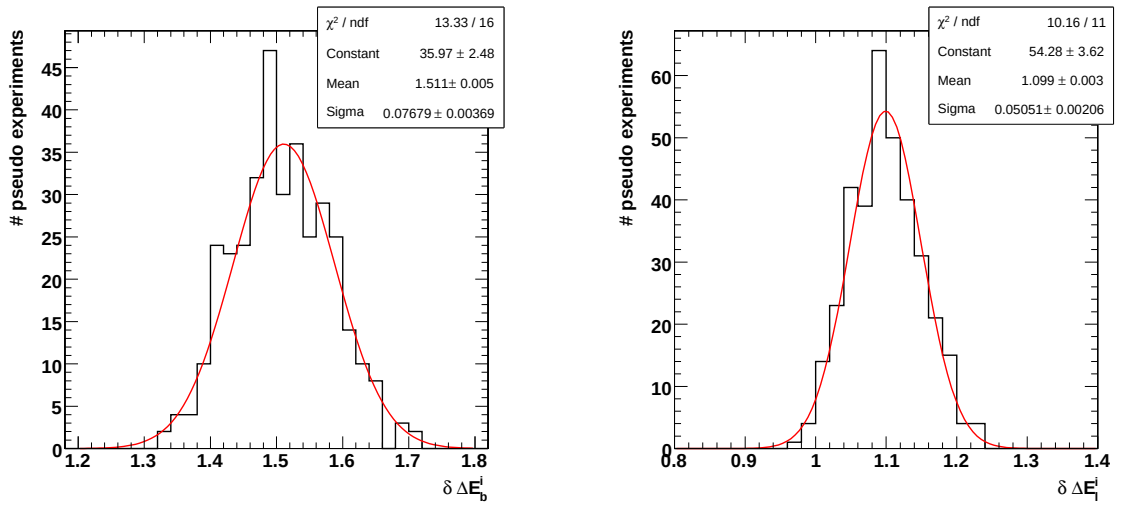


Figure 6.19: Distribution of the uncertainties constructed from about 400 pseudo-experiments for the b quark jet energy correction estimator (left) and the light quark jet energy correction estimator (right).

jet combinations with the highest MVA discriminant value. If we consider only the

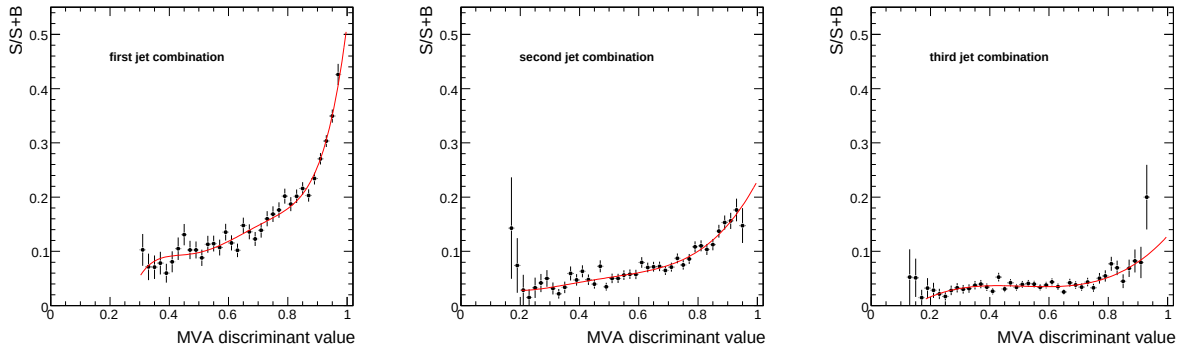


Figure 6.20: Probability to select a correct jet combination as a function of the MVA discriminant value of the first (left), second (middle) and third (right) jet combination.

events for which three of the four leading jets can be matched to the quarks in the $t \to bW \to bq\bar{q}$ decay, the probability to select the correct jet combination when taking the first, second or third jet combination ordered in decreasing MVA discriminant value is respectively 48.3 %, 19.5 % and 9.6 % when no cut is applied on the MVA discriminant values. Thus, by taking into account the first three jet combinations, the correct jet combination is selected for 77.4 % of the events, provided that it exists. A lower threshold of 0.2 on the MVA discriminant values of the three jet combinations, would select 95 % of the events for which three of the four leading jets can be matched to the quarks in the hadronic top quark decay. At the same time, for the selected events, the probability to find the correct jet combination among the first three jet combinations in the MVA ordering increases from 77.4 % to 81.6 %. In the analysis, the jet combinations with an MVA discriminant value above 0.2 are taken into account, with a maximum of

three jet combinations for each event. The weight for each jet combination is defined as the probability of having the correct jet combination corresponding to the MVA discriminant value of the jet combination. These probabilities are provided by the polynomial fits in Figure 6.20. The pull distributions are obtained using 350 pseudo-experiments each representing a data set with an integrated luminosity of $50\,\mathrm{pb}^{-1}$. An improvement in the statistical uncertainty of about 20 % is observed for the light quark jet energy corrections. No significant improvement is observed for the b quark jet energy corrections.

6.6 Robustness of the method and systematic uncertainties

In this section various possible contributions to the systematic uncertainty are considered. The robustness of the method to estimate the jet energy corrections with respect to the amount of background, jet reconstruction issues and the event generator used to simulate the events, is checked.

6.6.1 Robustness with respect to the amount of background events

There are two main contributions to the background events, namely what is called combinatorial background and process background. The amount of events for which the wrong jet combination is chosen is related to the efficiency of the MVA discriminant to choose the jet combination. Two aspects are therefore important, the cut that is made on the MVA discriminant value and the performance of the MVA discriminant to choose the correct jet combination. For the process background, i.e. non-$\mathrm{t\bar{t}}$, the total amount of events passing all selection criteria depends strongly on our knowledge of the cross sections. If these cross sections are significantly underestimated, much more background events will be selected in the analysis applied on real collision data.

Effect of the cut on MVA discriminant value

Ideally, the estimation of the jet energy correction factors should not depend on the cut that is applied on the MVA discriminant value of the chosen jet combination. The cut value on the MVA discriminant is varied between 0.2 and 0.8 and the expected and estimated jet energy corrections are calculated. The resulting values are shown in Figure 6.21. The uncertainties have been corrected for the non-unity width of the pull distributions. The estimation is flat with respect to the cut on the MVA discriminant value of the chosen jet combination. No bias on the estimation is expected due to the cut value. Therefore, there is no need in increasing the cut value. A loose cut can be applied on the MVA discriminant because the cut based on the region around the central value rejects already possible outliers.

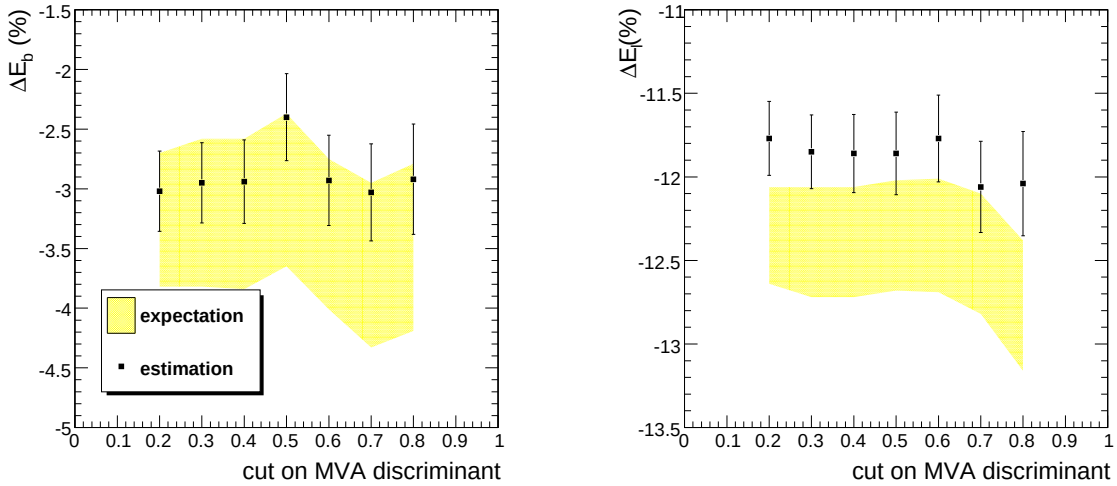


Figure 6.21: Effect of a cut on the MVA discriminant value of the chosen jet combination on the estimation of the jet energy corrections for b quark jets (left) and light quark jets (right).

Increasing the jet combinatorial background

The effect of an increased amount of wrong jet combinations is estimated by looking only at the $t\bar{t}$ events and chosing for 50 % of the events a random jet combination instead of the one with the highest MVA discriminant value. To make a random choice, each jet combination gets a probability of 1/12. The random jet combination can be correct or wrong, but in general more events with a wrong jet combinations will contribute to the total number of events. A cut on the MVA discriminant value of the chosen jet combination is applied of 0.2. In Table 6.5 the estimated jet energy corrections and bias are shown for the nominal chosen jet combination and for 50 % random jet combinations. The expected corrections might change as different jet combinations are taken into account while taking the chosen jet combination or 50 % random jet combinations. The uncertainty on the estimation slightly increases when taking random jet combinations. This can be understood because less events which fulfill the mass constraints contribute to the estimation, thus resulting in a higher uncertainty. The bias between the estimation and the expectation of the residual jet energy corrections for light and b quark jets is compatible with zero. No bias is expected because the events that do not fulfill the mass constraints are equally distributed over the full $(\Delta E_b, \Delta E_l)$ space. Although a possible bias is dominated by the statistical uncertainty, the fact that the bias is compatible with zero indicates that a decreased performance of the MVA discriminant to choose the correct jet combination should not constrain the potential of the described method. The nominal sample is at least 50 % correlated with the sample where half of the events have a random jet combination choice. Therefore the uncertainty on the difference in bias between both samples is smaller than the square root of the sum of variances on the bias of each sample. For both b quark and light quark jets correction estimators a conservative systematic uncertainty of 0.5 % is quoted dominated by the statistical precision on the estimate of the systematic effect.

	$\Delta E_{b,\text{exp}}$ (%)	$\Delta E_{b,\text{est}}$ (%)	$\Delta E_{b,\text{est}} - \Delta E_{b,\text{exp}}$ (%)
nominal	-3.08±0.55	-2.42±0.31	+0.66±0.63
50 % random	-2.66±0.64	-3.26±0.38	+0.60±0.74
	$\Delta E_{l,\text{exp}}$ (%)	$\Delta E_{l,\text{est}}$ (%)	$\Delta E_{l,\text{est}} - \Delta E_{l,\text{exp}}$ (%)
nominal	-12.32±0.29	-11.83±0.21	+0.49±0.36
50 % random	-12.18±0.37	-12.18±0.26	+0.00±0.45

Table 6.5: Estimated jet energy corrections and bias for the chosen jet combination (nominal) and when 50 % random jet combinations are used.

Possible underestimation of the cross section of background processes

The theoretical cross sections of the background processes can be underestimated as often only LO cross section calculations are available from literature. Therefore, the robustness with respect to the amount of events from process background is checked by increasing the cross sections for the different background processes with a factor of two. The estimation was then performed with a doubled amount of background events. The results are summarized in Table 6.6. No difference can be observed within

	$\Delta E_{b,\text{exp}}$ (%)	$\Delta E_{b,\text{est}}$ (%)	$\Delta E_{b,\text{est}} - \Delta E_{b,\text{exp}}$ (%)
nominal	-3.26±0.56	-3.02±0.34	+0.24±0.66
background × 2	-3.20±0.54	-3.05±0.43	+0.15±0.69
	$\Delta E_{l,\text{exp}}$ (%)	$\Delta E_{l,\text{est}}$ (%)	$\Delta E_{l,\text{est}} - \Delta E_{l,\text{exp}}$ (%)
nominal	-12.35±0.29	-11.77±0.22	+0.58±0.36
background × 2	-12.36±0.29	-11.87±0.30	+0.49±0.42

Table 6.6: Expected and estimated jet energy corrections for the nominal situation and a process background that is increased by a factor of two.

the statistical uncertainties with respect to the nominal situation. Therefore, it is concluded that an underestimation of the number of background events with respect to the possible number of background events in real proton collisions has a negligible effect on the estimation of the jet corrections. The uncertainties are higher when the background is increased as the size of the samples is rescaled according to the integrated luminosity of the smallest background sample, resulting in less $t\bar{t}$ events. The sample for which the background was increased with a factor of two is almost maximally correlated with the nominal sample. Therefore, a conservative systematic uncertainty is quoted for both b quark and light quark jet energy corrections of 0.1 %.

6.6.2 Jet reconstruction issues and overlapping jets

In this section the effect of the angular bias on the jet reconstruction described in Section 4.1.2 on the estimation of the residual jet energy correction factors is determined. Furthermore, the jet energies are smeared prior to the full analysis to investigate the robustness of the estimation with respect to this additional smearing. Also the event selection cut on the smallest ΔR between the jets is studied to get an idea of the possible influence on the estimation if this criterion is altered.

Smearing of jet energies

It is possible that the energies of the jets are not well measured in the first period of data taking at the LHC. Therefore, the effect of smeared jet energies on the estimation of the residual jet energy corrections is studied. Prior to the full analysis, the jet four-momentum is scaled according to a random number taken from a Gaussian distribution with a variance of $(1.3\sigma_E)^2 - (\sigma_E)^2$, where σ_E is the energy resolution. The procedure reflects an extra smearing of 30 % of the current resolution on the jet energy. This study was performed using only the $t\bar{t}$ events. Table 6.7 shows the expected and estimated jet energy corrections together with the bias. The smearing results in slightly less selected events and therefore slightly higher uncertainties as the cut on the maximum probability of the kinematic fit will reject more events because of larger deviations between the measured energy scale and the enforced mass constraints. Both samples, smeared and nominal are highly correlated. Therefore, the observed change in bias, $\Delta E_{\text{est}} - \Delta E_{\text{exp}}$, has a small uncertainty and is taken as a measure of the systematic effect. For the b quark jet energy corrections we quote 0.3 % and for the light quark jet energy correction a value of 0.1 % is taken.

Angular bias on the jet reconstruction

In Section 4.1.2 the bias on the reconstructed polar angle of the jets was discussed. The method to estimate the jet energy corrections might be sensitive to this bias because the W boson mass is used as a constraint. The mass of the reconstructed W boson is related with the angle between its decay products. Therefore, an angular bias on the reconstructed jets can induce a bias on the method. To check the sensitivity of the bias on the estimation, a function is fitted to the distribution of $\theta_{parton} - \theta_{CaloJet}$ as a function of $\theta_{CaloJet}$. Using the fitted function the bias on the polar angle of the

	$\Delta E_{b,\text{exp}}$ (%)	$\Delta E_{b,\text{est}}$ (%)	$\Delta E_{b,\text{est}} - \Delta E_{b,\text{exp}}$ (%)
nominal	-3.08±0.55	-2.42±0.31	+0.66±0.63
30 % smeared	-3.58±0.74	-2.62±0.32	+0.96±0.81
	$\Delta E_{l,\text{exp}}$ (%)	$\Delta E_{l,\text{est}}$ (%)	$\Delta E_{l,\text{est}} - \Delta E_{l,\text{exp}}$ (%)
nominal	-12.32±0.29	-11.83±0.21	+0.49±0.36
30 % smeared	-12.96±0.43	-12.38±0.22	+0.58±0.48

Table 6.7: Estimated jet energy corrections and bias for the nominal situation and after a smearing is applied to the jet energies.

jets is corrected prior to the full analysis. Only signal events were considered to study the effect of correcting for the angular bias, i.e. semi-muonic $t\bar{t}$ decays for which the jets can be matched to the quarks of the hadronically decaying top quark branch. The results are shown in Table 6.8. Due to the very strong correlation between the samples,

	$\Delta E_{b,\text{exp}}$ (%)	$\Delta E_{b,\text{est}}$ (%)	$\Delta E_{b,\text{est}} - \Delta E_{b,\text{exp}}$ (%)
nominal	-3.27±0.56	-2.69±0.62	+0.58±0.84
θ corrected	-3.07±0.52	-2.32±0.57	+0.75±0.77
	$\Delta E_{l,\text{exp}}$ (%)	$\Delta E_{l,\text{est}}$ (%)	$\Delta E_{l,\text{est}} - \Delta E_{l,\text{exp}}$ (%)
nominal	-12.44±0.29	-11.86±0.40	+0.58±0.49
θ corrected	-12.40±0.28	-11.43±0.38	+0.97±0.47

Table 6.8: Estimated jet energy corrections and bias for the nominal situation and after a correction for the angular bias on the jet direction is applied.

as a systematic uncertainty the difference in bias is quoted. This is 0.2 % and 0.4 % for respectively the b quark and light quark jet energy corrections.

Effect of overlapping jets

To study the effect of the applied criterion for the removal of possible overlapping jets, only $t\bar{t}$ events were used. A first check is performed by changing the transverse momentum threshold on the jets that are considered in the event for the calculation of the smallest ΔR. The transverse momentum threshold is altered from 15 GeV/c to 30 GeV/c in steps of 5 GeV/c. If the smallest ΔR in the event is between one of the four leading jets and another jet not among the leading four jets, the estimation of the jet energy corrections might be sensitive to this change of threshold. However, no difference on the bias is observed.

The influence of the requirement $\Delta R_{\text{smallest}} > 0.8$ or, in case $\Delta R_{\text{smallest}} < 0.8$, an overlap energy between the two closest jets of zero, on the estimated jet energy corrections is studied further by dropping the second part of the requirement. Therefore, the only remaining requirement was that $\Delta R_{\text{smallest}} > 0.8$, where $\Delta R_{\text{smallest}}$ is the smallest angle between the leading jets and the other jets in the event. The resulting jet energy corrections are compatible with the expectation and compatible with the default results.

The estimation is repeated by considering only the $t\bar{t}$ events with exactly four jets above 30 GeV/c and by applying the default requirement, where $\Delta R_{\text{smallest}}$ is calculated this time between the four jets. Also in this case no significant deviation is observed.

As a last check for the sensitivity with respect to the applied criterion, the criterion was dropped and the estimation was performed in bins of the smallest angle between the leading jet and any other jet in the event $\Delta R_{\text{smallest}}^{(\text{jet1,jeti}\neq 1)}$. Only jets with a transverse momentum exceeding 15 GeV/c are considered. The expected and estimated jet energy corrections for light and b quark jets are shown in Figure 6.22. For the b quark jet

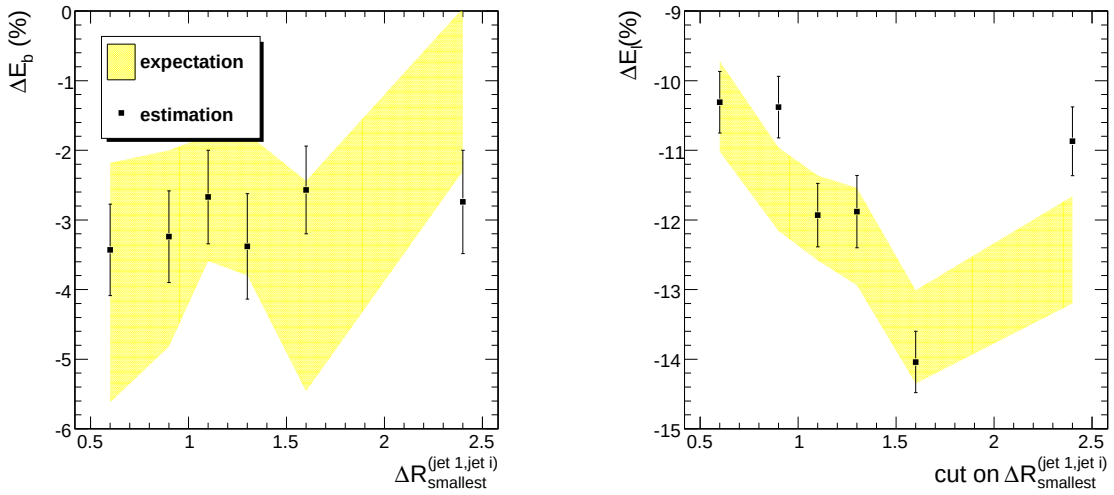


Figure 6.22: The expected and estimated jet energy corrections for b (left) and light (right) quark jets in bins of $\Delta R_{\text{smallest}}^{(\text{jet1,jeti}\neq 1)}$.

energy corrections, no real dependency is observed. One could think of seeing a possible dependency for the light quark jet energy corrections, but more simulated events are needed to determine this dependency correctly. One should however note that no real bias is seen between the estimated and expected jet energy corrections. In first order

it is thus concluded that the method is not sensitive to small variations of the criterion on overlapping jets.

6.6.3 Event simulation aspects

It is important to know if the method is sensitive to different aspects in the event simulation because it is not known yet how well the simulated proton collisions approximate the real proton collisions. By varying parameters in the simulation related to certain phenomena in the event simulation, a systematic uncertainty can be determined. In the following, the statistical uncertainty for both light and b quark jet energy correction estimations are corrected for the non-unity width of the pull distributions.

Initial and final state radiation

The PYTHIA event samples generated with different settings for the parameters controlling the amount of initial and final state radiation are used to determine the systematic uncertainty related to the amount of radiation. Table 6.9 summarizes the expected and estimated jet energy corrections for b and light quark jets for the different settings of the radiation parameters, which were discussed in Chapter 3. The expected values for

	$\Delta E_{b,\text{exp}}$ (%)	$\Delta E_{b,\text{est}}$ (%)	$\Delta E_{b,\text{est}} - \Delta E_{b,\text{exp}}$ (%)
nominal	-1.43±0.26	-1.30±0.17	+0.13±0.31
large ISR/FSR	-2.31±0.39	-2.95±0.24	-0.64±0.46
small ISR/FSR	-1.07±0.32	-1.22±0.22	-0.15±0.39
	$\Delta E_{l,\text{exp}}$ (%)	$\Delta E_{l,\text{est}}$ (%)	$\Delta E_{l,\text{est}} - \Delta E_{l,\text{exp}}$ (%)
nominal	-9.31±0.16	-8.94±0.12	+0.37±0.20
large ISR/FSR	-9.52±0.24	-8.95±0.17	+0.57±0.29
small ISR/FSR	-9.58±0.21	-8.89±0.16	+0.69±0.26

Table 6.9: Expected and estimated jet energy corrections and the bias on the estimation for different settings of the initial and final state radiation parameters.

the b quark jet energy corrections change as a function of the amount of radiation. The estimation follows the trend of the expected values and the bias on the estimation is compatible with zero within the statistical uncertainty. For both light and b quark jet energy corrections, a possible bias due to the amount of radiation is dominated by the statistical precision. For this study the samples are not correlated.

The MadGraph/MadEvent event samples generated with different settings for Λ_{QCD} are also related to the amount of initial and final state radiation. As a different top

quark mass was used to generate this sample, a direct comparison can not be made. However, the aim of this section is to show that the method is able to estimate the expected jet energy corrections and thus, that the resulting bias is compatible with zero. Table 6.10 lists the expected and estimated jet energy corrections. Also in this

	$\Delta E_{b,\text{exp}}$ (%)	$\Delta E_{b,\text{est}}$ (%)	$\Delta E_{b,\text{est}} - \Delta E_{b,\text{exp}}$ (%)
nominal	-0.97±0.41	-1.28±0.20	-0.31±0.46
Q^2 up	-1.10±0.45	-1.65±0.28	-0.55±0.53
Q^2 down	-2.41±0.48	-2.22±0.29	+0.19±0.56
	$\Delta E_{l,\text{exp}}$ (%)	$\Delta E_{l,\text{est}}$ (%)	$\Delta E_{l,\text{est}} - \Delta E_{l,\text{exp}}$ (%)
nominal	-9.29±0.26	-9.10±0.13	+0.19±0.29
Q^2 up	-10.41±0.29	-10.54±0.20	-0.13±0.35
Q^2 down	-10.10±0.28	-9.83±0.21	+0.27±0.35

Table 6.10: Expected and estimated jet energy corrections and the bias on the estimation for different settings of the Q^2 parameters.

case the estimated values are perfectly compatible with the expectation and the bias is dominated by the statistical precision.

For the description of the radiation a conservative systematic uncertainty is assigned, reflecting the statistical precision of both tests. This is approximated by values of 0.5 % and 0.3 % for respectively the b quark and light quark jet energy corrections.

Matrix element and parton shower matching threshold

The effect of changing the matching threshold is studied with the `MadGraph/MadEvent` event samples that were generated for that purpose (cfr. Section 3.4.2). Table 6.11 summarizes the expected and estimated residual jet energy corrections together with the bias on the estimation. The bias is compatible with zero if the statistical uncertainties are taken into account. These are again uncorrelated samples. Conservative systematic uncertainties of 0.5 % and 0.3 % are assigned to respectively the estimated b quark and light quark jet energy corrections.

Estimation using events simulated with `ALPGEN`

Although the top quark mass used to simulate the `ALPGEN` event sample is different as the top quark masses used in both the `PYTHIA` and `MadGraph/MadEvent` event samples, the estimation was also performed on this sample. It has been studied if there is a bias on the estimation of the jet energy corrections. The different top quark mass was

	$\Delta E_{b,\mathrm{exp}}$ (%)	$\Delta E_{b,\mathrm{est}}$ (%)	$\Delta E_{b,\mathrm{est}} - \Delta E_{b,\mathrm{exp}}$ (%)
nominal	-0.97±0.41	-1.28±0.20	-0.31±0.46
E_T^{min}= 10 GeV	-2.33±0.47	-2.08±0.29	+0.25±0.55
E_T^{min}= 40 GeV	-1.42±0.39	-2.03±0.24	-0.61±0.46
	$\Delta E_{l,\mathrm{exp}}$ (%)	$\Delta E_{l,\mathrm{est}}$ (%)	$\Delta E_{l,\mathrm{est}} - \Delta E_{l,\mathrm{exp}}$ (%)
nominal	-9.29±0.26	-9.10±0.13	+0.19±0.29
E_T^{min}= 10 GeV	-10.97±0.30	-10.45±0.21	+0.52±0.37
E_T^{min}= 40 GeV	-10.17±0.25	-9.86±0.16	+0.31±0.30

Table 6.11: Expected and estimated jet energy corrections and the bias on the estimation for different settings of the matching threshold.

taken into account for the constraint in the kinematic fit. The jet energy corrections are estimated to be

$$\Delta E_{b,\mathrm{est}} = -4.87 \pm 0.84\,\% \quad \text{and} \quad \Delta E_{l,\mathrm{est}} = -11.75 \pm 0.60\,\%, \tag{6.22}$$

while the expected corrections are

$$\Delta E_{b,\mathrm{exp}} = -3.38 \pm 0.55\,\% \quad \text{and} \quad \Delta E_{l,\mathrm{exp}} = -11.29 \pm 0.37\,\%. \tag{6.23}$$

Expected and estimated jet energy corrections are compatible within the statistical uncertainties, indicating that the type of event generator has no influence on the predicted performance of the method.

6.6.4 Systematic uncertainty due to top quark mass precision

The most important systematic uncertainty comes from the precision on the direct top quark mass measurements. While the W boson mass is measured with a precision of 0.03 %, the precision on the top quark mass is of the order of 0.7 % or about 1.25 GeV/c^2 [15]. Therefore, the method is repeated with a higher (+1.25 GeV/c^2) and a lower (-1.25 GeV/c^2) top quark mass as a constraint in the kinematic fit. In Table 6.12 the estimation of the jet energy corrections for the different top quark masses is given. The estimations of the light quark jet energy corrections for the different top quark masses are compatible with each other, because the light quark jet energy corrections are fully determined by the W boson mass. The difference between the estimations of the b quark jet energy corrections for the different top quark masses with respect to the nominal value is calculated. A shift of 1.77 % is observed when a lower top quark mass is used, while a shift of 0.76 % is found in case a higher top quark mass is used.

	m_t=169.65 GeV/c^2	m_t=170.9 GeV/c^2	m_t=172.15 GeV/c^2
$\Delta E_{b,\text{est}}$ (%)	-4.19±0.31	-2.42±0.31	-1.66±0.32
$\Delta E_{l,\text{est}}$ (%)	-11.75±0.21	-11.83±0.21	-11.74±0.21

Table 6.12: Estimated b and light quark jet energy corrections for three different top quark mass constraints.

It can be assumed that the effect should be symmetric around the nominal top quark mass value. Hence, the average of both shifts, or 1.3 %, is taken as the systematic uncertainty due to the uncertainty on the top quark mass.

6.6.5 Summary of the systematic uncertainties

In the previous sections it was shown that the bias due to possible systematic effects is sometimes dominated by the statistical uncertainty by which we estimate the effect. Therefore, as a conservative choice and to give an indication, the systematic uncertainty related to each of those aspects is taken as an approximation of the statistical uncertainty on the bias. Table 6.13 summarizes the systematic uncertainties obtained. Although compatible with zero, the statistical uncertainty on the estimation of a possible bias after the pull correction is added to the list. Since some of the possible

	$\sigma^{\text{syst}}_{\Delta E_{b,\text{est}}}$ (%)	$\sigma^{\text{syst}}_{\Delta E_{l,\text{est}}}$ (%)
possible bias	≤0.7	≤0.4
jet combinatorial background	≤0.5	≤0.5
process background	0.1	0.1
smearing	0.3	0.1
angular bias	0.2	0.4
radiation	≤0.5	≤0.3
matching threshold	≤0.5	≤0.3
uncertainty on m_{top}	1.3	-

Table 6.13: Summary of the possible contributions to the total systematic uncertainty.

systematic effects might be related and because a conservative choice was made, the systematic uncertainties can not be combined as usual. The uncertainty on the top quark mass clearly induces a dominating uncertainty on the estimation of the b quark jet energy correction. To quantify possible other systematic effects, the largest number in Table 6.13 is summed in quadrature to the systematic uncertainty related to the top

quark mass precision. A systematic uncertainty of 1.5% and 0.5% is obtained for the estimation of respectively the b and light quark jet energy corrections.

6.7 Other jet finding algorithms

So far, the developed method was applied on jets reconstructed with the Seedless Infrared Safe Cone jet algorithm. The estimation of the jet energy corrections can also be performed for jets reconstructed with another jet algorithm. Therefore, the method was tested on jets reconstructed using the inclusive longitudinally invariant k_T algorithm. For this jet clustering algorithm, the expected jet energy corrections are

$$\Delta E_{b,\text{exp}} = -1.92 \pm 0.57\,\% \quad \text{and} \quad \Delta E_{l,\text{exp}} = -11.00 \pm 0.34\,\%. \tag{6.24}$$

The developed method is then applied to estimate the jet energy corrections and the following values are obtained

$$\Delta E_{b,\text{est}} = -2.28 \pm 0.39\,\% \quad \text{and} \quad \Delta E_{l,\text{est}} = -10.76 \pm 0.26\,\%, \tag{6.25}$$

resulting in the a bias of respectively -0.36±0.71 % on the estimation of the b quark jet energy correction and +0.24±0.43 % for the light quark jet energy correction. Therefore, it was concluded that no bias is found.

In general, the method can be applied to estimate the jet energy corrections for jets reconstructed with various jet reconstruction algorithms. Also for jet types other than CaloJets, such as PFJets and JPTJets, residual jet energy corrections can be estimated using the developed method. However, the residual jet energy corrections for jets of this types are expected to be closer to zero.

When I have fully decided that a result is worth getting I go ahead of it and make trial after trial until it comes.

Thomas A. Edison

Chapter 7

Differential estimation of the jet energy corrections

The precise calibration of the jet energies is crucial for many physics studies at hadron colliders. Ideally, jet energy corrections should be provided which can be applied for a wide range of physics topologies. In order to transfer the jet energy corrections obtained from the top quark decay to another event topology, the corrections need to be differentiated according to the kinematics of the jet and the flavour of the parton associated to the jet. Different methods are proposed to estimate these corrections both from simulated and real proton collisions. Despite the success of the available methods, alternative methods for the estimation of the dependency of the jet energy corrections on the transverse momentum and pseudo-rapidity need to be studied and validation techniques need to be developed. Using top quark events it is possible to perform a differential estimation of the jet energy corrections.

In Chapter 6 a method is developed to estimate the residual jet energy corrections for light quark jets coming from the decay of a W boson and b quark jets. The W boson and top quark mass constraints are applied in an event-by-event kinematic fit on the $t \to bW \to bq\bar{q}$ decays. The estimation of the residual jet energy correction factors is performed in an inclusive way obtaining one overall correction for both light and b quark jets. It is however possible with the same techniques, to perform a differential estimation of the jet energy correction factors. One can differentiate with respect to the jet kinematics, i.e. the jet transverse momentum and pseudo-rapidity. In addition a differentiation with respect to the number of muons in the b quark jet is also feasible. For the analysis presented in this chapter, only the $t\bar{t}$ events were used to demonstrate the performance of the method. The reason for the choice of working only with the $t\bar{t}$ events is fully related to the CPU time needed to perform the analysis. However, no bias is expected due to the addition of process background events, because, as demonstrated in the previous chapter, these events have no influence on the measurement. Section 7.1 elaborates on the differential estimation of the jet energy corrections. Furthermore, a factorized approach for diverse differential jet energy corrections is discussed in Section 7.2. To validate the method, the jet energy corrections obtained after the factorized approach are applied and the reconstructed

W boson and top quark masses are shown in Section 7.3. Finally, in Section 7.4, the proposed method is compared with other methods.

7.1 Differential jet energy corrections

In this section, the differential estimation of the jet energy corrections is described. The differentiation performs an estimation of the jet energy correction in bins of the transverse momentum or pseudo-rapidity of the jet, or in bins of the number of muons in the b quark jet. The method to estimate the jet energy corrections in a certain bin is identical to the estimation of the inclusive jet energy corrections.

7.1.1 Dependence on muons in b quark jets

Jets from b quarks may contain muons and neutrinos from the leptonic decay of heavy hadrons produced in the fragmentation of the b quark. Since the neutrinos escape detection and muons deposit relatively little of their energy in the calorimeters different energy corrections are expected depending on the presence of a muon in the b quark jet. To estimate the dependency of the residual jet energy corrections on the presence of a muon, the events are categorized in different samples according to the number of muons found within $\Delta R < 0.5$ of the b quark jet axis. The method discussed in Chapter 6 is applied on each sample. The resulting estimations and expected corrections are presented in Table 7.1. As a reference also the inclusive estimation of the b quark jet energy correction using the full $t\bar{t}$ sample is given. There is a clear difference between the expected jet energy corrections in the sample of events without a muon and the corrections in the sample with at least one muon in the vicinity of a b quark jet. The method is able to estimate the jet energy corrections correctly, because no bias is observed within the statistical uncertainty.

	$\Delta E_{b,\text{exp}}$ (%)	$\Delta E_{b,\text{est}}$ (%)	$\Delta E_{b,\text{est}} - \Delta E_{b,\text{exp}}$ (%)
inclusive	-3.08±0.55	-2.42±0.31	+0.66±0.63
no muon in b jet	-4.25±0.62	-3.88±0.34	+0.37±0.71
1 muon in b jet	+4.65±2.74	+3.11±0.99	-1.54±2.91
≥1 muon in b jet	+6.41±2.28	+2.76±0.83	-3.65±2.43

Table 7.1: Expected and estimated b quark jet energy corrections as a function of the number of muons in the b quark jet.

When the method to estimate the jet energy corrections is robust, no difference is expected for the estimation of the light quark jet energy corrections depending on the presence of a muon in the b quark jet. In Table 7.2 the jet energy corrections for the light quark jets coming from the W boson are summarized for the different

samples of events. Both the expected and estimated corrections are compatible with each other and for the different samples. The results presented reflect the uncertainty

	$\Delta E_{l,\text{exp}}$ (%)	$\Delta E_{l,\text{est}}$ (%)	$\Delta E_{l,\text{est}} - \Delta E_{l,\text{exp}}$ (%)
inclusive	-12.32±0.29	-11.83±0.21	+0.49±0.36
no muon in b jet	-12.19±0.29	-11.77±0.23	+0.42±0.37
1 muon in b jet	-11.55±0.83	-10.84±0.66	+0.71±1.06
≥1 muon in b jet	-11.92±0.72	-10.97±0.59	+0.95±0.93

Table 7.2: Expected and estimated light quark jet energy corrections as a function of the number of muons in the b quark jet.

using a dataset with an integrated luminosity of $2.3\,\text{fb}^{-1}$ at 10 TeV. Additionally, the uncertainty on the estimated value is corrected for the non-unity width of the pull using the widths determined in Section 6.5 for the inclusive estimator. It is reasonable to assume this invariance of the pull because the pull is not expected to depend on the possible presence of muons in the b quark jet.

7.1.2 Dependence on jet kinematics

Since the input jet energies used for the analysis are corrected for the dependency on the pseudo-rapidity (L2) and the transverse momentum (L3) of the jet, we expect the residual jet energy corrections to be flat when performing the method as a function of these kinematic variables. To check this hypothesis, the jet energy corrections are estimated in bins of the pseudo-rapidity $|\eta|$ and in bins of the transverse momentum p_T.

According to their pseudo-rapidity $|\eta|$, the three jets of the hadronic branch of the top quark decay $\text{t} \to \text{bW} \to \text{bq}\bar{\text{q}}$ are each put in a certain $|\eta|$ bin. The kinematic fit with the W boson and top quark mass constraint is performed for each event, resulting in a $\chi^2(\Delta E_b(|\eta_b|), \Delta E_q(|\eta_q|), \Delta E_{\bar{q}}(|\eta_{\bar{q}}|))$ for each event.

Consider first the estimation of the b quark jet energy correction in bins i of the b quark jet pseudo-rapidity, $|\eta_b|_i$. The χ^2 values are summed over all events $j \in \{1, \ldots, N\}$ if the pseudo-rapidity of the b quark jet of event j, $|\eta_b^j|$, is part of the $|\eta_b|_i$ bin, or $|\eta_b|_i^{\text{min}} \leq |\eta_b^j| \leq |\eta_b|_i^{\text{max}}$. The summation is independent of the pseudo-rapidities $|\eta_q^j|$ and $|\eta_{\bar{q}}^j|$ of the two light quark jets. Additionally in the sum, the two light quark jet energy corrections are required to be equal, or $\Delta E_q = \Delta E_{\bar{q}}$. Hence,

$$\sum_{j=1}^{N} \chi_j^2 \left(\Delta E_b(|\eta_b^j|_i), \Delta E_q = \Delta E_{\bar{q}}\right) = \chi_{\text{tot}}^2 \left(\Delta E_b(|\eta_b|_i), \Delta E_l\right). \tag{7.1}$$

For each bin i of the pseudo-rapidity the value $(\Delta E_l)_{i,\text{min}}$ is determined from the 2D minimum of $\chi_{\text{tot}}^2 \left(\Delta E_b(|\eta_b|_i), \Delta E_l\right)$ and from the 1D function

$$\chi_{\text{tot}}^2 \left(\Delta E_b(|\eta_b|_i), (\Delta E_l)_{i,\text{min}}\right), \tag{7.2}$$

the estimator for the b quark jet energy correction in bin i, $\Delta E_{b,\mathrm{est}}(|\eta_b|_i)$ is obtained.

Similarly, the estimator for the light quark jet energy correction in each bin of the pseudo-rapidity is obtained by summing the χ^2 values over all events $j \in \{1, \ldots, N\}$ if the pseudo-rapidity of the first light quark jet q of event j, $|\eta_q|^j$, is part of the $|\eta_l|_i$ bin, or $|\eta_l|_i^{\min} \leq |\eta_q^j| \leq |\eta_l|_i^{\max}$. The summation is independent of the pseudo-rapidities $|\eta_b^j|$ and $|\eta_{\bar{q}}^j|$ of the b quark jet and the second light quark jet $\bar{q}$. Hence,

$$\sum_{j=1}^{N} \chi_j^2 \left(\Delta E_b, \Delta E_q(|\eta_q^j|_i) = \Delta E_{\bar{q}}\right) = \chi^2_{\mathrm{tot,q}} \left(\Delta E_b, \Delta E_l(|\eta_q|_i)\right). \tag{7.3}$$

The same procedure is followed for the second light quark jet $\bar{q}$ of event j, which is part of the $|\eta_l|_k$ bin, when $|\eta_l|_k^{\min} \leq |\eta_{\bar{q}}^j| \leq |\eta_l|_k^{\max}$. The summation is performed independent of the pseudo-rapidities $|\eta_b^j|$ and $|\eta_q^j|$ of the b quark jet and the first light quark jet q. And therefore,

$$\sum_{j=1}^{N} \chi_j^2 \left(\Delta E_b, \Delta E_q = \Delta E_{\bar{q}}(|\eta_{\bar{q}}^j|_k)\right) = \chi^2_{\mathrm{tot,\bar{q}}} \left(\Delta E_b, \Delta E_l(|\eta_{\bar{q}}|_k)\right) \tag{7.4}$$

is obtained. For each bin i of the pseudo-rapidity the values $\chi^2_{\mathrm{tot,q}} \left(\Delta E_b, \Delta E_l(|\eta_q|_i)\right)$ and $\chi^2_{\mathrm{tot,\bar{q}}} \left(\Delta E_b, \Delta E_l(|\eta_{\bar{q}}|_i)\right)$ are summed and the sum is denoted as

$$\chi^2_{\mathrm{tot}} \left(\Delta E_b, \Delta E_l(|\eta_l|_i)\right). \tag{7.5}$$

For each bin i of the pseudo-rapidity the value $(\Delta E_b)_{i,\min}$ is determined from the 2D minimum of $\chi^2_{\mathrm{tot}} \left(\Delta E_b, \Delta E_l(|\eta_l|_i)\right)$ and from the 1D function

$$\chi^2_{\mathrm{tot}} \left((\Delta E_b)_{i,\min}, \Delta E_l(|\eta_l|_i)\right), \tag{7.6}$$

the estimator for the light quark jet energy correction in bin i, $\Delta E_{l,\mathrm{est}}(|\eta_l|_i)$ is obtained.

The procedure to estimate the jet energy corrections in bins of the transverse momentum p_T of the jets is completely analogue to the procedure for the estimation of the jet energy correction in bins of the pseudo-rapidity explained above.

Differential corrections with respect to the pseudo-rapidity

The upper plots of Figure 7.1 show the estimation of the jet energy corrections as a function of the pseudo-rapidity of the b and light quark jets. As expected since the L2 corrections were applied, the residual jet energy corrections are flat with respect to the pseudo-rapidity. The large fluctuations in the expected correction is attributed to the instability of the Gaussian fit as mentioned in Section 6.4.

The origin of the non-unity width of the pull in Section 6.5 is partially related to an underestimation of the resolutions on the jet parameters. It is assumed that the underestimation is independent of the pseudo-rapidity of the jets, therefore the uncertainties on the estimated values are rescaled using the width of the inclusive estimation of the jet energy corrections. The uncertainties on the estimated values correspond to the statistical uncertainty using a data set with an integrated luminosity of $2.3\,\mathrm{fb}^{-1}$.

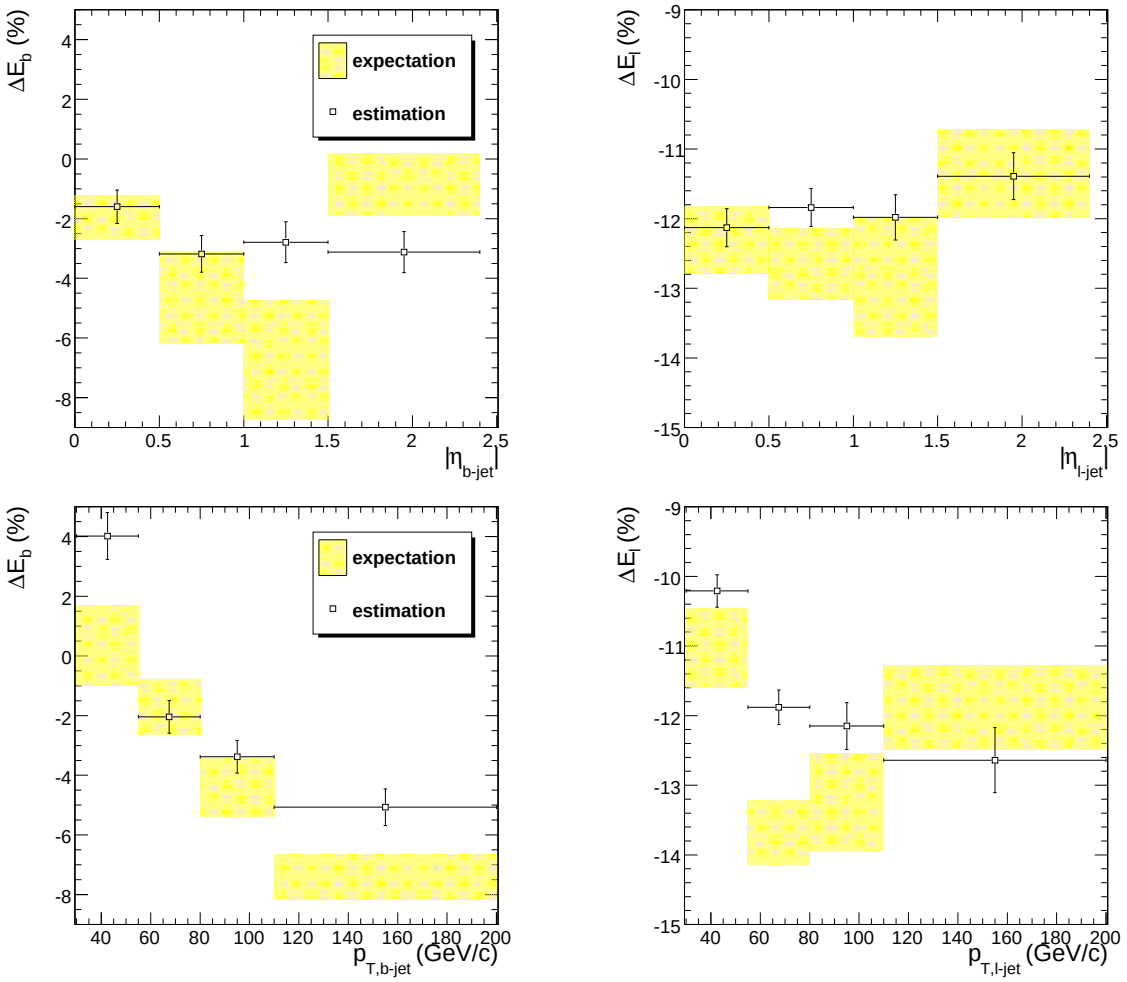


Figure 7.1: Expected and estimated jet energy corrections as a function of the jet pseudo-rapidity (upper plots) and jet transverse momentum (lower plots) for b (left) and light (right) quark jets.

Differential corrections with respect to the transverse momentum

One can also perform the method in bins of the transverse momentum of the jets. The residual p_T dependent jet energy corrections are shown in the lower plots of Figure 7.1. The jet energy corrections are expected not to depend on the transverse momentum as the L3 jet energy corrections were applied. However, while this is reasonably true within the statistical uncertainties for the light quark jet energy corrections, a dependency of the corrections for b quark jets is observed. The estimated corrections are consistent with the expected values. The origin of the dependency is related to the different flavour mixture of the di-jet sample from which the L3 corrections were derived compared to the $t\bar{t}$ sample. A similar assumption as for the pseudo-rapidity dependency is made, namely that the underestimation of the resolutions on the jet parameters does not depend on the transverse momentum of the jets. The uncertainties on the estimation are therefore rescaled using the width of the inclusive estimation of the jet energy corrections.

It is important to know if the method works in case a dependency of the corrections on the transverse momentum of the light quark jets would exist. Therefore, using only the semi-muonic $t\bar{t}$ events, the four-momenta of the light quark jets are initially rescaled according to some arbitrary p_T dependent corrections. The method is then applied on the jets with the rescaled four-momenta to check if an estimation of the jet energy corrections as a function of the transverse momentum of the light quark jets is possible. Figure 7.2 shows the expected and estimated jet energy corrections.

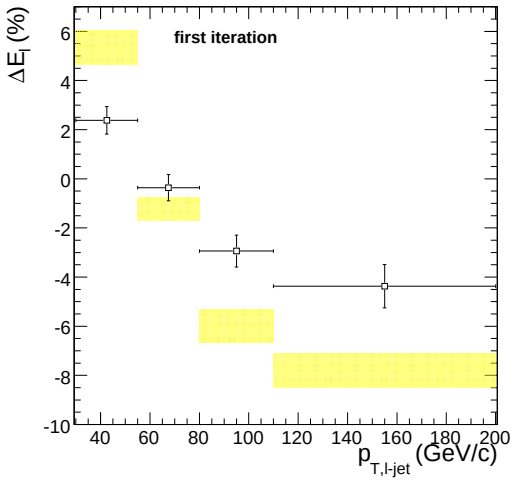


Figure 7.2: Expected and estimated jet energy corrections after an arbitrary p_T dependent rescaling of the jet four-momenta as a function of the jet transverse momentum for light quark jets.

The expected and estimated values are not entirely consistent, therefore one can apply the estimated jet energy corrections to rescale the jet four-momenta and repeat the method. By iterating several times the expected and estimated jet energy corrections become consistent. The second and third iteration are shown in Figure 7.3. The jet energy corrections estimated in the different iterations can then be combined to obtain the total jet energy correction for each p_T-bin.

Systematic uncertainty

The effect of the event simulation parameters on the measured jet energy corrections is checked by altering the values of the parameters and by comparing the obtained corrections with those of a nominal sample. It is interesting to know if the expected values of the residual jet energy corrections change with respect to the nominal values. On the other hand, independent of a change in the expected values, the developed method is performing well if the estimation of the jet energy corrections is unbiased with respect to these expected corrections. These two aspects are studied by performing the $|\eta|$ and p_T dependent estimation for various event samples with altered event simulation parameters. Within the statistical uncertainties, no significant difference is found in the expected values of the p_T and $|\eta|$ dependent jet energy corrections, both for light

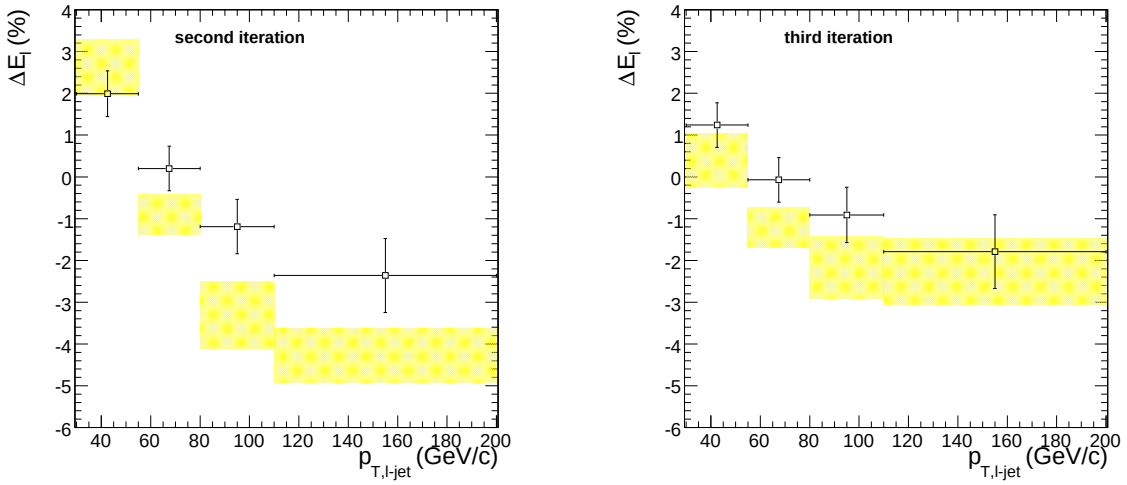


Figure 7.3: Expected and estimated light quark jet energy corrections for an arbitrary p_T dependent rescaling of the jet four-momenta after the second and third iterations.

and b quark jets. Table 7.3 summarizes the bias for the different p_T-bins of the b and light quark jets for various event samples. In most cases the bias is compatible with zero, although a few outliers with three to four standard deviations are visible. However, these deviations can be resolved by iterating several times. The statistical uncertainties on both the expected and estimated values and therefore on the bias are rather large. As a conservative choice, the largest statistical uncertainty in each p_T-bin is taken as the systematic uncertainty. These values are also given in the table.

In Table 7.4 the bias for the measurements in different $|\eta|$-bins of the b and light quark jets is given. Also in this case the statistical uncertainty is the limiting factor to determine the uncertainty related to a possible systematic effect due to the event simulation. The bias on the $|\eta|$ dependent jet energy correction estimation is for most bins and most event samples compatible with zero, although a few outliers can be found for the same reason as explained above. The largest statistical uncertainty in each bin is again taken as the systematic uncertainty.

Smeared jet energies

Another possible systematic effect may come from a worse measurement of the jets for the energy resolution. To test this effect, the jet energies and momenta were smeared according to a Gaussian distributed random number with a variance of $(1.3\sigma_E)^2-(\sigma_E)^2$, where σ_E is the assumed energy resolution. The event selection, with the exception of the pre-selection, is performed using the new smeared jet four-momenta. The method was repeated to estimate the p_T dependent jet energy corrections for b and light quark jets. The resulting plots are shown in Figure 7.4. The upper plots of the figure show the expected and estimated jet energy corrections after smearing the jet energies. Due to the large uncertainties and the instability of the Gaussian fit, the expected jet energy corrections are found to be consistent for most p_T-bins before and after the smearing. For some bins a bias of two to four standard deviations can be observed, these larger

b jets	$30 < p_T < 55$	$55 < p_T < 80$	$80 < p_T < 110$	$110 < p_T < 200$
nominal `PYTHIA`	+0.23±0.74	+1.07±0.60	+1.69±0.52	+0.20±0.49
large ISR/FSR	+1.86±1.52	-0.01±0.86	+1.10±0.70	-0.45±0.81
small ISR/FSR	+0.41±0.96	+0.91±0.68	+1.48±0.63	+0.75±0.74
nominal `MadGraph/MadEvent`	-1.12±0.82	+0.03±0.73	+0.66±0.59	+1.85±0.86
Q^2 up	+0.16±1.66	+0.90±1.00	+0.41±0.84	+0.05±0.94
Q^2 down	+1.22±2.23	+0.31±0.99	+0.65±0.88	+1.24±0.99
$E_T^{min} = 10\,\text{GeV}$	+1.57±1.96	+0.94±0.98	+1.44±1.02	+1.60±0.86
$E_T^{min} = 40\,\text{GeV}$	-0.42±1.23	+0.78±0.80	+1.28±0.76	-0.04±0.85
"total" assigned systematic uncertainty	≤ 2.2	≤ 1.0	≤ 1.0	≤ 1.0
light jets	$30 < p_T < 55$	$55 < p_T < 80$	$80 < p_T < 110$	$110 < p_T < 200$
nominal `PYTHIA`	+0.29±0.36	+1.30±0.33	+1.00±0.37	+0.59±0.44
large ISR/FSR	+0.55±0.58	+1.94±0.54	+0.99±0.59	+0.15±0.60
small ISR/FSR	+1.31±0.51	+1.36±0.44	+1.33±0.47	-0.10±0.57
nominal `MadGraph/MadEvent`	-0.08±0.42	+1.40±0.50	+0.75±0.44	+0.16±0.53
Q^2 up	+0.32±0.61	-0.20±0.61	+0.78±0.61	+0.46±0.73
Q^2 down	+0.70±0.63	+1.01±0.67	+1.33±0.76	+0.06±0.85
$E_T^{min} = 10\,\text{GeV}$	+1.19±0.86	+1.09±0.55	-0.20±0.78	+0.17±0.78
$E_T^{min} = 40\,\text{GeV}$	+1.16±0.58	+1.38±0.49	+0.80±0.56	-0.24±0.72
"total" assigned systematic uncertainty	≤ 0.9	≤ 0.7	≤ 0.8	≤ 0.8

Table 7.3: The bias $\Delta E_{\text{est}} - \Delta E_{\text{exp}}$ and the assigned systematic uncertainty on the estimated values in % in the different p_T-bins (GeV/c) of the b or light quark jet for the various event samples in the study of the systematic effects.

b jets	$0 < \|\eta\| < 0.5$	$0.5 < \|\eta\| < 1.0$	$1.0 < \|\eta\| < 1.5$	$1.5 < \|\eta\| < 2.4$
nominal PYTHIA	+0.28±0.49	+0.54±0.61	+0.40±0.81	-0.50±0.80
large ISR/FSR	+0.13±0.75	+0.76±0.84	-0.80±1.08	-1.84±0.90
small ISR/FSR	+0.43±0.69	-0.01±0.68	-0.11±0.98	-0.81±0.88
nominal MadGraph/MadEvent	-0.19±0.73	+0.57±0.74	+0.74±1.00	-0.90±0.84
Q^2 up	-1.30±0.94	-0.69±0.97	+0.31±1.26	+1.16±1.91
Q^2 down	-0.40±0.89	-0.20±1.17	-0.27±1.28	+0.19±1.03
$E_T^{min} = 10\,\text{GeV}$	-0.11±0.89	+1.09±1.05	+0.50±1.61	+0.66±1.24
$E_T^{min} = 40\,\text{GeV}$	-0.27±0.84	+1.00±0.84	-0.93±1.26	-2.03±0.96
"total" assigned systematic uncertainty	≤0.9	≤ 1.2	≤ 1.6	≤ 1.9

light jets	$0 < \|\eta\| < 0.5$	$0.5 < \|\eta\| < 1.0$	$1.0 < \|\eta\| < 1.5$	$1.5 < \|\eta\| < 2.4$
nominal PYTHIA	-0.12±0.32	+0.80±0.32	+1.84±0.43	+0.11±0.42
large ISR/FSR	+0.68±0.53	+0.77±0.52	+1.80±0.69	+0.72±0.61
small ISR/FSR	-0.01±0.44	+1.63±0.45	+1.91±0.59	-0.40±0.57
nominal MadGraph/MadEvent	-0.82±0.41	+0.61±0.48	+2.12±0.58	-0.61±0.45
Q^2 up	-2.00±0.58	+1.15±0.54	+1.01±0.81	-0.33±0.69
Q^2 down	-0.15±0.61	+0.64±0.76	+2.65±0.79	-0.21±0.64
$E_T^{min} = 10\,\text{GeV}$	-0.27±0.62	+0.06±0.59	+2.53±0.77	+0.54±0.79
$E_T^{min} = 40\,\text{GeV}$	-0.61±0.48	-0.23±0.51	+2.21±0.77	+0.04±0.54
"total" assigned systematic uncertainty	≤ 0.6	≤ 0.8	≤ 0.8	≤ 0.8

Table 7.4: The bias $\Delta E_{\text{est}} - \Delta E_{\text{exp}}$ and the assigned systematic uncertainty on the estimated values in % in the different $|\eta|$-bins of the b or light quark jet for the various event samples in the study of the systematic effects.

deviations can however be resolved by iterating several times. The estimated jet energy

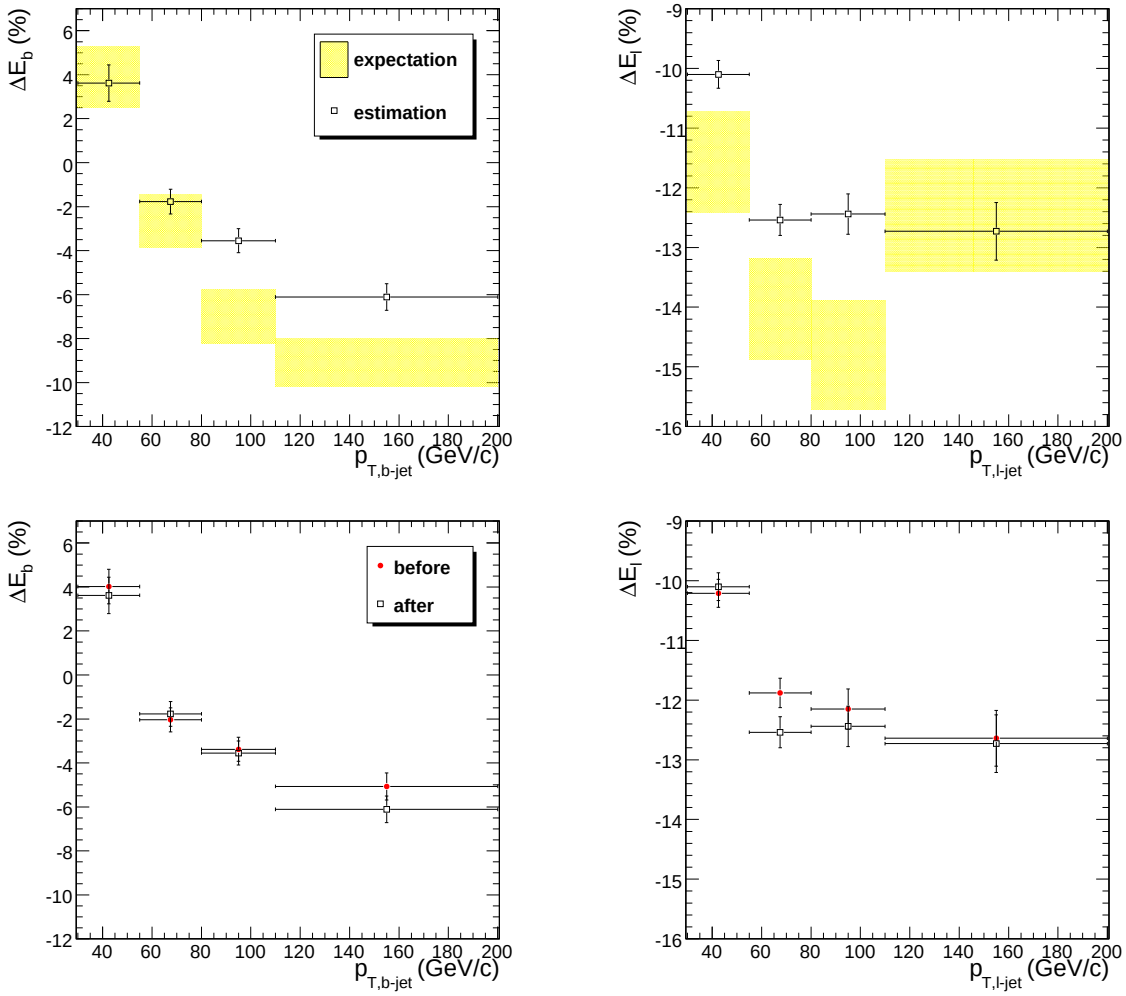


Figure 7.4: In the upper plots, the expected and estimated p_T dependent jet energy corrections for b (left) and light (right) quark jets is shown, after a smearing on the jet energies was applied. The lower plots show the estimated corrections before and after the smearing was applied.

corrections before and after the smearing are consistent as shown in the lower plots in Figure 7.4. No systematic uncertainty due to the smearing of the jet energies is accounted for.

Summary of the systematic uncertainty

A conservative estimation of the systematic uncertainty related to the event simulation aspects is given in Tables 7.3 and 7.4 for respectively the estimation of the p_T and $|\eta|$ dependent jet energy corrections. Nevertheless, the most important systematic uncertainty on the estimation of the b quark jet energy correction is related to the precision of the top quark mass measurement at the Tevatron collider because this value

is used in the kinematic fit to determine the jet energy corrections. This systematic uncertainty was estimated to be 1.3 % (cfr. Chapter 6). Hence, in each p_T- and $|\eta|$-bin of the b quark jet this systematic uncertainty on the jet energy correction should be taken into account. In Table 7.5 the systematic uncertainty related to the precision on the top quark mass and the systematic uncertainty related to the event simulation are combined for the b quark jet energy correction measurement. Also the systematic uncertainties for the different $|\eta|$- and p_T-bins of the light quark jet are summarized. For the b quark jets, the lower number is the systematic uncertainty related to the

	$0 < \|\eta\| < 0.5$	$0.5 < \|\eta\| < 1.0$	$1.0 < \|\eta\| < 1.5$	$1.5 < \|\eta\| < 2.4$
b jets	1.3 - 1.6	1.3 - 1.8	1.3 - 2.1	1.3 - 2.3
light jets	≤ 0.6	≤ 0.8	≤ 0.8	≤ 0.8
	$30 < p_T < 55$	$55 < p_T < 80$	$80 < p_T < 110$	$110 < p_T < 200$
b jets	1.3 - 2.6	1.3 - 1.6	1.3 - 1.6	1.3 - 1.6
light jets	≤ 0.9	≤ 0.7	≤ 0.8	≤ 0.8

Table 7.5: The absolute systematic uncertainty in % on the estimated values in the different $|\eta|$ and p_T-bins of the jet.

top quark mass, while the upper number is obtained by quadratically summing the former with the systematic uncertainty due to the event simulation aspects. The real systematic uncertainty will be somewhere in between those numbers.

Other jet reconstruction algorithms

The estimation of the jet energy corrections in bins of the pseudo-rapidity and transverse momentum of the jets is also performed using another jet reconstruction algorithm. The inclusive longitudinally invariant k_T algorithm is used to reconstruct the jets and the measurement is performed in $|\eta|$- and p_T-bins. The upper plots in Figure 7.5 show the expected and estimated $|\eta|$ dependent jet energy corrections for b and light quark jets. Expected and estimated corrections are consistent. Also the p_T dependent jet energy corrections are determined for this algorithm as shown in the lower plots of Figure 7.5. Also in this case the expectations and estimations agree.

The method is expected to provide corrections for any jet reconstruction algorithm because no algorithm dependent cuts are applied. However, for jet reconstruction

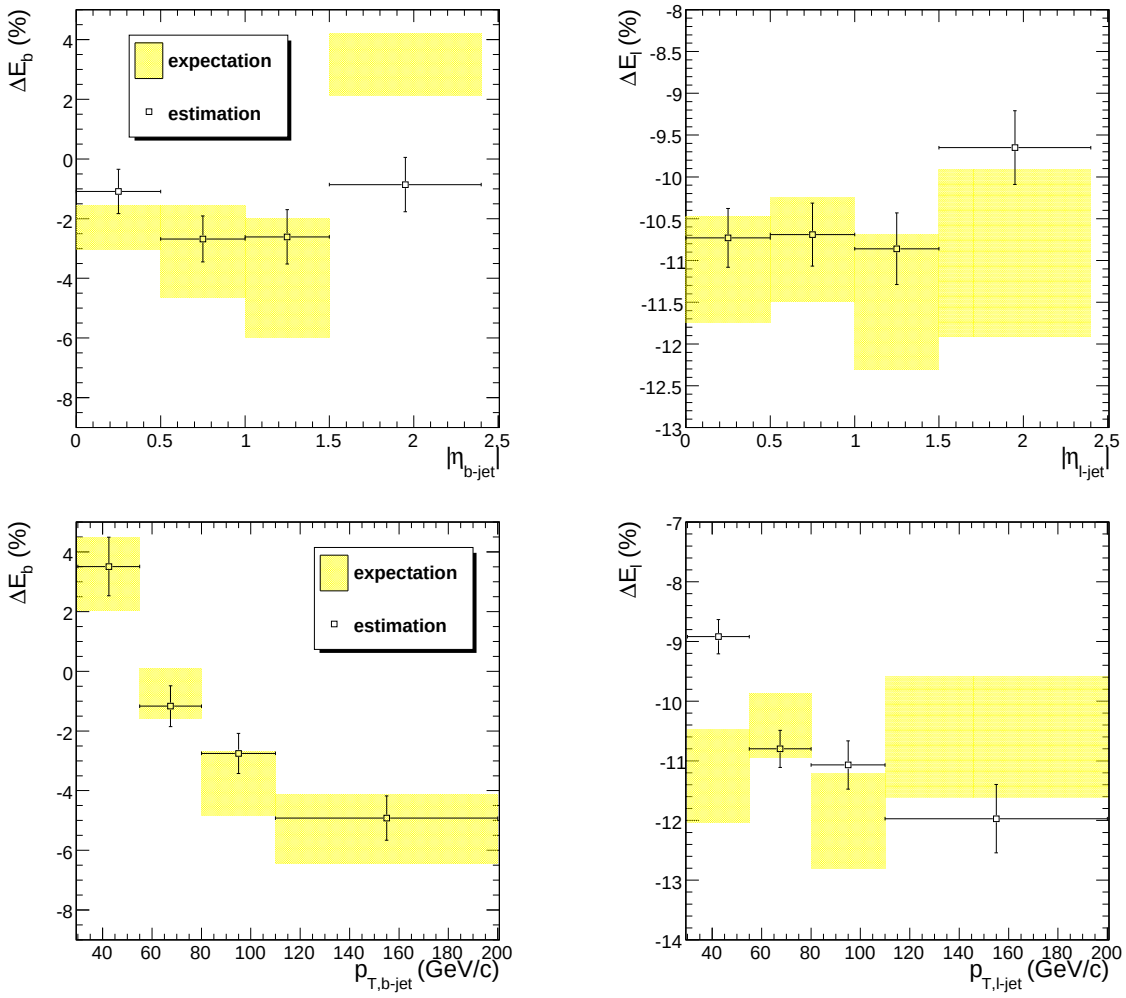


Figure 7.5: Expected and estimated $|\eta|$ (upper plots) and p_T (lower plots) dependent jet energy corrections for b (left) and light (right) quark jets reconstructed with the k_T algorithm.

algorithms using no splitting and merging, with as an example the Iterative Cone algorithm, overlapping jets may introduce a bias on the measurement. One should thus be careful and revisit the method to measure the jet energy corrections if overlapping jets are intrinsic to the jet reconstruction algorithm.

7.2 Factorized residual corrections

In Section 4.1.5 the factorized approach adopted by the CMS collaboration to estimate the jet energy corrections was discussed. Using $\mathrm{t\bar{t}}$ events a similar approach can be envisaged for the estimation of the b and light quark jet energy corrections separately. The strategy is to perform an inclusive estimation of the jet energy corrections, which can then be applied on the four-momenta of the jets obtaining $\mathrm{L}^{\mathrm{top}}_{\mathrm{incl}}$ corrected jets. Using the $\mathrm{L}^{\mathrm{top}}_{\mathrm{incl}}$ corrected jet four-momenta, in a next step the $|\eta|$ dependent corrections

can be determined, as shown in Figure 7.6. These $|\eta|$ dependent jet energy correc-

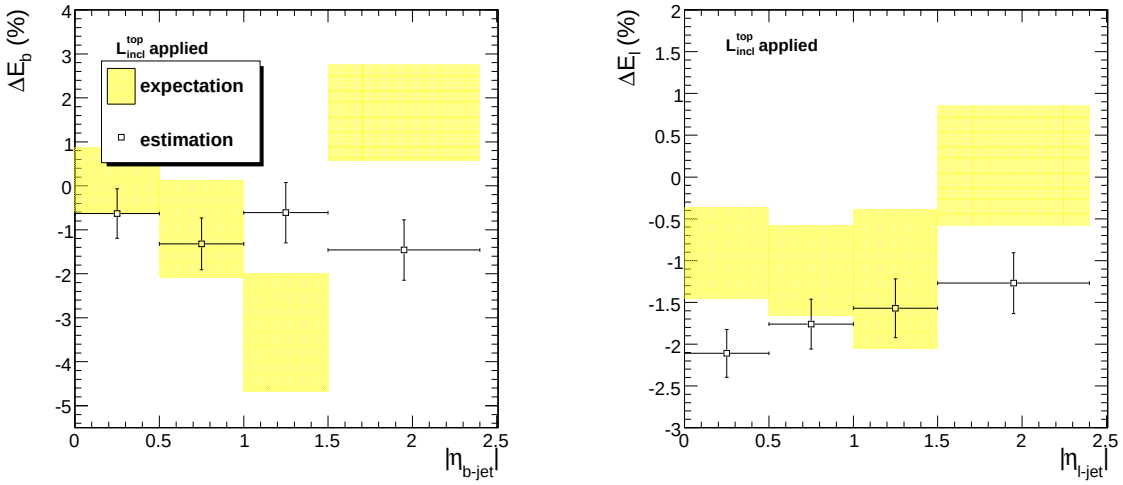


Figure 7.6: Expected and estimated $|\eta|$ dependent jet energy corrections for b (left) and light (right) quark jets when using $\mathrm{L}^{\mathrm{top}}_{\mathrm{incl}}$ corrected jets.

tions for light and b quark jets can then be used to rescale the four-momenta of the jets which are used as an input for the estimation. The corrected jets are denoted as $\mathrm{L}^{\mathrm{top}}_{\mathrm{incl}} + \mathrm{L2}^{\mathrm{top}}$ corrected jets. In a next step, the $\mathrm{L}^{\mathrm{top}}_{\mathrm{incl}} + \mathrm{L2}^{\mathrm{top}}$ corrected jets are used to estimate the p_T dependent jet energy corrections, shown in Figure 7.7. After applying

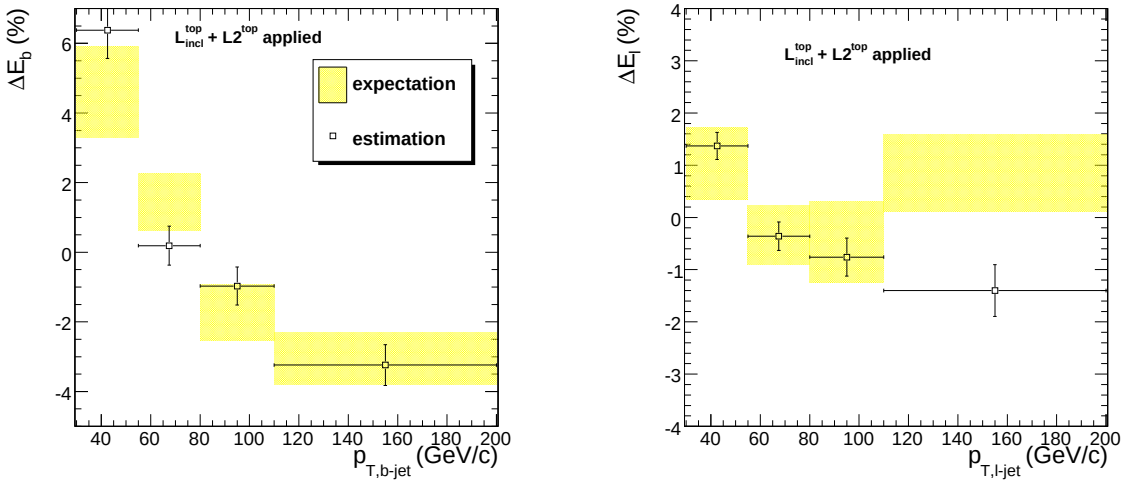


Figure 7.7: Expected and estimated p_T dependent jet energy corrections for b (left) and light (right) quark jets when using $\mathrm{L}^{\mathrm{top}}_{\mathrm{incl}} + \mathrm{L2}^{\mathrm{top}}$ corrected jets.

the estimated corrections, $\mathrm{L}^{\mathrm{top}}_{\mathrm{incl}} + \mathrm{L2}^{\mathrm{top}} + \mathrm{L3}^{\mathrm{top}}$ corrected jets are obtained. They additionally differentiate in flavour (L5) and correct to the parton level (L7), hence these correction levels are implicitly included.

After these iterations, the residual jet energy corrections with respect to $|\eta|$ and p_T should be flat around zero. The jet energy corrections obtained when using the $\mathrm{L}^{\mathrm{top}}_{\mathrm{incl}} + \mathrm{L2}^{\mathrm{top}} + \mathrm{L3}^{\mathrm{top}}$ corrected jets are shown in the upper plots of Figure 7.8 for the

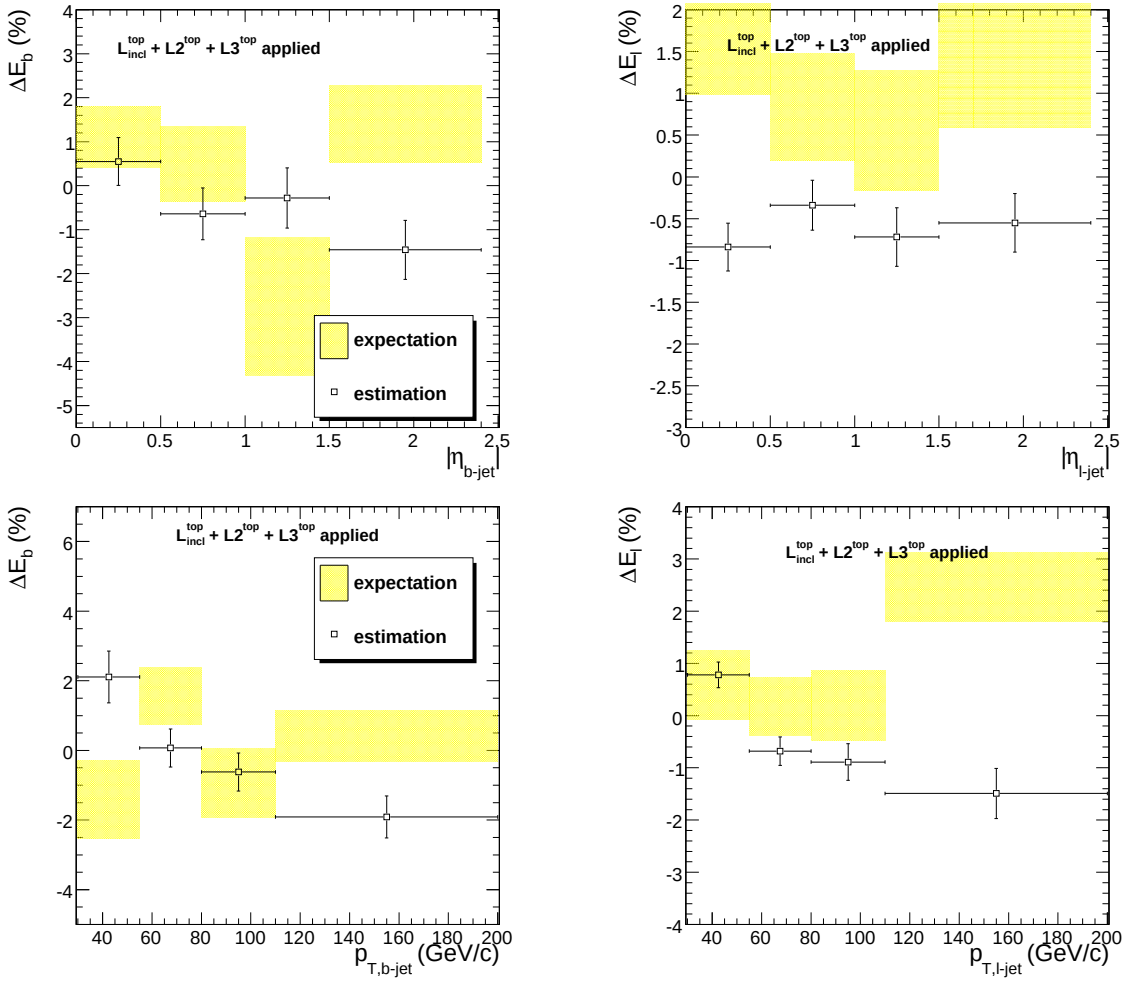


Figure 7.8: Expected and estimated $|\eta|$ (upper plots) and p_T (lower plots) dependent jet energy corrections for b (left) and light (right) quark jets when using $\mathrm{L}^{\mathrm{top}}_{\mathrm{incl}} + \mathrm{L2}^{\mathrm{top}} + \mathrm{L3}^{\mathrm{top}}$ corrected jets.

$|\eta|$ dependence and in the lower plots of Figure 7.8 for the p_T dependence. The jet energy corrections are more or less flat with respect to the pseudo-rapidity or transverse momentum of the jets. Possible biases or deviations from zero can be resolved by iterating a few times at each stage of the factorized approach.

7.3 Validation of the method

In Section 6.1 the shift on the reconstructed W boson and top quark mass spectra is discussed. It is explained that this shift is related to the presence of residual jet energy corrections. The method presented in this thesis estimates these residual corrections. It is studied if the reconstructed shift disappears when the estimated $\mathrm{L}^{\mathrm{top}}_{\mathrm{incl}} + \mathrm{L2}^{\mathrm{top}} + \mathrm{L3}^{\mathrm{top}}$ jet energy corrections are applied before the W boson and top quark mass spectra are reconstructed. Figure 7.9 shows the two dimensional W boson and top quark mass spectrum before and after the inclusive, $|\eta|$ and p_T dependent jet energy corrections

are applied. Although the peak of the two dimensional reconstructed W boson and top

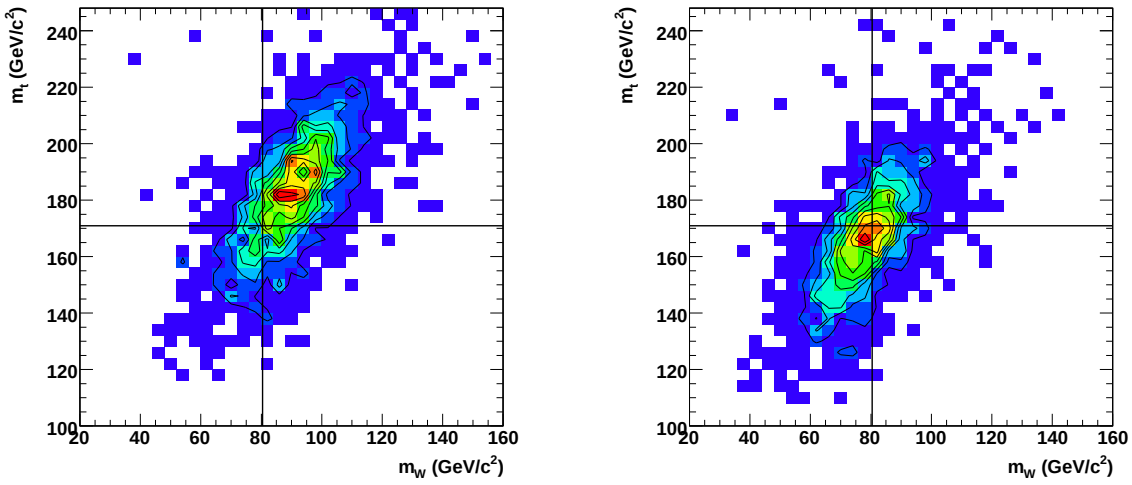


Figure 7.9: Two dimensional W boson and top quark mass spectrum before (left) and after the inclusive, $|\eta|$ and p_T dependent jet energy corrections are applied (right).

quark mass spectrum is close to the generated values after the estimated jet energy corrections are applied, there is still a difference. This difference is related to the statistical precision with which the jet energy corrections were estimated. One should also keep in mind that a possible bias disappears when several iterations are performed as discussed in Section 6.5. Therefore, it is concluded that the method to estimate the jet energy corrections is valid.

Because not only an inclusive correction is applied, but also a correction based on the properties of the jet, it is expected that the resolution on the reconstructed W boson and top quark mass after applying the differential corrections is improved. A significant improvement is observed for the mass resolutions obtained with the $\mathrm{L}^{\mathrm{top}}_{\mathrm{incl}} + \mathrm{L2}^{\mathrm{top}} + \mathrm{L3}^{\mathrm{top}}$ corrected jets compared to the event kinematics before applying the method. An improvement of 14 % in the resolution of the reconstructed top quark mass is obtained, while for the resolution of the W boson mass the improvement is 15 %.

7.4 Performance of the method and comparison with other methods

Figure 7.10 shows the expected statistical uncertainty using a dataset with an integrated luminosity of $1\,\mathrm{fb}^{-1}$ together with the systematic uncertainty as a function of either the p_T or $|\eta|$ of the jet. For $1\,\mathrm{fb}^{-1}$ the total uncertainty on the estimation of the b quark jet energy corrections is clearly dominated by the current precision on the top quark mass measurement at the Tevatron collider. The statistical uncertainty is around 1 % in each p_T or $|\eta|$ bin for b quark jets, while it is around 0.5 % in each p_T or $|\eta|$ bin for the light quark jets. Furthermore, since the statistical uncertainty is smaller than the conservatively expected systematic uncertainty, with more than $1\,\mathrm{fb}^{-1}$ of data, the number of p_T- or $|\eta|$-bins can be increased.

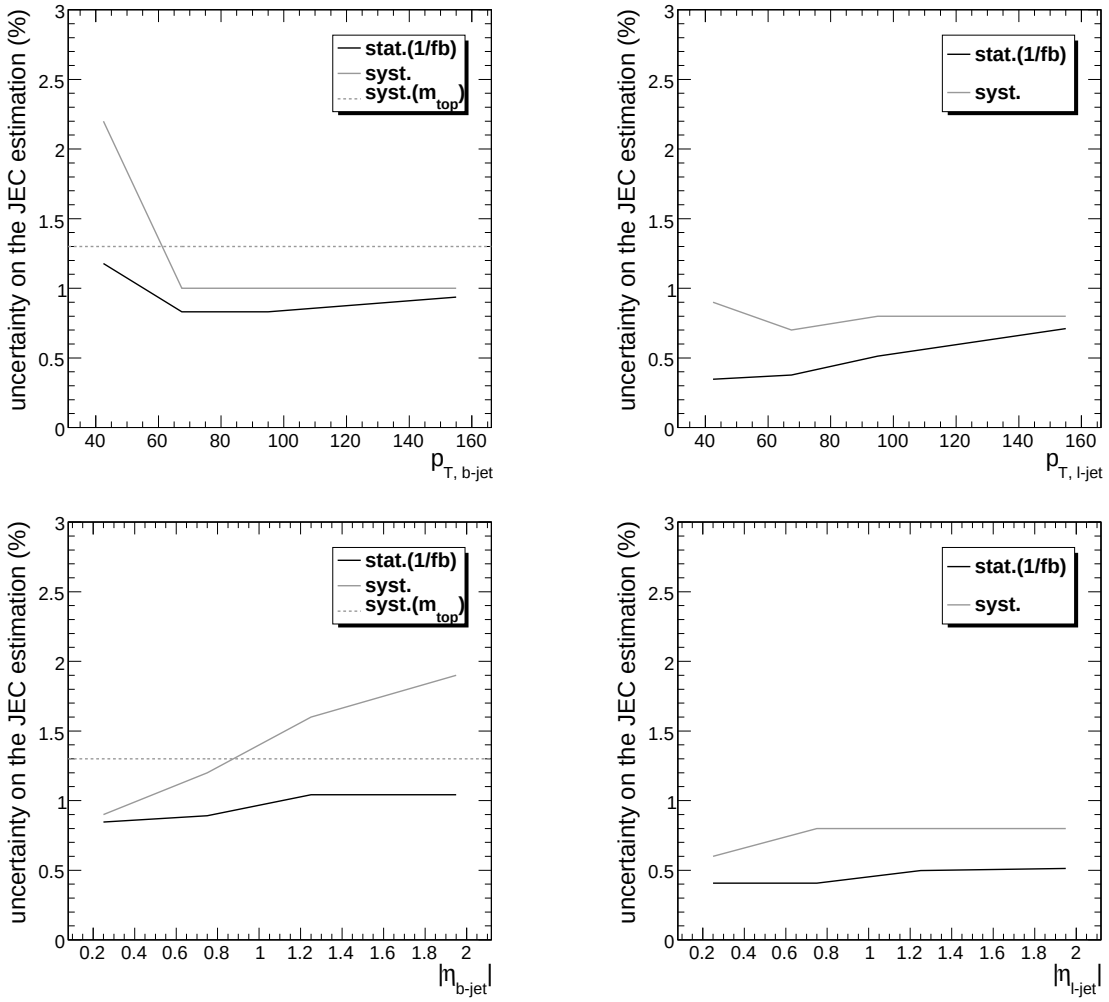


Figure 7.10: The statistical uncertainty for an integrated luminosity of $1\,\mathrm{fb}^{-1}$ and the combined systematic uncertainty on the estimated p_T (upper plots) and $|\eta|$ (lower plots) dependent jet energy corrections for b (left) and light (right) quark jets.

One can compare the developed method with the methods described in Section 4.1.5 to determine the η (L2) and p_T (L3) dependent corrections. However, it is important to realize that there are two important differences. The jet energy corrections for the methods discussed in Section 4.1.5 aim in a correction back to the particle or GenJet level, while the current method corrects back to the parton level. Secondly, a differentiation is obtained in the current method for b quark jets and a flavour mixture of quark jets from the decay of the W boson, while the other methods determine inclusive flavour corrections that are depending on the specific flavour mixture of the sample.

With an integrated luminosity of $0.5\,\mathrm{pb}^{-1}$, the η dependent jet energy correction can be determined from di-jet events to better than 1 %. A systematic uncertainty up to 3 % in the forward region and for jets with a low transverse momentum is expected [126]. Obviously, using top quark events, only a validation of these corrections

can be envisaged because an integrated luminosity of $1\,\mathrm{fb}^{-1}$ is needed to obtain the same precision.

The p_T dependent jet energy correction can be obtained either from p_T balance in γ+jet or Z+jet events. Using the γ+jet events in $100\,\mathrm{pb}^{-1}$ of data, one can determine the jet energy correction up to $600\,\mathrm{GeV/c}$ with a statistical uncertainty below 1 % for jets with $40 < p_T < 100$ GeV/c and about 1.5 % for jets with a transverse momentum lower than $40\,\mathrm{GeV/c}$ or $100 < p_T < 200\,\mathrm{GeV/c}$. The systematic uncertainty however is estimated to be about 5 % for jets with $p_T = 40\,\mathrm{GeV/c}$ and are mainly related to topological cuts and the flavour mapping of the different flavour composition between jets in γ+jet and multi-jets events [127]. With top quark events, the systematic uncertainty is clearly reduced, but the statistical uncertainty is significantly higher. With $1\,\mathrm{fb}^{-1}$ of data, the total uncertainty obtained with γ+jet events is clearly dominated by the systematic uncertainty, while the total uncertainty obtained with top quark events is well below the 5 % level. It should however be mentioned that more p_T-bins can be used in the method which makes use of γ+jet events. The same is true when Z+jet events are used to determine the p_T dependent jet energy correction. With Z+jet events jet energy corrections for jets up to $250\,\mathrm{GeV/c}$ can be obtained. The statistical uncertainty is less than 1 % for jets with $p_T = 30\,\mathrm{GeV/c}$ and increases to about 3 % for jets with $p_T = 250\,\mathrm{GeV/c}$ using $100\,\mathrm{pb}^{-1}$ of data. Systematic uncertainties are related to the topological cuts and the flavour mapping. The uncertainty due to the topological cuts is estimated to be about 5 % for jets with $p_T = 30\,\mathrm{GeV/c}$ and 2 % for jets with $p_T = 250\,\mathrm{GeV/c}$ [128, 129]. Also in the case of Z+jet events the method is quickly dominated by systematic uncertainties. The differential method based on top quark events is not limited by the systematic uncertainty during early data taking and therefore complements the above mentioned methods. Furthermore, it is interesting for different physics studies to apply the η and p_T dependent jet energy corrections obtained by other methods and apply the inclusive $\mathrm{L}^{\mathrm{top}}_{\mathrm{incl}}$ jet energy corrections from top quark events to correct for the flavour assigned to the jets and perform a correction back to the parton level.

You must look within for value, but must look beyond for perspective.

Denis Waitley

Chapter 8

Conclusions and perspectives

In the field of experimental particle physics at colliders and in particular for precision physics studies and searches at the LHC it is essential to have well calibrated jet energies. Therefore, several methods to determine the correction factors on the energy of reconstructed jets are examined and new methods are gradually developed. The considered methods should be able to perform well during the early data taking at the LHC and result in the lowest possible statistical and systematical uncertainties. For the CMS experiment at the LHC techniques are implemented to estimate the jet energy correction factors from the first data. At the relatively high center of mass energies reached at the LHC, top quark pair events can be used for the first time in the history of collider physics as a calibration tool for both b quark and light quark jets. In this thesis a method is presented to derive jet energy correction factors using constraints on the W boson and top quark masses reconstructed in the semi-muonic $t\bar{t}$ decay.

After the study of the event simulation, reconstruction and selection in Chapters 3, 4 and 5, the method developed to measure the residual jet energy corrections from top quark pair events was extensively discussed in Chapters 6 and 7. An overview of the main results is provided in Section 8.1 and the minimal integrated luminosity needed to perform the measurement is discussed. Additionally, possible applications of the jet energy corrections obtained with top quark events are listed. The developed method to estimate the jet energy corrections can be extended and this is the topic of the discussion in Section 8.2. The technique can be used to estimate jet energy corrections for any jet reconstruction algorithm. Also a differentiation with respect to the number of primary vertices can be obtained to provide jet energy correction factors as a function of the amount of pile-up collisions in the event. Furthermore, a combined measurement of the top quark mass and jet energy correction factors can be envisaged. An extrapolation of the results obtained from simulated proton collisions at a center of mass energy of 10 TeV to the center of mass energy of the proton collisions at 7 TeV foreseen to be accumulated in 2010 and 2011 is presented.

8.1 Conclusions

For the CMS experiment a factorized jet calibration strategy is deployed as described in Section 4.1.5. Before performing the developed method based on top quark events, the

jet energies are corrected for the dependency on the pseudo-rapidity (L2 correction) and transverse momentum (L3 correction). An event-by-event kinematic fitting technique, incorporating the W boson and top quark masses as constraints in the $t \to bW \to bq\bar{q}$ decay, is used to estimate the residual jet energy corrections for both b quark jets and light quark jets. Both an inclusive and a differential estimation of the jet energy corrections is performed using jets reconstructed with the Seedless Infrared Safe Cone algorithm and a cone opening angle of $\Delta R = 0.5$. Due to the dependency of the corrections on the flavour composition of the sample used to derive the L2 and L3 corrections, the residual jet energy corrections are non-zero. This section provides a summary of the measurement of these residual jet energy corrections and the potential of the method.

8.1.1 Inclusive measurement of the jet energy corrections

In Chapter 6 an inclusive measurement of the jet energy corrections was performed. The estimated relative jet energy corrections for b quark jets is

$$\left(\frac{\Delta E}{E}\right)_b = \Delta E_{b,\text{est}} = -3.0 \pm 2.0\,(\text{stat}) \pm 1.3\,(\sigma_{m_{\text{top}}}) \pm 0.7\,(\text{syst})\,\%, \tag{8.1}$$

and for light quark jets is

$$\left(\frac{\Delta E}{E}\right)_l = \Delta E_{l,\text{est}} = -11.8 \pm 1.3\,(\text{stat}) \pm 0.5\,(\text{syst})\,\%. \tag{8.2}$$

They were found to be compatible with the expected values. While the magnitude of the jet energy corrections is well estimated, it is important to have a closer look at the uncertainty. The statistical uncertainty reflects the uncertainty to be obtained with $50\,\text{pb}^{-1}$ of data at a center of mass energy of 10 TeV rescaled from a sample with a higher integrated luminosity. The value of the statistical uncertainty was corrected for the non-unity width of the pull distribution obtained from pseudo-experiments reflecting data sets with an integrated luminosity of $50\,\text{pb}^{-1}$. These uncertainties for $50\,\text{pb}^{-1}$ of 10 TeV data are obtained from on average 120 selected signal events and 30 selected background events.

The main contribution to the systematic uncertainty on the estimated b quark jet energy correction is induced by the uncertainty on the top quark mass measurement at the Tevatron collider, which is currently $1.25\,\text{GeV/c}^2$. It is expected that this uncertainty will be further reduced to about $1\,\text{GeV/c}^2$ [135], which would result in a 1 % contribution to the systematic uncertainty of the estimated b quark jet energy correction. Other contributions to the systematic uncertainty were investigated and within the statistical uncertainties no important systematic effect was identified. As a conservative choice the largest statistical uncertainty on the estimated systematic effects was taken as an estimate of the systematic uncertainty.

In Figure 8.1 the statistical and systematical uncertainties on the estimation of the b and light quark jet energy corrections are shown as a function of the integrated luminosity. For each integrated luminosity, the pull distribution was obtained to correct the statistical uncertainty. The statistical uncertainty is taken from the resampling

exercise rather than a rescaling from a sample with a higher integrated luminosity. Within the uncertainty, no significant change of the width of the pull distributions is observed for different integrated luminosities. Using a dataset with an integrated

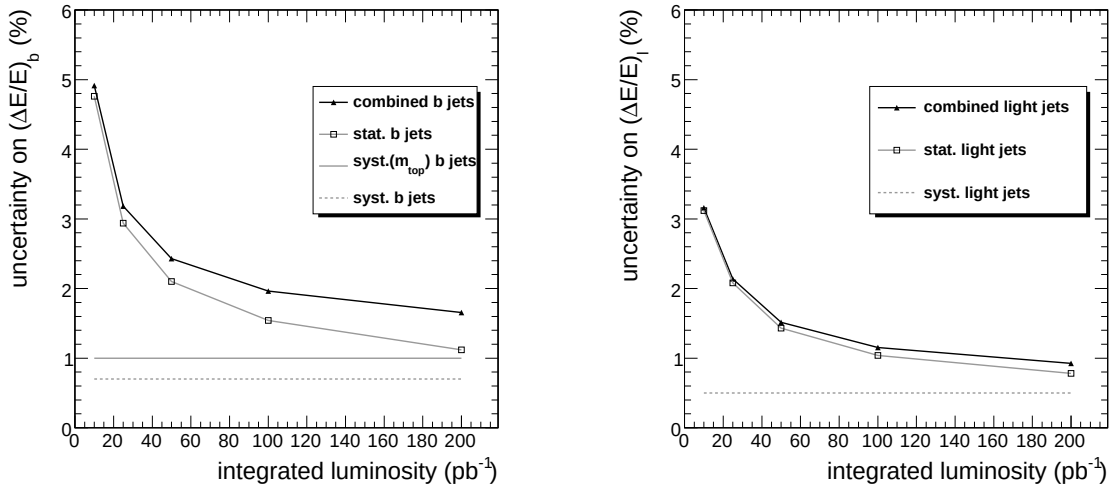


Figure 8.1: Expected uncertainty on the estimation of the jet energy correction factors for b (left) and light (right) quark jets as a function of the integrated luminosity for 10 TeV collisions. An improved precision of $1\,\mathrm{GeV/c^2}$ on the top quark mass measurement at the Tevatron collider is assumed.

luminosity of $10\,\mathrm{pb}^{-1}$ at 10 TeV the jet energy correction factors can be estimated with a statistical precision of about 5 % for b quark jets and 3 % for light quark jets. Taking into account an improved precision on the top quark mass, a systematic uncertainty of 1 % is expected on the b quark jet energy correction estimation. Other systematic effects are expected to be smaller than 0.5 % and 0.7 % for the light and b quark jet energy corrections respectively. Therefore for integrated luminosities below $200\,\mathrm{pb}^{-1}$ the statistical uncertainty dominates.

The method can be applied on jets that received an initial calibration for the jet energies, such that the corrected jet energies deviate less than 20 to 30 % from the true parton energies. Larger deviations can induce a bias on the estimation of the jet energy corrections or non-linearities on the estimator. A possible bias is however avoided by applying an initial calibration and performing successive iterations of the developed method. In each iteration, the residual jet energy corrections are estimated, whereafter these estimated corrections are used to rescale the four-momenta of the jets as described in Section 6.5.1. The next iteration will determine the residual corrections after the first estimation. Furthermore, in the same section, it was shown that by performing several iterations, possible non-linearities are resolved. When the method is applied on real collision data, it is therefore advised to perform successive iterations until the residual jet energy corrections vanish and remain invariant.

This indicates that the method can also serve as a closure test for jet energy corrections obtained with other methods, but additionally the residual jet energy corrections on top of these methods can be estimated.

8.1.2 Differential measurement of the jet energy corrections

A differentiated measurement of the jet energy corrections with respect to the number of muons in the b quark jet, or the transverse momenta or pseudo-rapidity of the b or light quark jets was discussed in Chapter 7.

It is important to know which integrated luminosity is minimally needed to measure the jet energy correction factors as a function of the transverse momentum or pseudo-rapidity. Therefore, it was checked for which integrated luminosity the statistical and systematical uncertainties are comparable. At 10 TeV a minimal integrated luminosity of $500\,\mathrm{pb}^{-1}$ is needed when four bins are used in either the p_T or η dimensions. However, it is important to note that the largest systematic uncertainty is related to the statistical uncertainty on the possible systematic effects. A real limitation of the method comes from the precision on the top quark mass measurement for the estimation of the b quark jet energy correction. With the current and future data recorded by the Tevatron experiments, this precision is expected to go down to $1\,\mathrm{GeV/c^2}$. In Figure 8.2 the different contributions to the total uncertainty taking into account the higher experimental precision on the top quark mass, is shown. Taking into account only the statistical uncertainty and the systematic uncertainty related to the precision on the top quark mass, a total uncertainty of about 1.8 % and 0.7 % on respectively the b and light quark jet energy correction measurements as a function of the pseudo-rapidity or transverse momentum of the jets can be obtained using $500\,\mathrm{pb}^{-1}$ of data.

The dependency on the pseudo-rapidity or transverse momentum of the jet can be estimated separately or successively as explained in Section 7.2. During this so-called factorized approach, first an inclusive estimation of the jet energy corrections is performed and the jet four-momenta are corrected with these factors. Then the estimation is performed as a function of the pseudo-rapidity, while the estimation as a function of the transverse momentum is only performed after correcting for the pseudo-rapidity dependence. Using this approach the different contributions to the jet energy corrections are factorized which allows for more transparency.

8.1.3 Applications of the estimated jet energy corrections

When the estimated residual jet energy corrections are applied on the jets in the jet combination chosen by the MVA method, the reconstructed top quark and W boson mass spectra are peaked at the correct value (cfr. Section 7.3). This validates the developed method for the estimation of the residual jet energy corrections for b and light quark jets. Moreover, since a dependency of the jet energy corrections on the transverse momentum, p_T, and pseudo-rapidity, η, of the jets can be estimated, the p_T and η dependent estimated jet energy corrections can be applied on various event topologies for which a flavour dependent correction to parton level is needed.

8.2 Perspectives

The method to measure the jet energy corrections from $\mathrm{t\bar{t}}$ events presented in this thesis can be extended for different purposes. In the current section, examples are

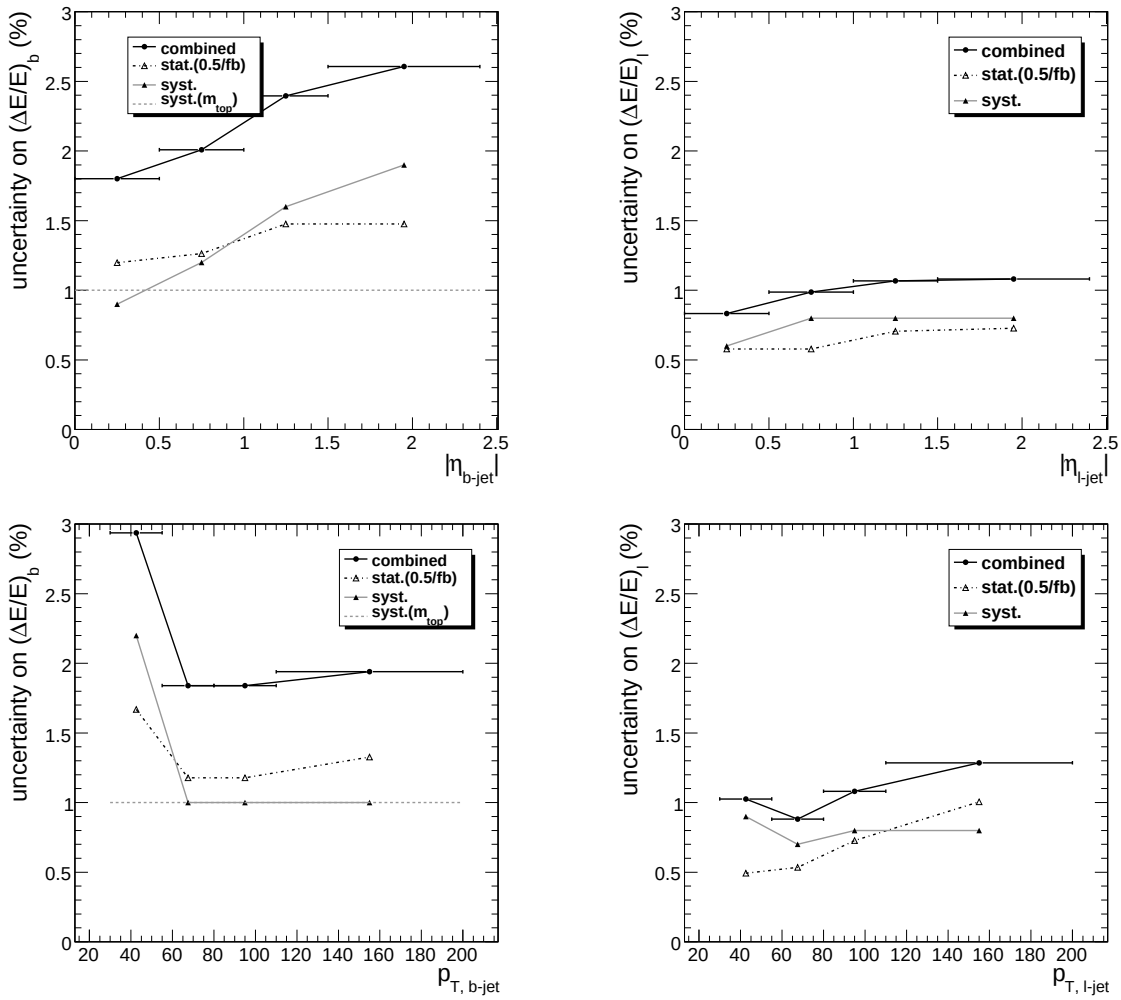


Figure 8.2: Expected uncertainty on the estimated jet energy corrections using $500\,\mathrm{pb}^{-1}$ of 10 TeV data as a function of the jet pseudo-rapidity (upper plots) and jet transverse momentum (lower plots) for b (left) and light (right) quark jets.

given of various possible extensions. Also the potential of the method for the foreseen data taking period at 7 TeV in 2010 and 2011 is discussed.

8.2.1 Other jet algorithms

The method to estimate the jet energy corrections from top quark pair events is developed on the jets reconstructed with the Seedless Infrared Safe Cone algorithm with a cone radius in $(\eta \times \phi)$ space of $\Delta R = 0.5$. It was shown for both the inclusive estimation of the jet energy corrections in Chapter 6 as well as for the differential estimation of the jet energy corrections in Chapter 7, that the method can also be applied on jets clustered with the inclusive longitudinally invariant k_T algorithm. More generally, the method can be used to estimate jet energy corrections for both cone-type jets as well as jets reconstructed by sequential clustering algorithms, such as the k_T algorithm. It is observed that the expected statistical precision on the estimated jet energy corrections is independent of the jet algorithm used.

Beside CaloJet reconstruction techniques, the method is expected to work also for JPTJets and PFJets introduced in Section 4.1.1. The JPTJets and PFJets will have a better energy resolution because also the information of the silicon tracker is taken into account. The jet energy corrections are therefore expected to be smaller. More importantly, the expected statistical precision on the estimated jet energy corrections will be slightly smaller with JPTJets and PFJets compared to the CaloJet reconstruction techniques because of the improved jet energy resolutions and b-tagging performance.

8.2.2 Differentiation with respect to the number of primary vertices

The jet energy corrections depend on the number of proton collisions in the same bunch crossing. Therefore, jet energy corrections are needed that can correct for the dependency on the amount of pile-up collisions in the event. The developed method for the estimation of the jet energy corrections can be performed as a function of the number of primary vertices. The estimated jet energy corrections can then be applied on any event topology according to the number of primary vertices observed in each event.

8.2.3 Combining with the estimation of the top quark mass

The method to estimate the jet energy corrections can be extended to measure the top quark mass together with the b and light quark jet energy corrections. To achieve this, the kinematic fit is performed on an event-by-event basis, using an arbitrary top quark mass and the world average value of the W boson mass as constraints. The idea is then to perform the kinematic fit for different top quark masses in a broad range such that a probability distribution $P^{\text{fit}}(\Delta E_b, \Delta E_l, m_{\text{top}})$ is constructed. For the true top quark mass, the value S of the χ^2 expression in Section 6.2.1 should be minimal. This technique is, however, expected to be very sensitive to the precise determination of the resolutions on the parameterized jet four-momenta. When the resolutions are

not well-known, this extended method will result in a bias on both the estimation of the top quark mass as well as on the estimation of the b quark jet energy correction.

8.2.4 Extrapolation towards other center of mass energies

The results presented in this thesis are obtained for simulated proton collisions at a center of mass energy of 10 TeV. Initially, the LHC is operating however at a center of mass energy of 7 TeV. For the 2010-2011 run it is planned to collect $1\,\text{fb}^{-1}$ of data at 7 TeV. Therefore, it is useful to know what the results would be at this lower center of mass energy. In Section 3.2 it was discussed that the cross section for top quark pair production is reduced by a factor of 2.4 when the protons collide at a center of mass energy of 7 TeV instead of 10 TeV. However, the cross sections of the main background processes are reduced by smaller factors. The relatively lower signal to background ratio should not change the potential of the method, because in Section 6.6.1 the method was found to be robust with respect to increased cross sections of background processes. The potential of the method at 7 TeV is obtained by rescaling the integrated luminosity in Figure 8.1 with a factor of $\sqrt{2.4}$. Using a dataset with an integrated luminosity of about $15\,\text{pb}^{-1}$ at 7 TeV, an inclusive estimation of the jet energy corrections can be performed with a precision of about 5 % and 3 % for the b and light quark jet energy corrections respectively.

Similarly, when the LHC operates at the design center of mass energy of 14 TeV, the $\text{t}\bar{\text{t}}$ production cross section increases with a factor of about 2.2 compared to the cross section at 10 TeV. Therefore, it is expected that a precision of about 5 % and 3 % for the estimation of respectively the b and light quark jet energy corrections can be obtained using a data set with an integrated luminosity of about $7\,\text{pb}^{-1}$.

At different center of mass energies, the criteria to select the events and the MVA techniques to choose the correct jet combination have to be revisited. For instance, the $\frac{p_T^{\text{t,h}}}{\sum p_T^{\text{3jets}}}$ variable used in the MVA discriminator is based on the transverse momentum of the top quark with respect to the background. For smaller center of mass energies, this variable is expected to have less discriminating power.

Bibliography

[1] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory*, USA: Addison-Wesley (1995) 842 p.

[2] F. Halzen and A. D. Martin, *Quarks and Leptons: an Introductory Course in Modern Particle Physics*, New York, Usa: Wiley (1984) 396p.

[3] *Particle Data Group*, `http://pdg.lbl.gov/`.

[4] C. Quigg, *Gauge Theories of the Strong, Weak and Electromagnetic Interactions*, USA: Addison-Wesley (1997) 334 p.

[5] F. Englert and R. Brout, Phys. Rev. Lett. **13** (1964) 321–322.

[6] P. Higgs, Phys. Rev. Lett. **13** (1964) 508–509.

[7] G. Guralnik, C. Hagen, and T. Kibble, Phys. Rev. Lett. **13** (1964) 585–587.

[8] The CDF Collaboration, Phys. Rev. Lett. **74** (Apr, 1995) 2626–2631.

[9] The D0 Collaboration, Phys. Rev. Lett. **74** (Apr, 1995) 2632–2637.

[10] *The CDF experiment*, `http://www-cdf.fnal.gov/physics/physics.html`.

[11] *The D0 experiment*, `http://www-d0.fnal.gov/`.

[12] *The Tevatron Collider*, `http://www-bdnew.fnal.gov/tevatron/`.

[13] A. Quadt, Eur. Phys. J. **C48** (2006) 835–1000.

[14] The Tevatron Electroweak Working Group, *Combination of CDF and D0 Measurements of the Single Top Production Cross Section*, `arXiv:hep-ph/0908.2171`.

[15] The Tevatron Electroweak Working Group, *Combination of CDF and D0 Results on the Mass of the Top Quark*, `arXiv:hep-ex/0903.2503`. FERMILAB-TM-2427-E.

[16] F. Fiedler, *Precision Measurements of the Top Quark Mass*, `arXiv:hep-ex/1003.0521`.

[17] The LEP Electroweak Working Group `http://lepewwg.web.cern.ch/LEPEWWG/`.

[18] R. Frederix and F. Maltoni, JHEP **01** (2009) 047.

[19] H.-J. He, N. Polonsky, and S. Su, Phys. Rev. **D64** (2001).

[20] C. Balázs, M. Carena, and C. E. M. Wagner, Phys. Rev. D **70** (Jul, 2004) 015007.

[21] D. J. H. Chung *et al.*, Phys. Rept. **407** (2005) 1–203.

[22] H. Bachacou and P. Van Mulders, Nuovo Cim. **123B** (2008) 1197–1204.

[23] P. Van Mulders, Nuovo Cim. **123B** (2008) 1325–1326.

[24] The CMS Collaboration, *Measurement of jet energy scale corrections using top quark events*, CMS PAS TOP-07/004.

[25] S. Lowette, J. D'Hondt, J. Heyninck, and P. Vanlaer, *Offline Calibration of b-Jet Identification Efficiencies*, CMS-NOTE-2006-013.

[26] V. E. Bazterra and T. Speer, *Strategies for btagging calibration using collider data at CMS*, CMS-CR-2009-032.

[27] B. S. Acharya, F. Cavallari, G. Corcella, R. Di Sipio, and G. Petrucciani, *Commissioning ATLAS and CMS with top quarks*, CMS-CR-2008-014.

[28] L. Evans and P. Bryant, JINST **3 S08001** (2008).

[29] J. P. Blewett, *200 GeV intersecting storage accelerators*, Proceedings of the 8^{th} International Conference on High-Energy Accelerators.

[30] The ATLAS Collaboration, JINST **3 S08003** (2008).

[31] The CMS Collaboration, JINST **3 S08004** (2008).

[32] The LHCb Collaboration, JINST **3 S08005** (2008).

[33] The ALICE Collaboration, JINST **3 S08002** (2008).

[34] The TOTEM Collaboration, JINST **3 S08007** (2008).

[35] The LHCf Collaboration, JINST **3 S08006** (2008).

[36] M. Della Negra, L. Foà, A. Hervé, and A. Petrilli, *CMS physics Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 2006.

[37] The CMS Collaboration, *Track reconstruction in the CMS Tracker*, Note in preparation.

[38] R. Frühwirth, Nuclear Instruments and Methods in Physics Research A **262** (1987) 444.

[39] P. Azzurri, *Track reconstruction Performance in CMS*, CMS CR-2008/110.

[40] The CMS HCAL Collaboration, Eur. Phys. J. **C55** (2008) 159.

[41] J. Knobloch *et al.*, *LHC Computing Grid Technical Design Report*, CERN LHCC-2005-024.

[42] A. Fanfani *et al.*, J. Grid Computing (2010) 1572–9814.

[43] G. Codispoti *et al.*, IEEE Trans. Nucl. Sci. **56** (2009) 2850–2858.

[44] G. Bagliesi *et al.*, NSS IEEE (2008).

[45] G. Bagliesi *et al.*, *Debugging Data Transfers in CMS*, CMS CR-2009/093.

[46] D. Bonacorsi, Production Grids in Asia (2010) 59–69.

[47] M. Dobbs, S. Frixione, E. Laenen, and K. Tollefson, *Les Houches Guidebook to Monte Carlo Generators for Hadron Collider Physics*, `arXiv:hep-ph/0403045`.

[48] T. Sjöstrand, S. Mrenna, and P. Skands, JHEP **05** (2006) 026.

[49] G. Corcella *et al.*, JHEP **01** (2001) 010.

[50] M. Mangano *et al.*, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, `arXiv:hep-ph/0206293`.

[51] T. Stelzer and W. F. Long, Comput. Phys. Commun. **81** (1994) 357.

[52] F. Maltoni and T. Stelzer, JHEP **0302** (2003) 027.

[53] E. Boos *et al.*, *Generic User Process Interface for Event Generators*, `arXiv:hep-ph/0109068`.

[54] J. C. Collins, D. E. Soper, and G. Sterman, *Factorization of Hard Processes in QCD*, `arXiv:hep-ph/0409313`.

[55] D. Stump *et al.*, JHEP **0310** (2003) 046.

[56] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, Phys. Lett. **B 604** (2004) 61.

[57] *Online Plotting and Calculation of Parton Distributions*, `http://durpdg.dur.ac.uk/hepdata/pdf3.html`.

[58] D. Stump *et al.*, Phys. Rev. **D65** (2002) 014012.

[59] J. Pumplin *et al.*, Phys. Rev. **D65** (2002) 014013.

[60] J. M. Campbell, J. W. Huston, and W. J. Stirling, Rept. Prog. Phys. **70** (2007) 89.

[61] Y. L. Dokshitzer, Sov. J. Phys. JETP **46** (1977) 641.

[62] V. N. Gribov and L. N. Lipatov, Sov. J. Nucl. Phys. **15** (1972) 438.

[63] G. Altarelli and P. G, Nucl. Phys. **B126** (1977) 298.

[64] V. V. Sudakov, Zh. E. T. F. **30** (1956) 87.

[65] P. Bartalini, R. Chierici, and A. De Roeck, *Guidelines for the estimation of theoretical uncertainties at the LHC*, CMS NOTE-2005/013.

[66] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber, JHEP **0111** (2001) 063.

[67] F. Krauss, JHEP **0208** (2002) 015.

[68] S. Hoche, F. Krauss, N. Lavesson, L. Lönnblad, M. Mangano, A. Schalicke, and S. Schumann, *Matching Parton Showers and Matrix Elements*, `arXiv:hep-ph/0602031`.

[69] M. Mangano, M. Moretti, F. Piccinini, and M. Treccani, *Matching matrix elements and shower evolution for top-quark production in hadronic collisions*, `arXiv:hep-ph/0611129`.

[70] S. Frixione and B. R. Webber, JHEP **0206** (2002) 029.

[71] B. Andersson, Acta Phys. Polon. **B32** (2001) 3993–4011.

[72] J. Schwinger, Phys. Rev. **82** (1995) 664.

[73] C. Peterson, D. Schlatter, I. Schmitt, and P. M. Zerwas, Phys. Rev. **D27** (1983) 105–111.

[74] The OPAL Collaboration, Z. Phys. **C69** (1996) 543.

[75] The ALEPH Collaboration, Z. Phys. **C55** (1992) 209.

[76] The DELPHI Collaboration, Z. Phys. **C73** (1996) 11.

[77] A. Moreas *et al.*, Acta Phys. Polon. **B35** (2004) 433.

[78] P. Bartalini *et al.*, *Bottom Production*, `arXiv:hep-ph/0003142`.

[79] *Peter's Pythia Plots*, `http://home.fnal.gov/~skands/leshouches-plots/`.

[80] T. Sjöstrand and P. Skands, JHEP **03** (2004) 053.

[81] The CMS Collaboration, JHEP **02** (2010) 041.

[82] N. Kidonakis and R. Vogt, *The theoretical top quark cross section at the Tevatron and the LHC*, `arXiv:hep-ph/08053844`.

[83] M. Cacciari, S. Frixione, M. Mangano, P. Nason, and G. Ridolfi, *Updated predictions for the total production cross sections of top and of heavier quark pairs at the Tevatron and at the LHC*, `arXiv:hep-ph/08042800`.

[84] N. Kidonakis, *Higher-order corrections to top-antitop pair and single top quark production*, `arXiv:hep-ph/09090037`.

[85] The GEANT4 Collaboration, Nucl. Instrum. Meth. **A506** (2003) 250–303.

[86] A. Giammanco and A. Perrotta, *Fast Simulations of the ATLAS and CMS experiments at LHC*, CMS CR-2007/010.

[87] D. Evans *et al.*, *The CMS Monte Carlo Production System: Development and Design*, Presented at the Hadron Collider Physics Symposium HCP 2007, Italy.

[88] J. M. Hernández *et al.*, Journal of Physics: Conference Series **119** (2008) 052019.

[89] A. Mohapatra *et al.*, Nucl. Phys. **B177-178** (2008) 324–325.

[90] *Reference Page for Top-Quark Analyses: Samples and Cross-Sections*, `http://www.ge.infn.it/~tosi/cms/topMC.html`.

[91] B. Harris, E. Laenen, L. Phaf, and Z. Sullivan, *The fully differential single-top-quark cross section in next-to-leading order QCD*, `arXiv:hep-ph/0207055`.

[92] Z. Sullivan, Phys. Rev **D70** (2004) 114012.

[93] J. Campbell, R. Ellis, and F. Tramontano, Phys. Rev **D70** (2004) 094012.

[94] J. Campbell and F. Tramontano, Nucl. Phys. **B726** (2005) 109.

[95] *MadGraph Samples for LHC physics*, `http://cp3wks05.fynu.ucl.ac.be/twiki/bin/view/Library/MadGraphSamples`.

[96] S. Frixione and M. Mangano, *How accurately can we measure the W cross section?*, `arXiv:hep-ph/0405130`.

[97] *CMSSW Application Framework*, `https://twiki.cern.ch/twiki/bin/view/CMS/WorkBookCMSSWFramework`.

[98] M. Hansen *et al.*, *The Top Quark Analysis Framework*, CMS IN-2007/068.

[99] R. Wolf, *A tour of the CMS Physics Analysis Model*, CMS CR-2009/350.

[100] W. Adam *et al.*, *PAT: the CMS Physics Analysis Toolkit*, CMS CR-2009/083.

[101] *TQAF classes*, `https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideTQAFClasses`.

[102] P. Vanlaer, *Electron and photon reconstruction in CMS*, CMS CR-2009/201.

[103] S. Baffioni *et al.*, *Electron reconstruction in CMS*, CMS NOTE-2006/40.

[104] The CMS Collaboration, *CMS Strategies for tau reconstruction and identification using particle-flow techniques*, CMS PAS PFT-08/001.

[105] C. Buttar *et al.*, *Standard Model Handles and Candles Working Group: Tools and Jets Summary report*, arXiv:hep-ph/0803.0678v1.

[106] *The SISCone Jet Algorithm*, http://projects.hepforge.org/siscone.

[107] *FastJet*, http://www.lpthe.jussieu.fr/~salam/fastjet.

[108] G. P. Salam and G. Soyez, *A practical Seedless Infrared-Safe Cone jet algorithm*, arXiv:hep-ph/0704.0292.

[109] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, Nucl. Phys. **B406** (1993) 187.

[110] R. Demina *et al.*, *Calorimeter Energy Thresholds for Jet Reconstruction in CMS*, CMS NOTE-2006/020.

[111] V. Chetluru, F. Pandolfi, P. Schieferdecker, and M. Zielinski, *Jet Reconstruction Performance at CMS*, CMS AN-2009/067.

[112] The CMS Collaboration, *Jet Plus Tracks Algorithm for Calorimeter Jet Energy Corrections in CMS*, CMS PAS JME-09/002.

[113] The CMS Collaboration, *Particle-Flow Event Reconstruction in CMS and Performance for Jets, E_T^{miss}, and Taus*, CMS PAS PFT-09/001.

[114] Particle Data Group, Journal of Phys. **G33** (1993) 861–867.

[115] The CMS Collaboration, *Algorithms for b Jet Identification in CMS*, CMS PAS BTV-09/001.

[116] W. Waltenberger, *Development of Vertex Finding and Vertex Fitting Algorithms for CMS*, CMS TS-2006/012.

[117] W. Waltenberger, R. Frühwirth, and P. Vanlaer, Journal of Phys. **G34** (2007) N343–356.

[118] W. Waltenberger, R. Frühwirth, and P. Vanlaer, *Adaptive Vertex Fitting*, CMS NOTE-2007/008.

[119] T. Speer, K. Prokofiev, R. Frühwirth, W. Waltenberger, and P. Vanlaer, *Vertex Fitting in the CMS Tracker*, CMS NOTE-2006/032.

[120] Private communication with C. Saout, February 2010.

[121] *TMultiLayerPerceptron*, http://root.cern.ch/root/html/TMultiLayerPerceptron.html.

[122] The CMS Collaboration, *Plans for Jet Energy Corrections at CMS*, CMS PAS JME-07/002.

[123] A. Bhatti *et al.*, *Plans for Jet Energy Corrections at CMS*, CMS AN-2007/055.

[124] The CMS Collaboration, *Offset Energy Correction for Cone Jets*, CMS PAS JME-09/003.

[125] V. Chetluru *et al.*, *Update of calorimeter cell and tower thresholds for jet reconstruction in first CMS collision data*, CMS AN-2010/024.

[126] The CMS Collaboration, *Determination of the Relative Jet Energy Scale at CMS from Dijet Balance*, CMS PAS JME-08/003.

[127] The CMS Collaboration, *Jet energy calibration with photon+jet events*, CMS PAS JME-09/004.

[128] The CMS Collaboration, *Calibration of the absolute jet energy scale with $Z(\rightarrow \mu^+\mu^-)$+jet events at CMS*, CMS PAS JME-09/009.

[129] V. Buge *et al.*, *Jet Energy Correction Using $Z(\rightarrow e^+e^-)$+Jet p_T Balance and the Method for Combining Data Driven Corrections*, CMS AN-2009/004.

[130] The CMS Collaboration, *Jet Corrections to Parent Parton Energy*, CMS PAS JME-08/002.

[131] N. Amapane *et al.*, *Muon Identification in CMS*, CMS AN-2008/098.

[132] The CMS Collaboration, *Prospects for the first Measurement of the $t\bar{t}$ Cross Section in the Muon-plus-Jets Channel at $\sqrt{s}$ = 10 TeV with the CMS Detector*, CMS PAS TOP-09/003.

[133] *TMVA Toolkit for Multivariate Data Analysis with ROOT*, `http://tmva.sourceforge.net`.

[134] J. D'Hondt, S. Lowette, *et al.*, *Fitting of Event Topologies with External Kinematic Constraints in CMS*, CMS Note-2006/023.

[135] CDF and D0 Collaboration, L. Brigliadori, Nuovo Cim. **032C** (2009) 265–272.

Summary

The high center of mass energy reached in the proton collisions at the Large Hadron Collider is opening a new era in particle physics for the study and discovery of new phenomena at the TeV scale. Many physics analyses at hadron colliders are based on the observation of jets for which the precise reconstruction of the four-momenta is a key aspect for the success of the experiment. The study of different jet reconstruction algorithms indicates that a detailed calibration procedure for the obtained jet energies is crucial for the physics program of the LHC. For this purpose a factorized approach is deployed at the Compact Muon Solenoid experiment.

Although the top quark sector is expected to give an important window on new phenomena it is also used to commission and calibrate the detector and reconstruction tools. The method developed in this thesis uses the $pp \to t\bar{t} \to bW\bar{b}W \to bq\bar{q}\bar{b}\mu\bar{\nu}_\mu$ process. Dedicated event selection criteria are deployed, in particular criteria based on the isolation of the muon in the final state. To correctly reconstruct the W boson and top quark mass spectra in the $t \to bW \to bq\bar{q}$ decay, the observed jets need to be assigned to the correct quarks from which they are originating. Diverse multi-variate analysis tools are considered to combine the variables which are able to discriminate between wrong and correct jet-quark assignments. With these tools a dataset of 690 signal events and 560 background events is expected for an integrated luminosity of $50\,\mathrm{pb}^{-1}$ at 10 TeV. Of the signal events about 17 % has a correct assignment of observed jets to the quarks in the $t \to bW \to bq\bar{q}$ decay.

Using an event-by-event kinematic fitting technique with Lagrange multipliers to incorporate the W boson and top quark mass constraints in the fitted $t \to bW \to bq\bar{q}$ decay, a novel method is deployed for the estimation of the jet energy corrections for both b and light quark jets. With an integrated luminosity of $10\,\mathrm{pb}^{-1}$ at 10 TeV the jet energies can be calibrated with a statistical precision of about 5 % for b quark jets and 3 % for light quark jets relative to the energy of the jets. The most important systematic uncertainty on the b quark jet energy correction factor is related to the precision of the top quark mass measurements obtained at the Tevatron collider. Taking into account an improved precision of $1\,\mathrm{GeV/c^2}$ on the top quark mass measurement, the related systematic uncertainty on the b jet energy scale is about 1 %. Other systematic effects are expected to be smaller than 0.7 % and 0.5 % relative to the energy of the jets, for the b and light quark jet energies respectively.

With the same method, also a differential estimation of the jet energy corrections is performed. The differentiation is performed with respect to the number of muons in the b quark jet and with respect to the jet kinematics, in particular the pseudo-rapidity and transverse momentum. Using a dataset with an integrated luminosity of $500\,\mathrm{pb}^{-1}$ at

10 TeV and taking into account the statistical uncertainty together with the systematic uncertainty related to the precision on the top quark mass, a total uncertainty relative to the jet energy of about 1.8 % and 0.7 % on respectively the b and light quark jet energies as a function of the pseudo-rapidity or transverse momentum of the jets can be obtained. The estimated b and light quark jet energy corrections can be applied on various event topologies for which a flavour dependent correction to parton level is required.

The method to calibrate reconstructed jets developed in this thesis is potentially the best performant in the kinematic range $30 \leq p_T \leq 200\,\mathrm{GeV/c}$. Only 15 years after its discovery and even without a profound understanding of the top quark sector, the top quark is becoming one of the key handles for calibration at the Large Hadron Collider.

Samenvatting

Calibratie van de jet energie met behulp van top quark gebeurtenissen aan de LHC

De hoge massamiddelpuntsenergie die bereikt wordt in de proton botsingen aan de Large Hadron Collider (LHC) laat toe om nieuwe fenomenen nabij de TeV energieschaal te ontdekken en te bestuderen. De recente start van dit groots experiment bij massamiddelpuntsenergieën van 7 TeV markeert een nieuw tijdperk in de elementaire deeltjesfysica. Verschillende studies nabij hadron versnellers zijn gebaseerd op de observatie van jets voortkomend uit de fragmentatie van quarks of gluonen. Een precieze reconstructie van de vier-impuls van deze objecten is cruciaal voor het uitvoeren van het onderzoeksprogramma van het experiment. De studie van verschillende algoritmen voor de reconstructie van jets toont aan dat een gedetailleerde procedure voor het calibreren van de energieën van de gereconstrueerde jets noodzakelijk is. Voor deze calibratie werd in de Compact Muon Solenoid (CMS) collaboratie geopteerd voor een gefactoriseerde aanpak.

Hoewel de top quark sector belangrijk is om nieuwe fysica fenomenen te ontdekken, kan deze ook gebruikt worden voor het begrijpen en het calibreren van de CMS detector en de reconstructie algoritmen. De methode die werd uitgewerkt in deze thesis is opgesteld voor het $pp \to t\bar{t} \to bW\bar{b}W \to bq\bar{q}\bar{b}\mu\bar{\nu}_\mu$ proces en is getest op gesimuleerde proton botsingen bij een massamiddelpuntsenergie van 10 TeV. Specifieke criteria voor het selecteren van deze gebeurtenissen werden ontworpen, in het bijzonder criteria gebaseerd op de isolatie van het muon. Om de massaspectra van het W boson en de top quark in het $t \to bW \to bq\bar{q}$ verval correct te reconstrueren, moeten de waargenomen jets toegekend worden aan de quarks waar ze van komen. Verschillende multivariate technieken werden beschouwd om variabelen te combineren waarmee een onderscheid gemaakt kan worden tussen correcte en verkeerde jet-quark combinaties. Na het toepassen van de selectie criteria worden er 690 signaal en 560 achtergrond gebeurtenissen verwacht voor een geïntegreerde luminositieit van $50\,\text{pb}^{-1}$ bij een massamiddelpuntsenergie van 10 TeV. Voor 17 % van de signaal gebeurtenissen worden de quarks in het $t \to bW \to bq\bar{q}$ verval juist toegekend aan de waargenomen jets.

Een nieuwe en polyvalente calibratietechniek werd ontworpen voor de schatting van de correctiefactoren op de jet energieën voor zowel b quark jets als de jets geassocieerd met de quarks van het verval van het W boson. Hiervoor wordt gebruik gemaakt van een kleinste kwadraten methode waarbij de hypothesen van de W boson en de top quark massa getoetst worden voor elke gebeurtenis afzonderlijk. Met een geïntegreerde luminositeit van $10\,\text{pb}^{-1}$ bij 10 TeV kunnen de jet energieën gecalibreerd worden met

een relatieve statistische onzekerheid van ongeveer 5 % voor b quark jets en 3 % voor jets geassocieerd met de quarks van het verval van het W boson. De belangrijkste systematische onzekerheid op de correctiefactor van de b quark jet energie is gerelateerd met de nauwkeurigheid van de top quark massa metingen bekomen aan de Tevatron experimenten. Indien rekening gehouden wordt met een verbeterde onzekerheid op de top quark massa van $1\,\mathrm{GeV/c^2}$ is de overeenkomende systematische onzekerheid op de energieschaal van de b quark jets ongeveer 1 %. Andere systematische effecten op de energieschaal zijn kleiner dan 0.7 % en 0.5 % voor respectievelijk de energieën van de b quark jets en de jets geassocieerd met de quarks van het verval van het W boson.

Met dezelfde methode kan ook een gedifferentieerde schatting van de correctiefactoren op de jet energieën bekomen worden. De differentiatie werd uitgevoerd als functie van het aantal muonen in de b quark jet en als functie van de pseudo-rapiditeit en de transverse impuls van de jets. Met behulp van een hoeveelheid botsingsgegevens overeenkomend met een geïntegreerde luminositeit van $500\,\mathrm{pb^{-1}}$ bij een massamiddelpuntsenergie van 10 TeV en rekening houdend met zowel de statistische onzekerheid als de systematische onzekerheid gerelateerd aan de nauwkeurigheid op de top quark massa meting wordt een totale onzekerheid van ongeveer 1.8 % en 0.7 % bekomen op de energieschaal van respectievelijk de b quark jets en de jets geassocieerd met de quarks van het verval van het W boson in vier klassen van de pseudo-rapiditeit of de transverse impuls. Deze geschatte correctiefactoren kunnen toegepast worden op de topologieën van verschillende botsingsprocessen waarvoor een correctie van de jet energieën naar de quark energieën vereist is.

De ontworpen methode om de gereconstrueerde jets te calibreren heeft het potentieel om optimaal te presteren voor jets met een transverse impuls tussen 30 en $200\,\mathrm{GeV/c}$. Slechts 15 jaar na de ontdekking van de top quark en zelfs zonder de top quark sector te doorgronden, is de top quark één van de belangrijkste ingrediënten geworden voor calibratietechnieken nabij de LHC.

Acknowledgements

There are many different people that I would like to acknowledge for the sometimes very different ways in which they contributed to the realization of this book. First of all, I would like to thank the IIHE and the IWT for offering me the opportunity to start a PhD. The inspiration and enthousiasm but also the support and patience of my promotors Jorgen D'Hondt and Nick Van Remortel were of invaluable aid during this project. I appreciated the constructive comments and questions by the jury members and would like to express my gratitude to them for the evaluation of the manuscript.

There are also many people at the IIHE that helped me in various ways, for instance Rosine, Marleen, Daisy and Annie for the help on the "practical aspects" of doing a PhD. From the side of computing and the T2-related aspects at the IIHE, I would like to thank Danny, Stéphane, Olivier, Shkelzen, Abdel and Stijn. While doing service work for CMS, I had the opportunity to work together with many people which made the work pleasant, such as for example Peter Kreuzer, Daniele Bonacorsi, Christoph Wissing, Nicoló Magini, James Letts, Jukka Klem, Wolf Behrenhoff and Lukas Vanelderen. Thank you for your help and patience with the many computing issues. I'm also grateful to the people answering the many e-mails concerning CRAB problems. Then there are the people which I met only a few times (or even never) and that helped me with all the various aspects one encounters when developing an analysis in the CMS collaboration: Roberto Chierici, Roger Wolf, Sebastian Naumann-Emme, Christophe Saout and many other individuals in the Top, JEC and JetMET groups.

The social aspect of doing a PhD should not be underestimated. I met great people in the Top Brussels group: during the first months of the PhD, Steven and Jan were there to guide me. Jan, your help during the first year was important to me. I suspect you still have blue knees from all the times you hit the desks in our office and probably you still have a pencil which you forgot to give back. During the last four years I practically lived together with three amazing people. Joris, we shared 8 years or about a third of our lives, thanks for being there all that time! Grégory and Ilaria, we will probably never know if Paris or Roma is the center of the world, but I enjoyed the many lively discussions. I appreciate enormously the support and friendship from the three of you during the turbulent periods of the PhD, thanks! Catherine, thank you for the positive feedback and interest you were always showing in my work. Stéphanie, thanks for providing us with the "hot" news and all the updates of what was going on with the LHC. Volker, I always enjoyed the chats and constructive comments. Eric, having you at Brussels significantly improved the work and social life. Stijn, Michael and Alexis, your enthousiasm and willingness to listen to the "stories" of a senior PhD student helped me more than you probably realize! I also enjoyed working together or

chatting with Maryam and Nadjieh.

During the four years of studying physics and the subsequent years while doing a PhD I met people which enriched my life. Besma, Deniz, Julie, Otman, Mateusz, Vincent, Eileen, Bram and Evi, Laurène, and many others, thank you for the nice moments we spent together! My dearest friends, Wim, Marie, Sylvia and Antony, I will try to make up for the lack of time. Sylvia, I owe you so much, our "lunch dates" every week were the motivational boosts which were absolutely crucial to keep me going! I hope I can do the same for you during the next months and keep breathing, I know you can do it. I would also like to thank my family for their love, patience and stubborn belief in me. In particular Annarella because she showed me that physics was my real cup of tea, thanks for setting me back on track! Irina, thank you for the ultra-enthousiastic ("MIER!") e-mails you sent to me during the redaction of my thesis and for continuing doing so even when I did not answer.

I am eagerly looking forward to the new opportunities and the many experiences that lie ahead of me, but I will never forget this nice period which is coming to an end.

List of publications

1. Fanfani, A. *et al.*, *Distributed analysis in CMS*, J.Grid Comput. **8**, (2010) 159-179.

2. Bagliesi, G. *et al.*, *Debugging Data Transfers in CMS*, CMS Conference Report CR-2009/093, (2009).

3. Andreeva, J. *et al.*, *CMS Analysis Operations*, CMS Conference Report CR-2009/088, (2009).

4. Hernandez, J. M. *et al.*, *CMS Monte Carlo production in the WLCG computing grid*, J.Phys.Conf.Ser.119:052019, (2008).

5. Bachacou, H. and Van Mulders, P., *The top quark as a calibration tool at the LHC*, Nuovo Cim. **123B**, (2008) 1197-1204.

6. Van Mulders, P., *Estimation of the jet energy scale corrections using top quark events*, Nuovo Cim. **123B**, (2008) 1325-1326.

7. Mohapatra, A. *et al.*, *CMS Monte Carlo production operations in a distributed computing environment*, Nucl.Phys.Proc.Suppl.177-178, (2008).

8. Buttar, C. *et al.*, *Standard Model Handles and Candles Working Group: Tools and Jets Summary Report*, Proceedings of Les Houches 2007, Physics at TeV colliders, (2008) 121-214.

9. D'Hondt, J. and Van Mulders, P., *Measurement of jet energy scale corrections using top quark events*, CMS Physics Analysis Summary PAS-TOP-07-004 and Analysis Note AN-2007/029, (2008).

10. Esen, S. *et al.*, *Plans for Jet Energy Corrections at CMS*, CMS Physics Analysis Summary PAS-JME-07-002 and Analysis Note AN-2007/055, (2008).

11. Hansen, M. *et al.*, *The Top Quark Analysis Framework* , CMS Internal Note IN-2007/068, (2007).

12. D'Hondt, J. *et al.*, *Observability of same-charge lepton topology in di-leptonic t anti-t events*, CERN-CMS-NOTE-2006-065, (2006).

Co-author of 29 papers published by the CMS experiment.

No amount of experimentation can ever prove me right; a single experiment can prove me wrong.

Albert Einstein