Vrije Universiteit Brussel

Faculteit Wetenschappen en
Bio-Ingenieurswetenschappen
Departement Natuurkunde

Isfahan University of
Technology

Department of Physics

Measurement of the Jet Energy Scale in the CMS experiment with the First LHC Proton Collisions

Maryam Zeinali

Promotors
Prof. Dr. Jorgen D'Hondt
Prof. Dr. Mansour Haghighat

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

September 2011

Doctoral examination commission:

Prof. Dr. Freya Blekman (Vrije Universiteit Brussel, CERN)
Prof. Dr. Ben Craps (Vrije Universtiteit Brussel, *co-chair*)
Prof. Dr. Daniel Denegri (Saclay, CERN)
Prof. Dr. Jorgen D'Hondt (Vrije Universiteit Brussel, CERN)
Prof. Dr. Mansour Haghighat (Isfahan University of Technology)
Prof. Dr. Alexandre Nikitenko (Imperial College London, CERN)
Prof. Dr. Ahmad Shirzad (Isfahan University of Technology, *co-chair*)

Contents

Introduction

The Standard Model of particle physics describes the elementary particles and their interactions, with the exception of gravity which is not formulated within the model. The model has been tested in diverse experiments and its parameters have been measured with great precision in high energy physics laboratories. However, there is still a missing particle which has not yet been detected, being the Standard Model Higgs boson. Also there is evidence which makes the Standard Model only an effective theory and motivates the need to have new physics beyond it. Therefore, efforts are made to build the most fundamental theory describing nature. To search for the Higgs particle and test the new physics ideas beyond the Standard Model, the Large Hadron Collider (LHC) was constructed.

The LHC, located at the CERN laboratory near Geneva, has been colliding beams of protons at a center of mass energy of 7 TeV since March 2010. The proton-proton collisions are detected with the use of detectors installed at the interaction points, such as the CMS (Compact Muon Solenoid) experiment, which is one of the two general-purpose detectors at the LHC. The energy and momentum of particles which are produced in the final state of proton-proton interactions, are measured by the detectors. While the energy of leptons can be measured with high precision, the measurement of the energy of the quarks, which hadronize to make so-called jets, is quite challenging. While strategies are developed to calibrate the jet energies within the CMS collaboration, searching for new physics benefits from a more precise measurement of the jet energies than the ones obtained by applying the baseline methods. In this analysis, a method to calibrate the energy of the jets is applied on the proton collisions containing two top quarks in the final state, $pp \rightarrow t\bar{t}$. The high rate of the production of the top quark pairs in proton-proton collisions provides a huge amount of statistics. This allows the top quark pairs to be used as a tool to perform the calibration studies.

This thesis is organized as follows. In Chapter 1, the Standard Model of particle physics is reviewed with emphasis on the top quark sector followed by a motivation to construct the LHC. The description of the Large Hadron Collider and the CMS experiment, is given in Chapter 2. The CMS experiment detects the emerging particles from the proton collisions through electronic signals which are subsequently used to reconstruct the physics objects as described in Chapter 3. The simulation of the proton collisions is described in Chapter 4 followed by the description of the procedure to select top quark events on which the calibration method is based. The method to estimate the residual jet energy corrections is described in detail in Chapter 5. The estimated results of the method are obtained based on both the simulated proton collisions as well as the real proton collisions collected in 2010 by the CMS experiment. Finally, a

summary of the analysis and a conclusion are given in Chapter 6.

Chapter 1

The Standard Model of Particle Physics

Elementary particle physics aims to understand the structure of matter at the microscopic level and explain the interactions between them. The most thorough theory currently known which describes particles and forces is the Standard Model of elementary particle physics [1, 2]. Despite of some unanswered questions, the Standard Model has been able to describe the experimental data since it was proposed in sixties. Three out of four interaction forces which have been observed in nature are formulated within the Standard Model. The particles predicted within the model have already been discovered in research laboratories, except for a missing particle which is the mass generator of the model. The heaviest discovered elementary particle is the top quark which was observed in 1995 [3–7] at the Tevatron collider [8]. The parameters of the model have been measured with great precision and efforts are made in order to discover the last predicted particle of the model, being the Higgs particle. The most recent experiment which has been constructed to search for the missing piece of the Standard Model as well as providing answers to other formulated theoretical models beyond the Standard Model, is the Large Hadron Collider (LHC). The two incoming proton beams are currently collided at the center of mass energy of 7 TeV to provide the experimental data for further analysis.
In this chapter, the Standard Model of particle physcis is reviewed. The way of formulating the particles and forces into the model is described in Section 1.1. Special emphasis is given to the top quark sector of the Standard Model since the top quark is the heaviest elementary particle discovered so far with distinct features compared to the other known elementary particles. Hence, Section 1.2 is dedicated to the physics of the top quark and elaborates its properties.

1.1 Elementary Particles and Their Interactions

1.1.1 Electroweak Theory

The fundamental particles, namely quarks and leptons, which are fermions with half-integer spin are represented by Dirac spinors ψ, that fulfill the Dirac equation [1]

$$i\gamma^{\mu}\partial_{\mu}\psi = 0.$$

In the presence of an external electromagnetic field A_{μ}, the fermions couple to the quantum of the field, which is called the photon γ, and their interaction are described by a Lagrangian term expressed as

$$\mathcal{L}_{e.m}^{int} = \sum_{f} \bar{\psi}_f i\gamma^{\mu} D_{\mu}\psi_f,$$

where the sum is over all fermions and D_{μ} is the covariant derivative defined as $D_{\mu} = \partial_{\mu} + ieQA_{\mu}$. In the language of group theory, the dirac spinors are representations of the $U(1)_{e.m}$ group since the Lagrangian is invariant under the gauge transformation

$$\psi \rightarrow e^{ieQ}\psi.$$

According to Noethers's theorem, there is a conservation law that can be associated to this gauge symmetry. Therefore the electric charge Q, which is expressed in units of quantum number e, is constant during the electromagnetic interactions. The Dirac spinors associated to the elementary particles and their corresponding charge are listed in Table 1.1.

As expressed in Table 1.1, the leptons as well as the quarks appear in three generations whose constituents are equal in all quantum numbers but differentiated in terms of mass. The most stable particles are the ones listed in the first generation for each of the two category of fermions. The elementary particles interact with the electromagnetic field via their charge, hence no interaction vertex is present between the photon and the neutrino in the Standard Model since the neutrinos are neutral fermions.
In addition to the electromagnetic interactions which are formulated in a simple Lagrangian term and can fully describe the interaction of the charged particles, there is another kind of interaction describing the phenomena like beta decay $p \rightarrow n + e + \nu_e$, which corresponds to the physics processes containing a neutrino as one of the components of the interaction. This kind of interaction, known as the weak interaction, contains a heavy particle as propagator. Because of the rather massive property of the propagator of the weak interaction, they are difficult to be produced during an interaction. Therefore, the weak interactions are rare within the Standard Model.
In order to describe the physics of the weak interactions, an appropriate term should be added to the Lagrangian. According to the experimental data, where only left-handed fermions contribute to the weak interaction, the weak current $\bar{\psi}_{\nu_e}\gamma^{\mu}(1-\gamma^5)\psi_e$

[1] The mass term is dropped from the Dirac equation for simplicity.

fermions		Q
leptons	first generation $\begin{pmatrix}\psi_e\\ \psi_{\nu_e}\end{pmatrix}$	-1 0
	second generation $\begin{pmatrix}\psi_\mu\\ \psi_{\nu_\mu}\end{pmatrix}$	-1 0
	third generation $\begin{pmatrix}\psi_\tau\\ \psi_{\nu_\tau}\end{pmatrix}$	-1 0
quarks	first generation $\begin{pmatrix}\psi_u\\ \psi_d\end{pmatrix}$	$+\frac{2}{3}$ $-\frac{1}{3}$
	second generation $\begin{pmatrix}\psi_c\\ \psi_s\end{pmatrix}$	$+\frac{2}{3}$ $-\frac{1}{3}$
	third generation $\begin{pmatrix}\psi_t\\ \psi_b\end{pmatrix}$	$+\frac{2}{3}$ $-\frac{1}{3}$

Table 1.1: The elementary particles, being quarks and leptons, predicted by the Standard Model and observed in the experiment. The electric charge Q, associated to each particle is also listed.

was proposed, as the operator $(1-\gamma^5)$ is the chirality projector that only keeps the left-handed part of the Dirac spinor, for example one can have

$$P_L\psi_e \equiv \frac{(1-\gamma^5)}{2}\psi_e = \psi_{e,L}.$$

In order to keep the Lagrangian invariant under the gauge transformation, the $SU(2)_L$ symmetry group is proposed for which the new doublets Ψ, constructed from the left-handed Dirac spinors, are the representations of the gauge group which are expressed explicitly for the first generation of the fermions as

$$\Psi_{\text{Lepton}} = \begin{pmatrix}\psi_{\nu_e,L}\\ \psi_{e,L}\end{pmatrix} \text{ and } \Psi_{\text{Quark}} = \begin{pmatrix}\psi_{u,L}\\ \psi_{d,L}\end{pmatrix},$$

and the Pauli matrices τ^i, are the generators of the group. The three massless gauge bosons which are introduced in this group are not able to explain the experimental data, being the three massive particles $W^\pm$ and Z. Hence the idea of unification of the electromagnetism and weak interation was made to combine the two type of physics interactions into one unified theory called electroweak which is explained in the following.
In order to combine the electromagnetic and weak interactions into a single formalism, a new symmetry group $U(1)_Y$, called hypercharge, is introduced and the Lagrangian is modified and expressed as

$$\mathcal{L}_{electroweak} = \sum_f \left[\bar{\Psi}_f i\gamma^\mu D_\mu \Psi_f + \bar{\psi}_{f,R} i\gamma^\mu D_\mu \psi_{f,R}\right] + \text{Kinetic terms.},$$

where the sum is over all fermions and R refers to the right-handed Dirac spinors which are obtained by applying the right-handed chirality operator $\frac{(1+\gamma^5)}{2}$, on the Dirac spinors. The definition of the covariant derivative D_μ, is different depending on which operand it is acting. The covariant derivatives of doublets and singlets are respectively expressed as

$$D_\mu = \partial_\mu + ig_Y Y B_\mu + ig_W \frac{\tau^i}{2} W^i_\mu,$$

and

$$D_\mu = \partial_\mu + ig_Y Y B_\mu,$$

where Y is the hypercharge operator. The constants g_Y and g_W represent respectively the coupling strengths of the hypercharge and weak intractions. The three W^i and one B gauge bosons are combined to express the observed physical intermediate bosons γ, $W^\pm$ and Z, as explained below.
As a consequence of introducing the new hypercharge symmetry group, the electroweak part of the Standard Model can be described on the basis of the $SU(2)_L \otimes U(1)_Y$ gauge group. Therefore, the left-handed fermions in doublets and right-handed fermions in singlets are representations of this new symmetry group. The quantum numbers which are assigned to the hypercharge operator Y, as well as the weak isospin operators, being the Pauli matrices, are determined in such a way that the appropriate electric

charge of fermions is recovered. As an example, only the first generation of the leptons is considered. The Lagrangian equation can be expanded as

$$\begin{pmatrix}\bar{\psi}_{\nu_e,L} & \bar{\psi}_{e,L}\end{pmatrix}\gamma^\mu\Big[-g_Y Y B_\mu - g_W\frac{\tau^3}{2}W_\mu^3\Big]\begin{pmatrix}\psi_{\nu_e,L}\\ \psi_{e,L}\end{pmatrix}+\bar{\psi}_{e,R}\gamma^\mu\Big[-g_Y Y B_\mu\Big]\psi_{e,R},$$

where only the part responsible to re-produce the electromagnetic interaction is kept. The two gauge bosons B_μ and W_μ^3 are combined to make the new physics gauge bosons A_μ and Z_μ, according to

$$\begin{pmatrix}A_\mu\\ Z_\mu\end{pmatrix}=\begin{pmatrix}\cos\theta_W & \sin\theta_W\\ -\sin\theta_W & \cos\theta_W\end{pmatrix}\begin{pmatrix}B_\mu\\ W_\mu^3\end{pmatrix},$$

where θ_W is the Weinberg mixing angle. Therefore the Lagrangian expression is simplified and given by

$$\begin{pmatrix}\bar{\psi}_{\nu_e,L} & \bar{\psi}_{e,L}\end{pmatrix}\gamma^\mu A_\mu\Big[-g_Y\cos\theta_W Y-\frac{1}{2}g_W\sin\theta_W\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\Big]\begin{pmatrix}\psi_{\nu_e,L}\\ \psi_{e,L}\end{pmatrix}+\bar{\psi}_{e,R}\gamma^\mu A_\mu\Big[-g_Y\cos\theta_W Y\Big]\psi_{e,R}.$$

With the assumption of $e=g_Y\cos\theta_W=g_W\sin\theta_W$ and assigning

$$Y\begin{pmatrix}\psi_{\nu_e,L}\\ \psi_{e,L}\end{pmatrix}=-\frac{1}{2}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}\psi_{\nu_e,L}\\ \psi_{e,L}\end{pmatrix},$$

and

$$Y\psi_{e,R}=-1\psi_{e,R},$$

the Lagrangian expression would be simplified to

$$+eA_\mu\Big[\bar{\psi}_{e,L}\gamma^\mu\psi_{e,L}+\bar{\psi}_{e,R}\gamma^\mu\psi_{e,R}\Big]=+eA_\mu\bar{\psi}_e\gamma^\mu\psi_e.$$

Therefore, the electromagnetic interaction term with the right electric charge is recovered given that a correct value for the hypercharge quantum number is chosen.
The problem that arises is that the observed weak bosons are massive while an explicit mass term cannot be introduced into the Lagrangian as it spoils the gauge symmetry. The Higgs mechanism [9–11], which is described in detail in the following, is supposed to be responsible for giving mass to the particles in the Standard Model.

Electroweak Symmetry Breaking

Since the observed weak bosons are massive, the electroweak symmetry must be broken. The formalism of symmetry breaking is introduced with the use of the Higgs mechanism. A new doublet containing scalar fields is introduced as

$$\phi=\begin{pmatrix}h^1\\ h^2\end{pmatrix}$$

whose hypercharge quantum number is chosen to be $+\frac{1}{2}$. The corresponding Lagrangian term which describes the dynamic of the scalar field ϕ is written as

$$\mathcal{L}_{Higgs}=\frac{1}{2}D_\mu\phi^\dagger D_\mu\phi-V(\phi),$$

where $V(\phi)$ is the scalar potential and is given by $-\mu^2\phi^2 + \lambda\phi^4$. The symmetry is spontaneously broken when the scalar field takes a value of v at the minimum of the potential, defined as $v^2 \equiv \frac{\frac{1}{2}\mu^2}{\lambda}$. Due to the quantum fluctuations around the minimum of the potential, one can have

$$\phi = \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix},$$

where $h(x)$ corresponds to the Higgs boson field, for which there is no experimental evidence yet. Expanding the Lagrangian expression around the vacuum state, the mass terms for the physical observables $W^{\pm}$ and Z are obtained to be

$$m_W = \frac{1}{2} g_W v; \;\; m_Z = \frac{1}{2}\sqrt{g_Y^2 + g_W^2}\, v,$$

while the photon remains massless, as aimed for. All the parameters g_W, g_Y and v are free parameters of the model [2].

The Higgs mechanism results in giving mass to the weak gauge bosons of the Standard Model. As a result, the fermions are still massless because an explicit mass term would violate the gauge invariance of the electroweak Lagrangian as it mixes the left and right-handed components of the Dirac spinors

$$m_f \bar{\psi}_f \psi_f = m_f \Big[\bar{\psi}_{f,L} \psi_{f,R} + \bar{\psi}_{f,R} \psi_{f,L} \Big].$$

Therefore the procedure of giving mass to the fermions proceeds by introducing the Yukawa interaction terms which make use of the Higgs doublet in order to construct the gauge invariant expressions out of Dirac spinors. For example, the first generation of the leptons acquire mass when introducing the following gauge invariant terms

$$\mathcal{L}_{Yukawa} = -G_e \Big[\bar{\Psi}_L \phi \psi_R + \bar{\psi}_R \phi^{\dagger} \Psi_L \Big],$$

where G_e is the Yukawa coupling constant for the electron field. The Yukawa coupling constants which are introduced to the theory through $\mathcal{L}_{Yukawa}$, are again arbitrary parameters of the Standard Model.

As a consequence, the Higgs particle couples to a fermion with a strength proportional to the mass of that fermion.

1.1.2 Strong Interactions

In addition to the electromagnetic and weak interactions of the particles within the Standard Model, there is another kind of interaction, called the strong interaction, which are specific to the quarks carrying color charges. There are three colors which resulted in proposing a new gauge group $SU(3)_c$ to be responsible for the strong interactions. The quarks are triplets under the gauge transformations of the $SU(3)_c$

[2] The three best measured electroweak parameters, being the electromagnetic coupling constant α, the Fermi constant G_F and the mass of the Z boson m_Z, are used to determine these three parameters g_W, g_Y and v. The parameter μ, representing the mass of the Higgs boson remains undetermined in the theory.

symmetry group and the eight gauge boson, G^a, which are introduced accordingly are the massless gluons, the intermediate bosons of the strong interactions.
The covariant derivative which contains the information of the strong interaction is then expressed as

$$D_\mu = \partial_\mu + ig_Y Y B_\mu + ig_W \frac{\tau^i}{2} W^i_\mu + ig_S \frac{\lambda^a}{2} G^a_\mu,$$

where the g_S, being another free parameter of the model, is the strong coupling constant and λ^a are the Gell-Mann matrices which are the generators of the color $SU(3)_c$ group.
The Higgs mechanism and the Yukawa interactions are as before and the $SU(3)_c$ remains unbroken after the electroweak symmetry breaking.

1.2 Top Quark Sector of the Standard Model

1.2.1 Indirect Evidence for the Existence of the Top Quark

Before its discovery in 1995, the top quark was predicted by the Standard Model. In order to obtain a renormalisable gauge theory, the anomalies arrising from the so-called triangle diagrams of which an example can be found in Figure 1.1, should cancel. Since each triangle is proportional to $c_A^f Q_f^2$, where Q_f is the charge of fermion and c_A^f is the axial coupling of the weak neutral current, the total anomaly can be obtained by

$$\sum_{i=1}^{N} = \Big(\frac{1}{2} \times (0)^2 - \frac{1}{2} \times (-1)^2 + \frac{1}{2} \times 3 \times (+\frac{2}{3})^2 - \frac{1}{2} \times 3 \times (-\frac{1}{3})^2\Big),$$

where N denotes an equal number of lepton and quark doublets. Therefore in order to have a vanishing anomaly, it was required that the quark generations come in pairs to compensate the divergencies from the fermion loops.

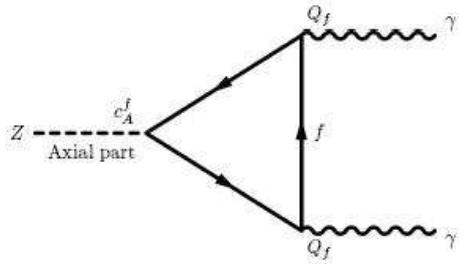

Figure 1.1: The triangle diagram containing a fermion loop which produces divergences in the quantum field theory.

Another indirect evidence for the existence of the third generation of quarks, hence the top quark as one of the components, was to provide a natural way to suppress theoretically flavour changing neutral current phenomenon (FCNC) which is not experimentally observed. Based on the GIM mechanism [12] which was originally proposed for two quark generations, the same argument which forbids flavour changing

transitions among down-type quarks, applies for three quark doublets.
Also there is lots of evidence for the existence of the top quark. There have been some detailed studies of the $Zb\bar{b}$ vertex using the data collected at the e^-e^+ colliders LEP and SLC, resulting in the measurement of the isospin of the b quark [13]. It has been measured [14] that the b quark has a left-handed third component of the isospin of $-\frac{1}{2}$ which implies that the b quark must have a weak isospin partner. In other words, the top quark must exists as a counterpart of the isospin doublet of the b quark.

1.2.2 Indirect Constraints on the Mass of the Top Quark

All electroweak quantities of the Standard Model depend, at leading order, on three parameters, namely two gauge coupling constants g_Y, g_W and vacuum expectation value of the Higgs boson field v. At higher order of corrections, two extra parameters, being the top quark mass m_t and the mass of the Higgs boson m_H, contribute to the radiative loop corrections and therefore can affect the values of the electroweak observables. Hence, the precision electroweak measurements can provide indirect constraints on the mass of both the top quark and the Higgs boson.
While the dependency of the electroweak observables such as the mass of the Z boson on the mass of the top quark is quadratic, the mass of the Higgs boson appears into the radiative corrections through logarithmic terms. Therefore, the inferred constraints on m_t are stronger than those on m_H and that is why there has been a good prediction on the mass of the top quark before its discovery. A global fit of the Standard Model to the precision data successfully predicts a value for the mass of the top quark as [15, 16]

$$m_t = 179.4^{12.1}_{-9.2}\, GeV\,^3,$$

which is in good agreement with the measured value of the top quark [17]

$$m_t = 173.2 \pm 0.9\, GeV.$$

The successful prediction of the mass of the top quark provided by the precision electroweak data indicates the predictive power of the radiative corrections which persuades one to obtain constraints on the other parameters of the model, such as Higgs boson mass. As already mentioned, because of the logarithmic dependence of the radiative corrections on the mass of the Higgs boson, it is not possible to derive stringent predictions on the Higgs boson mass. In the following, the results of the Standard Model fit to the precision electroweak data in order to extract constraints on m_H are discussed. Using the measured value of the top quark mass, the Standard Model fit to the electroweak precision data is performed in order to infer the mass of the Higgs boson [18]. The result of the fit, expressed in terms of $\Delta\chi^2$, is shown in Figure 1.2. The most likely value of the mass of the Standard Model Higgs boson is determined from the minimum of the $\Delta\chi^2$ curve to be $m_H = 91^{+45}_{-32}\, GeV$, which means the data prefers a light Higgs boson. The direct searches for the Higgs boson at LEP have exculded a Higgs boson with a mass below 114.4 GeV [19], therefore the preferred value is slightly above the exclusion limit.

[3] Throughout the thesis, the natural units are used.

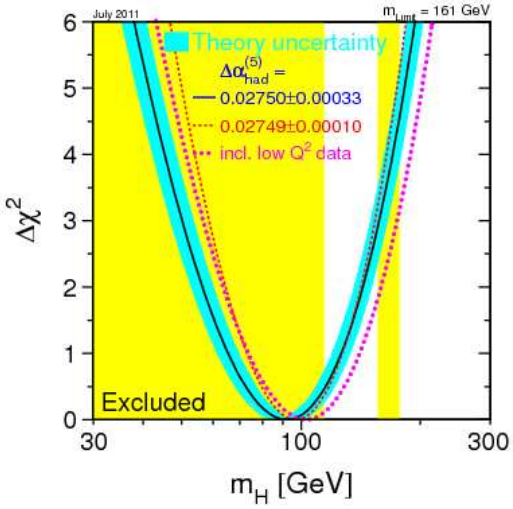

Figure 1.2: The indirect determination of the mass of the Higgs boson obtained from the Standard Model fit to the electroweak data assuming the world average mass of the top quark [18]. Also the exclusion region on m_H obtained from direct searches for the Higgs boson at LEP is also shown.

Figure 1.3 contains the information of the direct as well as the indirect measurements of both the mass of the top quark and the W boson. Also the lines indicating the various Higgs boson masses are overlaid. The upper limit of 1000 GeV on the mass of the Higgs boson comes from the theoretical calculation [20]. The direct searches for the Standard Model Higgs boson at the Tevatron, result in an exclusion of the mass of the Higgs particle in a narrow band specified on the plot [21].
As it can be seen from Figure 1.3, the direct and indirect measurements of m_t and m_W are in good agreement, confirming that the Standard Model of particle physics is not obviously wrong. It should also be noted that the contours of the W and top quark masses are calculated at 68% confidence level. In order to make stronger limits on the measured values, more data should be provided. Therefore, a more precise determination of either the mass of the top quark or the mass of the W boson, would make it possible to accept or reject the existence of the Higgs particle within the Standard Model.

1.2.3 Direct Searches of the Top Quark and the Discovery Era

As mentioned before, the top quark was finally discovered by the CDF [22] and D0 [23] experiments at the Tevatron. The latest world average value measured for the mass of the top quark can be found in Figure 1.4. The indirect measurements from precision data are also shown in the same plot. Besides the good agreement between the direct and indirect measurements of the mass of the top quark, it can be also seen that the mass of the top quark is measured with an uncertainty of 0.5% on the mean value which represents a precise measurement of the mass of the top quark. Even with more

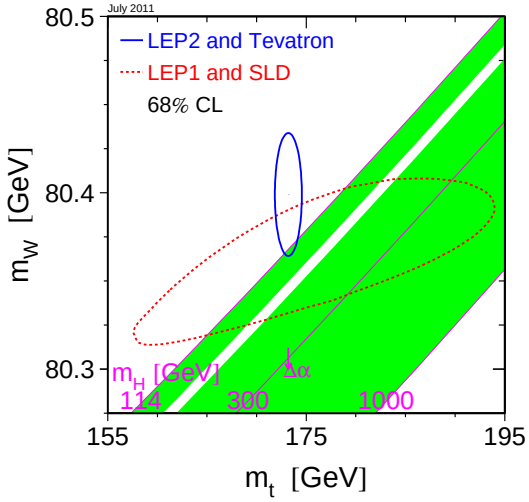

Figure 1.3: The direct (solid ellipse) and indirect (dotted ellipse) measurements of m_W and m_t determinde from the data collected at the Tevatron and LEP colliders. The solid lines represent the various configurations of the Higgs boson mass.

data available from the LHC, it is possible to improve the accuracy on the quoted value for the mass of the top quark.

In addition to the mass, the top quark has other quantum numbers which should be measured directly from the data. The electric charge, for example, is one of the important quantum numbers that must be checked. Since the center of mass energy provided by the LEP collider was around 200 GeV, much less than the energy needed to produce a pair of top quarks via the interaction $e^+e^- \rightarrow \gamma^* \rightarrow t\bar{t}$, it was not possible to use the collected data from the electron-positron collider to measure the electric charge associated to the top quark. At a proton-proton collider it is still possible to measure the electric charge. Because the electric charge measurement requires the top quark to interact with a photon, then it was suggested to search for the interactions in which a top is produced together with a radiative photon [24]. According to the studies performed based on the data accumulated from the Tevatron, it has been shown the possibility that the top quark is an exotic charge $Q_t = -\frac{4}{3}$ quark can be ruled out at about 95% confidence level [25]. Hence the top quark which is produced at hadron colliders and decays to a W boson and a b quark $t \rightarrow Wb$, cannot have $Q_t = -\frac{4}{3}$, but it is required to have $Q_t = +\frac{2}{3}$.

1.2.4 Top Production and Decay at Hadron Colliders

At hadron colliders, top quarks are produced predominantly via the strong interaction. Depending on the center of mass energy, either gluon-gluon fusion or quark-antiquark annihilation is the main channel which is responsible for top quark pair production.

Due to its huge mass, the top quark has a very short life time, being 10^{-25} s, which is an order of magnitude shorter than the time scale of the hadronization process. As

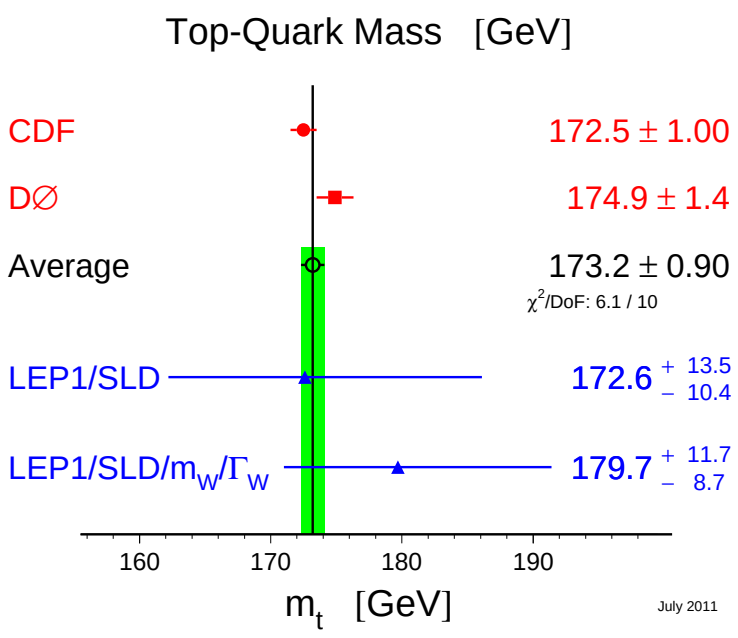

Figure 1.4: The latest results of the direct measurement on the mass of the top quark provided by both experiments at the Tevatron collider together with world average of the measured value of m_t. The indirect measurement from the Standard Model fit to the electroweak data is also quoted.

a result, the top quark decays before getting hadronized. According to the Cabibbo-Kobayashi-Maskawa (CKM) matrix [26, 27] which accounts for the possibility of quark mixing states, the probability of a top quark decaying into its counterpart is close to 100% [28]. Hence top quarks always decay to a W boson and a b quark, $t \rightarrow Wb$. The W boson can subsequently decay to quarks or leptons, which mark the various decay channels of a $t\bar{t}$ system. This is explained in more detail in Section 4.2.1.

1.3 Problems of the Standard Model

The Standard Model of particle physics is a self-consistent gauge field theory which is in good agreement with the observed experimental data. However, as there are so many free parameters within the theory, whose values cannot in principle be determined, this makes it hard to state that the Standard Model is the most fundamental theory which discribes nature. In other words, in order to understand why the masses of the quarks, leptons and the W and Z gauge bosons have their observed values, there might be a theory beyond the Standard Model from which the Standard Model Lagrangian can be derived.

Another question that is brought to mind is the following. The mass of the top quark, which is the heaviest observed elementary particle of the model, is so close to the scale of the electroweak symmetry breaking $v \sim 246\ GeV$. This opens the question if the top quark becomes massive during the spontaneous symmetry breaking mechanism or interacting with another gauge field results in giving a mass to the top quark.

Also, due to the so-called hierarchy problem, there is a huge gap between $M_Z \sim 90\ GeV$, the energy scale of the weak interaction of the model, and the Plank mass scale $M_{Pl} \sim 10^{19}\ GeV$, which is the fundamental scale of gravity. Then, there have been proposals that the current Standard Model is only an effective version of a unified theory with an energy scale of the order of TeV. Another question which remains unanswered within the Standard Model is the lack of a candidate for dark matter. Therefore, with so many open questions in the model, some new extentions to the Standard Model have been proposed, such as supersymmetry. The Minimal Supersymmetric Standard Model (MSSM), predicts a candidate for dark matter and also provides an answer to the hierarchy problem. In order to confirm the validity of such extentions to the model, colliders with higher energies are required.

Besides the above mentioned challenges of the model, there is another important one to be noted. Since there has not yet been any experimental evidence for the existence of the Higgs particle, the procedure of giving mass to the weak bosons is not experimentally confirmed. Given that a Higgs boson is not discovered, the Standard Model of particle physics is still correct but needs some mechanism to explain the origin of mass in nature. That is why the next generation of the accelerator machines, such as the LHC, is constructed to run at higher center of mass energies. One of the important goals of the LHC project, in addition to provide information on the presence of any new physics beyond the Standard Model, is to search for the only missing particle predicted by the Standard Model. If a Higgs particle exists with a mass less than 1 TeV, the LHC would be able to discover it.

Chapter 2

The CMS experiment at the LHC

The Standard Model of particle physics is the most precise theory describing the matter and the forces, which has been widely tested at high energy physics colliders. Despite of the good agreement presented between the theory and the experiment, there is still some part for which there has not yet been any experimental evidence, being the mechanism of the electroweak symmetry breaking. The discovery of the Higgs boson would confirm the way of giving mass to the particles described by the theory. Therefore, the idea of constructing the Large Hadron Collider (LHC) was introduced to mainly search for the missing Higgs particle at higher energy scales of physics which had not yet been reached.
The giant LHC machine is described in more detail in Section 2.1 while the description of one of the general-purpose detectors of the LHC, namely the CMS experiment, is given in Section 2.2.

2.1 The Large Hadron Collider

The Large Hadron Collider [29] (LHC) is a proton-proton accelerator installed in a tunnel of 26.7 km circumference which was originally built for the Large Electron-Positron collider [30] (LEP). The tunnel is constructed underground, on average about 100 m below the surface and is located on the border between Switzerland and France, near Geneva. The LHC, employed by the European Organisation of Nuclear Research (CERN), is currently running at the center of mass energy of $\sqrt{s} = 7\ TeV$ which is a factor of two below the design value of $\sqrt{s} = 14\ TeV$. Accelerating beams of particles up to several TeV, the LHC has become the most powerful collider machine. The first proton-proton collisions at a center of mass energy of 7 TeV happened in March 2010. In addition to colliding beams of protons, the LHC also provides ion-ion collisions [31]. Lead nuclei are used as the heavy ion in these collisions.
The idea of constructing a hadron collider was approved at 1995 by CERN. It was decided to use the same pre-accelerating and underground facilities which had already been built to serve the LEP experiment [32]. The choice of colliding protons instead of electrons is due to some physics reasons. An important source of energy loss of a

rotating particle when moving around a circle of radius R, is synchrotron radiation which is expressed as

$$\Delta E \propto \frac{1}{R}\left(\frac{E}{m}\right)^4,$$

where E is the particle energy and m is the particle mass. Therefore, it is obvious that heavier particles loose a smaller amount of energy compared to less heavier particles. As the bending radius was already fixed by choosing the same tunnel as LEP, using protons instead of electrons would yield to obtain a more stable beam of rotating particles and consume less power.
Despite of providing beams of particles with less energy loss per turn, proton-proton colliders have their own problems as well. Since protons are composite particles, then to achieve physics goals up to an energy scale of s_1, the collider needs to provide a center of mass energy of s_2 which must be higher, hence $s_2 \gg s_1$. This is due to the fact that the proton constituents carry a fraction of the energy of the proton and the interaction takes place between these constituents. The decision of operating the LHC at a center of mass energy of a few TeV is to make it possible to scan the energy scale of physics interactions up to 1 TeV scale. Since one of the main goals of the construction of the LHC is to search for the Standard Model Higgs boson, the theory implies an upper limit on the mass of the Higgs boson to be less than one TeV. Hence a hadron collider with a center of mass energy of a few TeV would allow to discover that missing piece, if it exists.
Prior to their path in the LHC, the proton beams go through the CERN pre-accelerator machines which are explained in Section 2.1.1. The LHC has various detectors, each of which is specialized for particular physics purposes that are discussed briefly in Section 2.1.2.

2.1.1 The Proton Accelerator Chain

Before being injected into the LHC, protons are pre-accelerated through the CERN accelerator complex [33] depicted in Figure 2.1.

The current acceleration facilities have served for decades when providing beams of electrons and positrons to the LEP collider. For the LHC operation, they have been upgraded to provide beams of protons for collisions at unprecedent energies.
A duoplasmatron, where emitted electrons from a cathode filament ionize the hydrogen gas, provides the source of protons for the LHC. The produced protons are then transfered to the first stage of the accelerator chain, the linear accelerator called Linac, which is based on radio frequency technology. In a Linac machine, the acceleration occurs by oscillating an electric field with an appropriate frequency. Due to this reason, particles traverse through the accelerator discontinuously while making so-called bunches, each of which consists of thousands of particles. When leaving the Linac, the protons reach an energy of 50 MeV.
Energetic protons are fed to a booster prior to be injected into the Proton Synchrotron (PS) for further acceleration. The booster, which is the first circular accelerator in the acceleration chain, enhances the energy of the protons to 1.4 GeV. Subsequently, the protons are more accelerated while circulating in the 630 m circumference of the PS ring

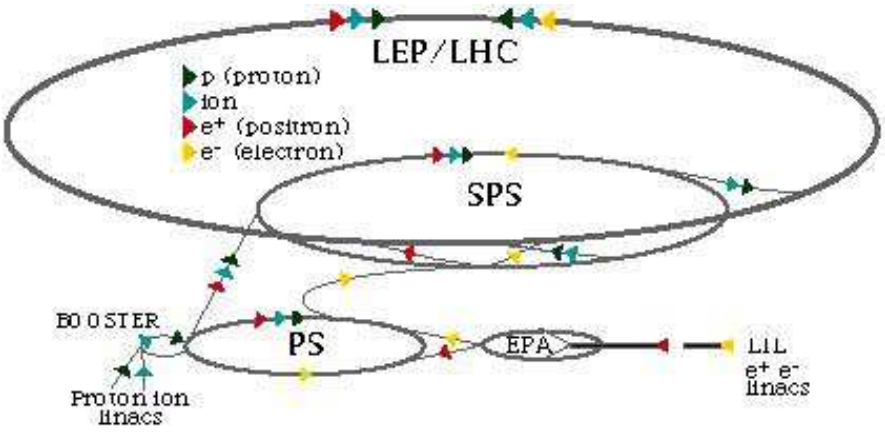

Figure 2.1: The CERN accelerator complex comprising the full particle accelerator chain. Starting on the left hand side, protons are accelerated while moving through the linear accelerator (Linac), the booster, the Proton Synchrotron (PS), and the Super Proton Synchrotron (SPS). Finally, the protons are injected into the Large Hadron Collider to reach the desired energy.

to reach an energy of 26 GeV, and are then fed to the Super Proton Synchrotron (SPS). The SPS, which is the final stage in the succession of the pre-accelerator complex, provides protons with an energy of 450 GeV and inject them into the LHC machine for final acceleration and subsequently collisions.

The proton beams finally enter the LHC tunnel which contains two beam pipes. The injected beams of particles move inside the pipes in opposite directions and cross each other at four points of the ring, where the detectors are located. The beams are guided to travel on a circular path with the use of a strong magnetic field, which is obtained using superconducting electromagnets. Each superconducting electromagnet, made from coils of superconducting wire, has a length of about 15 m and a weight of around 27 tonnes. To be operational, the temperature must be lowered to values of the order of a few Kelvin. The dipole magnets could provide a magnetic field of about 8.3 T which keeps the beams to move around the ring with a designed energy of about 7 TeV per beam.

While more than 1200 dipole magnets are used to make the proton beams circulated, an additional 400 quadrupole magnets are used in order to make the proton beams focused. It is important to have very narrow beams when they are crossing, as it increases an important parameter of the machine, which is called the luminosity $\mathcal{L}$ and defined as [34]

$$\mathcal{L} = f\frac{N_A N_B}{4\pi\sigma_x\sigma_y},$$

where σ_x and σ_y represent the beam profiles in horizontal and vertical directions at the interaction points. Also, N_A and N_B are the number of particles per bunch in the beams of type A and B, respectively. The f parameter is the frequency of crossing of the bunches of the two beams. It is obvious that an increased number of particles per bunch or more focused beams would result into a larger luminosity.

The luminosity of the machine is important since it represents how many physics events occur in an specific time period T. It is related to the production cross section σ, according to

$$N = \int_0^T \mathcal{L} dt \sigma,$$

where $\int_0^T \mathcal{L} dt$ represents the collected data in a period of time equivalent to T. Hence, for rare processes such as the interactions where a Higgs particle is produced, a higher luminosity would increase the probability of being observed in the experiment.
The LHC has been designed to operate with a luminosity of $\mathcal{L} \sim 10^{34}\ cm^{-2}s^{-1}$ which is at least two orders of magnitude higher than the luminosities which have been reached by other hadron colliders. In order to run at this high luminosity, it is required to obtain a number of protons per bunches to be of the order of 10^{11}. The bunches of protons are collided every 25 ns, corresponding to a bunch crossing frequency of 40 MHz. This would introduce the need for high-speed processors to select the interesting physics events and also large storage facilities to collect the selected data.
Since the start of the LHC operation in March 2010, it has been providing proton-proton collisions at a center of mass energy of 7 TeV. On November 2010, the machine stopped taking proton collision data and for the next one month, it has been colliding heavy ions. During the 2010 runs, the LHC machine delivered about 47.0 pb^{-1} of data to each of its experiments [35], while the data which have been recorded by the CMS experiment correspond to an integrated luminosity of 43.2 pb^{-1}, of which 36.1 pb^{-1} is certified to be used for analyses. The distribution of the total integrated luminosity delivered by the LHC and recorded by the CMS experiment as a function of time, is shown in Figure 2.2. As quoted on the plot, the total integrated luminosity corresponds to the proton-proton collisions taken in 2010.

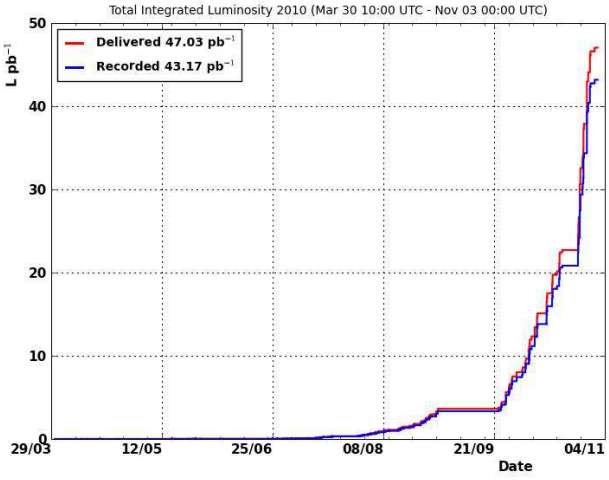

Figure 2.2: The total integrated luminosity delivered by the LHC and recorded by the CMS experiment during the proton-proton collisons in 2010.

The results of analyzing the 36.1 pb^{-1} of the proton-proton collision data in the framework of this thesis are given in detail in Section 5.6.
Currently the LHC machine is operating at a center of mass energy of 7 TeV and the CMS experiment has already been able to record an integrated luminosity of proton collisions equivalent to more than 2000 $pb^{-1} = 2\ fb^{-1}$ in 2011, hence 55 times more data than what had been collected in 2010. The LHC has been designed to operate at an even higher center of mass energy, being two times larger than the one achieved. It has been decided [36, 37] that the machine continues running at a center of mass energy of 7 TeV at least until the end of 2011.

2.1.2 Particle Detectors at the LHC

Four main experiments detect collisions which occur at the LHC. Two of them, being CMS (Compact Muon Solenoid) [38] and ATLAS (A Toroidal LHC ApparatuS) [39], are general-purpose detectors. They are designed to address a wide variety of physics questions. Searching for the Higgs particle, exploring physics at TeV scale and looking for evidence of physics beyond the Standard Model such as supersymmetry are some of the goals which are followed by the multi-purpose experiments of the LHC.
The other two, ALICE (A Large Ion Collider Experiment) [40] and LHCb (Large Hadron Collider beauty) [41], are special-purpose detectors. While the former is dedicated to heavy ion collisions, the latter is specialized to the studies of b quark physics. The location of these main experiments around the main ring of the LHC is shown in Figure 2.3.

In addition to the four main experiments, the LHC hosts two more detectors. A first one, which is named LHCf (Large Hadron Collider forward) [42], is intended to study the particles that are produced in the forward region of the interaction point of proton collisions. It consists of two pieces which are installed on either side of the ATLAS detector, 140 m away from the interaction point. A second one, which is named TOTEM (TOTal Elastic and diffractive cross section Measurement) [43], is designed to measure the total proton-proton cross section. It is located on the forward region of the CMS detector.

2.2 The CMS Experiment

Detectors are constructed in order to observe and analyze the collisions which are provided by the collider. Exploring the vast physics opportunities in the proton-proton collisions at high energy scales reachable at the LHC, requires to build general-purpose particle detectors such as the CMS experiment. Particle detectors are generally designed to be symmetric in space to make the physics analysis easier. The particle is detected when it interacts with matter. According to the various kind of particle interactions, the detectors contain appropriate materials, each of which is aimed to identify some special type of particles or to measure particular properties of the particles. The sub-systems are arranged in such a way which yields the detector to have an onion-like structure. A slice of the transverse view of the CMS detector is shown in Figure 2.5.

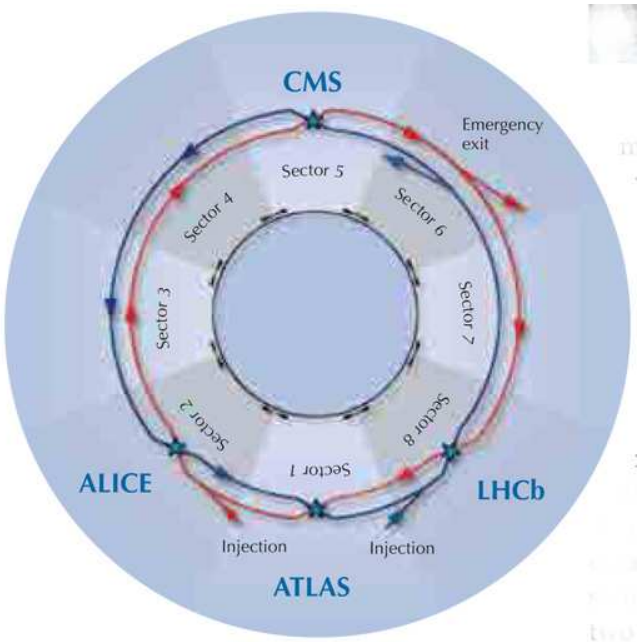

Figure 2.3: The location of the various detectors around the LHC main ring. Also shown are the injection points which are used to enter the proton beams into the LHC collider and the emergency exit points which are used to extract the beams from the machine.

Emerging particles from the hard interaction first enter the tracking system. The charged particles, when moving through the tracker material leave hits which are finally used to build the trajectory of that particle. The presence of a magnetic field with parallel field lines to the beam axis, bends the trajectories of the charged particles. The bending radius is subsequently used to determine the momentum of the particles as well as their charges.
Right after the tracker, the calorimetry system exists which comprises the electromagnetic calorimeter (ECAL) and the hadronic calorimeter (HCAL), aimed to measure the energy of the particles which interact via the electromagnetic and strong interactions, respectively. A high energy electron or photon, hits an absorber and initiates a cascade of secondary electrons and photons via bremstrahlung and pair production as depicted in Figure 2.4. Such an electromagnetic shower can be characterized with mainly two parameters, namely the radiation length X_0 and the Moliere radius R_M, which are explained in the following.

The longitudinal development of an electromagnetic shower is described by the parameter X_0 which is defined as the distance over which the electron loses about $\frac{2}{3}$ of its energy. This parameter is also a characteristic length scale which determines the thickness of the ECAL. The lateral development is described by the parameter R_M. An infinite cylinder centered on the electromagnetic shower initiator with a radius of $\sim$ 1 R_M, contains $\sim 90\%$ of the shower energy.
Similar to the electromagnetic showers, hadronic showers are formed by the strong

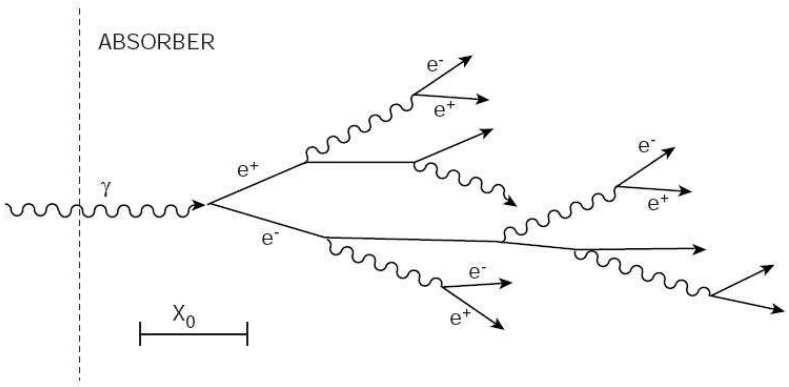

Figure 2.4: A typical electromagnetic shower which is initiated by a photon and is developping by bremstrahlung and pair production processes.

interaction of hadrons with the heavy nuclei of the material used in the HCAL subsystem. Analogous to X_0, the typical length scale parametrizing the thickness of the hadronic calorimeter is the nuclear interaction length λ which is defined as the mean free path length of the hadron before interacting with the material.
At the outermost part of the CMS detector, the muon system is used which is designed to detect and identify muons. Due to the rather large mean free path, muons are only interacting weakly with the calorimeter material and deposit a small amount of their energy in the calorimeter system and penetrate the last layer of the detector, which is aimed to detect muons. Since one of the important discovery channels of the Standard Model Higgs particle is via $pp \rightarrow H \rightarrow ZZ^* \rightarrow \mu\mu\mu\mu$, a sub-detector with a high resolution has been dedicated for muon identification and measurements. The importance of this has affected the naming of the detector.
Between the calorimetry and muon system, there is the solenoid magnet which generates a powerful magnetic field of 3.8 T. The magnetic field is confined using the iron yokes implemented outside of the magnet coil in the muon system.

The Compact Muon Solenoid experiment [44] is an apparatus with 21 m length, 15 m width and 15 m height, having a total weight of about 12500 tonnes, most of which is due to the iron yokes. Comparing with the other general-purpose detector of the LHC, ATLAS with 46 m length, 25 m width and 25 m height, CMS is much more compact. The CMS apparatus is the only detector of the LHC experiment which was constructed on the surface in fifteen sections, before being lowered underground into the cavern for assembly and installation. A schematic view of the Compact Muon Solenoid can be found in Figure 2.6.

The four-momenta of the particles which emerge in the proton-proton interactions are described using a spatial coordinate which is centered on the interaction point. The y-axis points vertically upward while the x-axis is chosen to point to the center of the LHC. The z-axis is then selected along the beam direction in such a way which yields to obtain a righthanded coordinate. Based on this choice of cartesian coordinates,

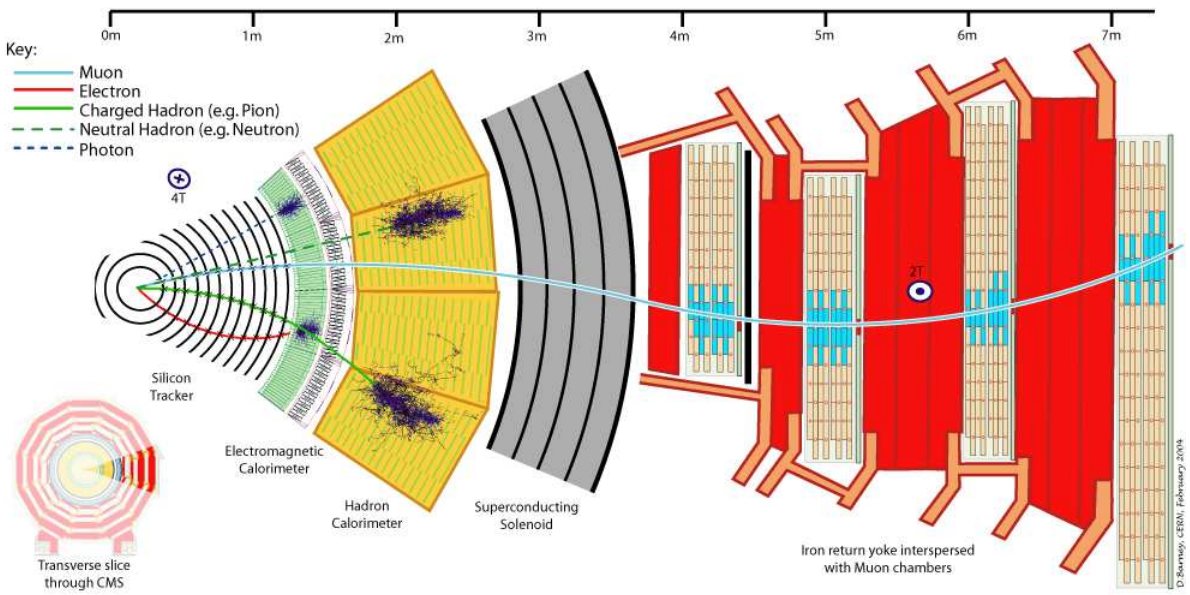

Figure 2.5: A slice of the transverse view of the CMS detector showing several particle trajectories while traversing the detector material. The tracker system, electromagnetic and hadron calorimeters together with the muon system are constituents of the CMS detector. The location of the superconducting magnet is also shown.

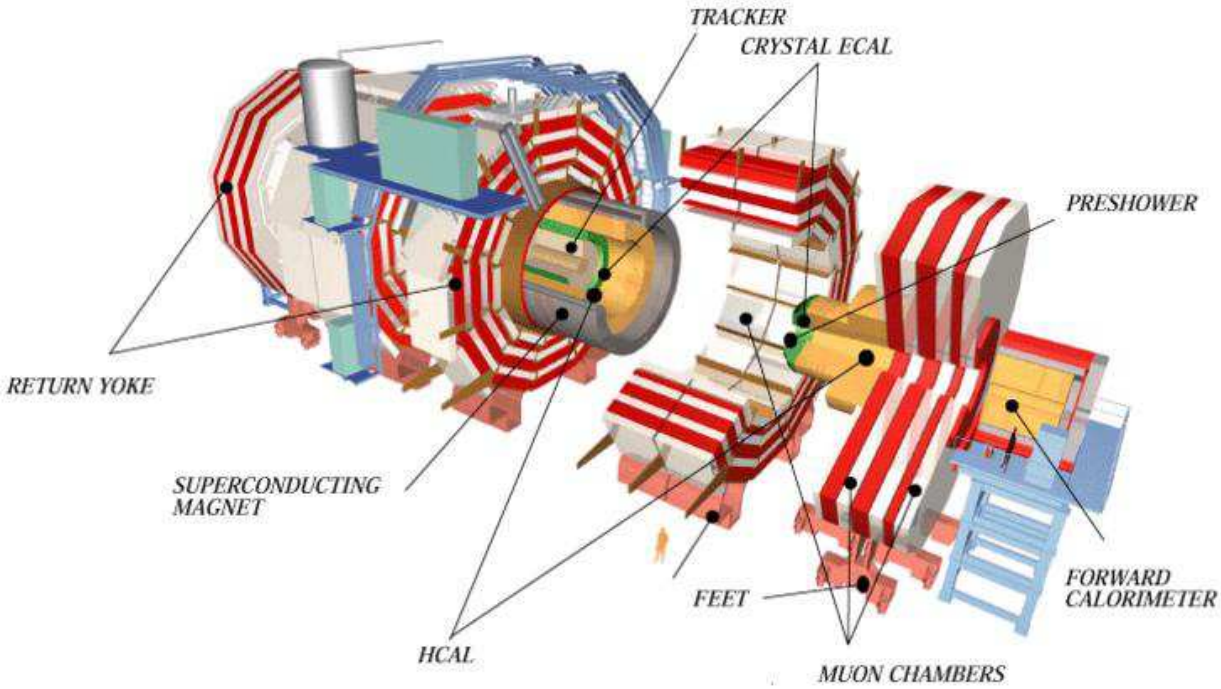

Figure 2.6: The overal size of the CMS detector, exhibiting the structure of the barrel and endcaps, is compared with the typical size of a human. Also shown are the various sub-detectors of the CMS experiment.

the spherical coordinates can also be introduced which are commonly used at hadron colliders. In addition to the polar angle θ and azimuthal angle ϕ, a new variable is of high importance at hadron colliders, namely the pseudorapidity η defined as

$$\eta \equiv \frac{1}{2} ln\Big(\frac{E + p_z}{E - p_z}\Big) = -ln \tan\Big(\frac{\theta}{2}\Big),$$

where the equation is true in the massless limit of the particle with an energy of E and a z component of the momentum vector of p_z. Since η is additive under longitudinal boosts along the beam axis, the pseudorapidity difference is invariant in the two frames. Therefore, in order to describe the angular coordinates of the particle in hadron collider physics, it is preferred to use (η, ϕ) parameters instead of (θ, ϕ) variables.
In the following sections the sub-detectors of the CMS experiment are briefly explained.

2.2.1 Tracker System

The tracker system [45, 46] of the CMS detector is made up of silicon based detectors to provide a high resolution detection required by the huge flux of charged particles produced in the high luminosity regime of the machine. The overall structure of the tracking system comprises a central cylindrical part, called tracker barrel, and two separate discs placed at either side of the barrel, called tracker endcaps. In total 13 layers in the central region and 14 layers in the endcaps form the CMS silicon tracker. They overall exhibit a diameter of 2.4 m and a length of 5.4 m resulting in a tracking coverage of $|\eta| \leq 2.5$. Closest to the interaction point is a silicon pixel detector, comprises the innermost 3 layers in the barrel and 2 layers in the endcaps. A total of 66 million pixels with a size of $100 \times 150\ \mu m^2$ cover a distance of 4 to 11 cm from the beam line on a cylindrical barrel and endcaps structure with total length of 92 cm. Besides the silicon pixel detectors, silicon strip detectors contribute in the tracker system with a variable size from $10\ cm \times 150\ \mu m$ to $25\ cm \times 150\ \mu m$. There are 9.6 million strip channels in total. A longitudinal view of the tracker system can be found in Figure 2.7, where only the silicon strip layers surrounding the pixel detector are shown.

The transverse momentum of a charged particle with an energy of about 100 GeV produced in the barrel part of the detector, is measured by the tracker system with a precision of about (1-2)%.

2.2.2 Calorimetry System

ECAL

The electromagnetic calorimeter (ECAL) [47, 48] is aimed to measure the energy of electrons and photons. Similar to the tracker system, the ECAL is also charactrized by the electromagnetic barrel (EB) and the electromagnetic endcap (EE) structure. Lead tungstate $PbWO_4$ crystals are used to construct the ECAL due to its high density and radiation hardness. It has a small radiation length of $X_0 = 0.89\ cm$ and a Moliere radius of $R_M = 2.2\ cm$. Each barrel crystal covers an area of $\Delta\eta \times \Delta\phi = 0.0174 \times 0.0174$

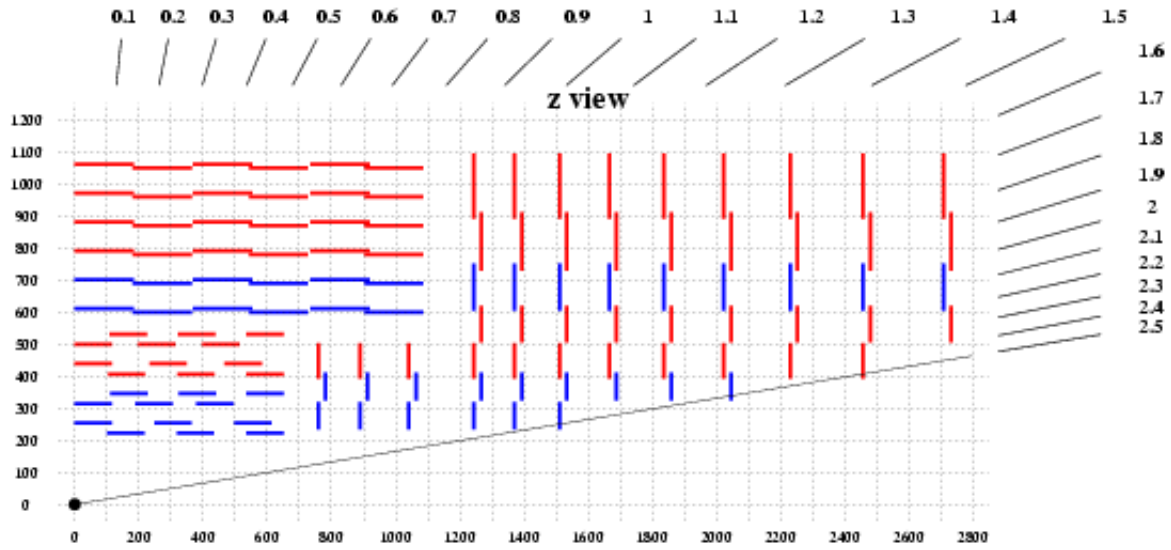

Figure 2.7: The silicon strip layers used in the tracker barrel (horizontal lines) and the tracker endcaps (vertical lines) of the CMS detector is shown. The distances from the interaction point are quoted in units of mm. Also shown are the various pseudorapidity regions covered by the tracker system.

with a length of 23 cm which corresponds to 25.8 X_0. The arrangement of the barrel crystals is quasi projective, which means they point almost to the interaction point to minimize the cracks. Each endcap crystal is 24.7 X_0 long. The angular coverage of the EB and EE is defined to be $|\eta| < 1.479$ and $1.479 < |\eta| < 3.0$, respectively. Also each endcap is partially covered by a preshower detector (ES) which extends the range of $1.653 < |\eta| < 2.6$. The ES is aimed to clearly distinguish π^0 from photons by identifying two closely spaced photons from π^0 decays. Since $pp \rightarrow H \rightarrow \gamma\gamma$ is one of the important decay channels of the Standard Model Higgs boson, π^0 decaying to two photons could mimic the final products of a hard scattering including a Higgs boson as the intermediate particle. The high granularity of the preshower would then improve the search for the Higgs particle.

The ECAL is calibrated using test beam electrons [49], resulting in a parametrized energy resolution which is expressed as [50]

$$\frac{\sigma(E)}{E} = \frac{2.8\%}{\sqrt{E/GeV}} \oplus 0.3\%.$$

Therefore, an electron with an energy of 100 GeV can be measured by the ECAL with a precision of about 0.4%.

HCAL

The hadronic calorimeter (HCAL) [51] is designed to measure the energy of hadrons. The hadronic barrel (HB) is located between the ECAL and the solenoid coil and is complemented by the outer part (HO) placed outside the magnet. The HO is designed to absorb the remnant of the hadronic shower which has not been fully absorbed in the HB. The HB covering the pseudorapidity range up to $|\eta| = 1.3$, provides a granularity of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. The hadronic endcap (HE) covers the region $1.3 < |\eta| < 3.0$ with the same granularity as HB up to $|\eta| = 1.6$ and a decreasing granularity to $\Delta\eta \times \Delta\phi = 0.17 \times 0.17$ for the rest of HE coverage. The forward hadronic calorimeter (HF) provides an extended coverage of the pseudorapidity from $|\eta| = 3.0$ up to $|\eta| = 5.0$. The HCAL sub-detector is a sampling calorimeter where layers of absorber, result in producing hadron showers, interleaving with layers of active material, which measure the energy of the showered particles. In the HB and HE, brass plates are used as absorber material while the plastic scintillators constitute the active material. The HF is constructed from steel as absorber and as active material, quartz fibers are used.

The initial calibration of the HCAL is performed using test beam data and a generic parametrization as of the ECAL can be obtained for the HCAL energy resolution. The combined energy resolution of the HCAL and the ECAL in the barrel region of the detector is parametrized as [52]

$$\frac{\sigma(E)}{E} = \frac{84.7\%}{\sqrt{E/GeV}} \oplus 7.4\%.$$

Therefore, a reconstructed jet with an energy of 100 GeV can be measured by the calorimeter system with a precision of about 11.2%.
The final calibration of the HCAL is achieved with the use of physics events based on the kinematical properties of the objects produced in the hard interaction which are explained in more detail in the next chapter.

A quarter of the longitudinal view of the ECAL and HCAL sub-detectors can be found in Figure 2.8 where the pseudorapidity boundaries are also overlaid.

2.2.3 Muon System

The muon chambers [53], the outermost sub-detector of the CMS experiment, is aimed to identify muons and measure their momenta. Three types of detectors are used in the muon system, being drift tubes (DT), cathode strip chambers (CSC) and resistive plate chambers (RPC). While the first and second ones provide precise trajectory measurements, the last one is characterized for its fast signal processing. The DTs are used in the barrel part of the detector, while the CSCs are used in the endcaps. The layers of RPCs are installed in both the barrel and the endcaps.

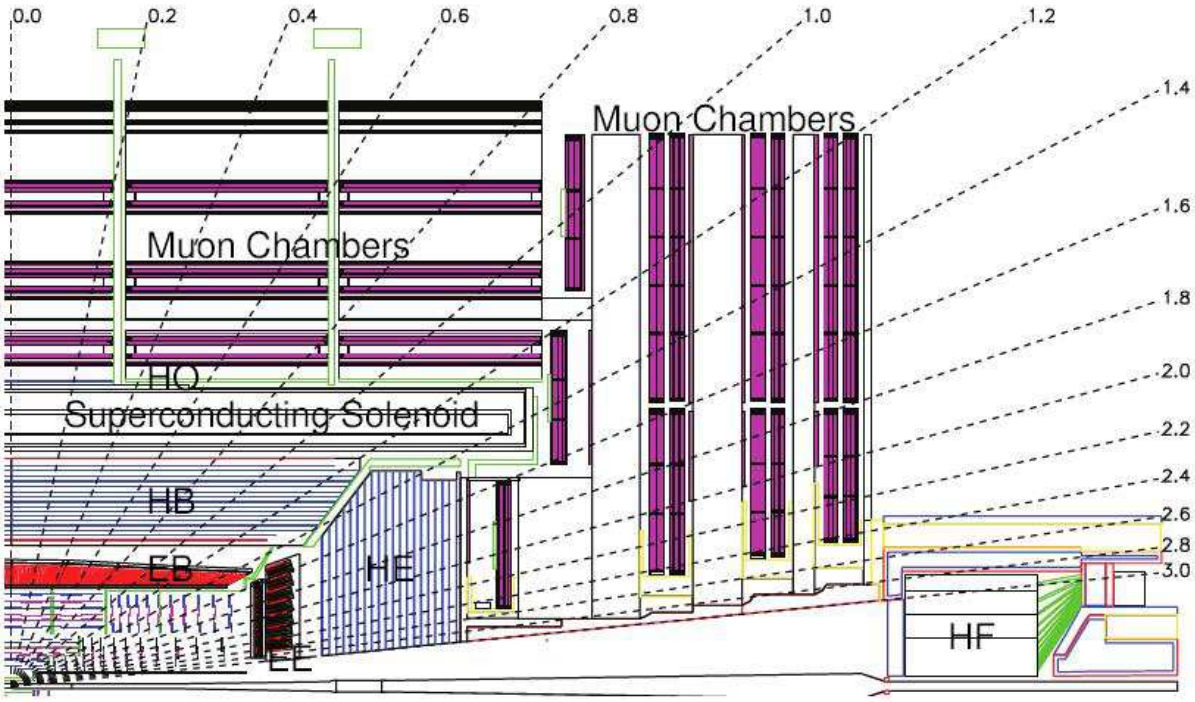

Figure 2.8: The longitudinal view of the CMS detector, where the pseudorapidity coverage of the electromagnetic barrel (EB), the electromagnetic endcap (EE), the hadronic barrel (HB), the hadronic endcap (HE), the forward hadronic calorimeter (HF) and the outer part (HO) are shown. Also shown is the muon system, the last layer of the CMS detector.

2.2.4 Triggering and Data Acquisition

In one hand, the LHC, while working at the design luminosity of $\mathcal{L} = 10^{34}$, collides bunch of protons every 25 ns resulting in a bunch crossing rate of 40 MHz as mentioned before. This corresponds to obtain 10^9 proton-proton interactions per second when the machine is operating at the center of mass energy of 14 TeV. In the other hand, the current technology is limited to process only about 100 interactions per second, hence 10^7 times less than what is produced. Since most of the occured interactions are soft collisions of protons, then one does not lose processes of physics interest by using an optimal rejection procedure. The process of online event selection is handled with the use of the trigger system [54, 55], which is performed in two stages. A first stage, which is called first level trigger (Level-1), is aimed to reduce the rate of the data taking down to 100 kHz. Because the decision made by the Level-1 trigger should be as fast as possible, the basic information from calorimeters and the muon system are used to decide whether an event has to be kept or not. A second stage, which is called High-Level Trigger (HLT), is designed for a further reduction of the rate down to 100 Hz. The interactions selected by the Level-1 trigger are then fed to the HLT for additional investigation and making a more robust decision. The HLT software contains more sophisticated algorithms which are very close to the ones applied in the offline analysis. The decision which is made at each stage of the trigger to keep or reject a specific interaction is based on the reconstruction and identification of the physics objects at the final state of that particular interaction. Because processes of physics interest, such as production and decay of rare particles like Higgs bosons, contain high energetic objects, such as electrons, photons, muons and jets, various trigger streams are defined, each of which requires some configurable physics conditions. An interaction which fires at least one of the trigger streams is archived permanently and kept for the offline analysis.
The data which pass the HLT filters are stored in computing centers distributed all around the world. According to the CMS computing model [56], the storage resources are connected and managed through the Worldwide LHC Computing Grid (WLCG) [57]. The WLCG project makes it possible for the users to access the information and data which have been distributed over the storage resources with the use of the Grid network. Computing centers are grouped in three categories referred to as Tier-0, Tier-1 and Tier-2, each of which provides some particular tasks. The raw data which are provided by the CMS detector are first moved to the Tier-0, which is located at CERN, in order to run the initial reconstruction algorithms and processing of the data. The reconstructed data are then transferred to the computer resources of the Tier-1 centers. Also the simulated proton collisions, which are produced at Tier-2 computing resources, are kept in Tier-1 storage services. In case it is needed, the re-reconstruction of the data is also performed at Tier-1 centers. The several Tiers are kept in a hierarchial level. The collected data by the CMS experiment are then stored in the form of ROOT files [58] and subsequently analyzed with some dedicated software, called CMSSW. The CMS computing model provides various data formats to facilitate the data transfer and processing. The data which come out of the detector, containing the electronic signals and hits, is called RAW data. They are huge and comprising the low level objects which are not suitable for the physics analaysis. After

running reconstruction algorithms, the RAW data format become RECO data format with a reduced size which contains the reconstructed physics objects, hence the high-level objects relevant for the analysis. Moreover, a sub-set of the RECO data, which is the Analysis Object Data (AOD) represents the third data format layer and is designed to improve the data transfer rate among the computing resources with a limited size of data events. In addition, a fourth data format is introduced by the Physics Analysis Toolkit (PAT) [59] which is aimed to facilitate the physics analysis by including the advanced properties to the reconstructed objects. Furthermore, it is still possible to improve the speed of the analysis by the use of Ntuples [58], which are kind of Tree objects officially produced in various collaborating institutes resulting to enhance the analysis speed. This thesis is based on the processing of the objects provided officially by the CMS Brussels team [60], called TopTrees. The version of the software which has been used to produce the PAT objects is `CMSSW_3_6_1_patch3` followed by the `CMSSW_36X_v1_patch2` version of the software used for the production package of the TopTrees.

Chapter 3

Object Reconstruction

The elementary particles, namely quarks and leptons, emerging from a hard interaction are observed in a detector experiment when they interact with the material of the detector, the neutrinos being an exception because they do not interact with the matter and their presence is understood by the missing amount in a balance of energy between initial and final state. According to the known physics processes [61], particles interact with matter and release their energy which subsequently is collected by the data acquisition (DAQ) system of the detector. The information that comes out of the detector includ electronic signals which are used when reconstructing the physics objects. Different objects have a distinct behaviour while traversing the detector. The detector has been designed in such a way that, for example, the electron objects are stopped in the ECAL and loose almost all their energy before reaching the HCAL. Since electrons are charged particles, they leave hits in the tracker. Therefore an electron is reconstructed by assigning a track in the tracker to a collection of energy deposit in the ECAL. A more precise description of the electron reconstruction algorithm used in the CMS collabration, is given in Section 3.1.

In case of quarks, as they are not observables before the hadronization process, they pass through the detector as a cascade of hadrons, called jets. Jet reconstruction is not as simple as reconstructing an electron and much more sophisticated algorithms are needed. A detailed explanation on how to reconstruct the jet objects is the subject of Section 3.2. In addition to the electrons and the jets, the reconstruction of the neutrinos is of great importance and is described in Section 3.3. Muons can also be reconstructed in the CMS experiment [62]. As muons are not processed in this analysis, no dedicated part is added to explain their reconstruction.

3.1 Electron Reconstruction

In order to reconstruct an electron in CMS, the information from both the tracker and the electromagnetic calorimeter is used. The first step is the clustering of the energy deposits in the ECAL. It has been shown that most of the energy of a single electron or an unconverted photon reaching the ECAL, can be collected in a cluster which is made of a few ECAL crystals. For instance, electrons with an energy of 120 GeV impinging at the center of an ECAL crystal, deposit about 97% of their energy

in a 5×5 crystal window around it [63]. Nominally electrons radiate photons while traversing the tracker material. About 35% of the electrons radiate more than 70% of their initial energy before reaching the ECAL. The radiated photons reaching the ECAL spread their energy in the ϕ direction and this makes the situation more complicated to cluster the energy of such electrons. The idea is to collect the electron energy by making a cluster of clusters, the so called supercluster, along a ϕ road which contains the bremsstrahlung photons emitted along the electron trajectory in the tracker.

3.1.1 Electron Clustering Algorithms

The basic Hybrid and Island clustering algorithms can be used for electrons in the ECAL barrel and endcap respectively [64]. The main step for both of these algorithms is to make clusters. A cluster is created by looking for a bump of deposited energy around a local energy maximum called the seed crystal. For the barrel electrons, the energy is collected in dominoes made from 3 or 5 crystals in η. Dominoes centered around the seed crystal make a cluster. For the endcap electrons, instead of making dominoes, the search for energy deposited around the seed crystal is started cell by cell. The energy of each crystal is added to the cluster until a rise in the energy is observed. The final step in the clustering algorithms is simply to collect the nearby clusters in the ϕ direction in order to recover the radiation emitted along the trajectory of the electron.

3.1.2 Electron Track Reconstruction

Having clustered the energy of an electron, the next step is to reconstruct and associate a track to it. The starting point for reconstructing a track in the tracker is to find two hits in the pixel detector. The two pixel hits that are found serve as a seed to build an electron track in the silicon tracker detector.
In order to reduce the number of possible hit combinations that can form a candidate seed for reconstructing the electron track, the search for seeds is restricted to a region which is compatible with a supercluster in the ECAL. This approach of supercluster-driven pixel seed finding has the advantage of increasing the purity of the sample of candidate electron tracks. Hits in the pixel layers are predicted by backward propagation of the position of the supercluster through the magnetic field towards the pixel detector. A first compatible hit is then looked for in the innermost pixel layer. The predicted trajectory is then propagated to look for a second pixel hit in the next pixel layer. When two hits are found, a seed is generated.
Starting from the seed and using a Bethe Heitler modeling of the electron energy losses, a trajectory is created and compatible hits on the next silicon layers are searched for. The procedure is stopped if no hit is found in two subsequent layers. The compatibility among the predicted trajectory and the measured hits is checked using a fit method called the Gaussian Sum Filter (GSF) and is defined in terms of a χ^2 value. Finally the best track candidate with the smallest χ^2 and having a minimum of five hits is kept and associated to the electron.
Besides to the supercluster-driven pixel seed finding method, there is a tracker-driven

seeding algorithm which has been developed for non-isolated electrons [65]. In a tracker-driven seeding procedure, the reconstruction of the track of an electron starts from the tracker and then a supercluster in the ECAL is matched to the reconstructed track. The performance plot which compares the efficiency to reconstruct the track of the electron is shown in Figure 3.1.

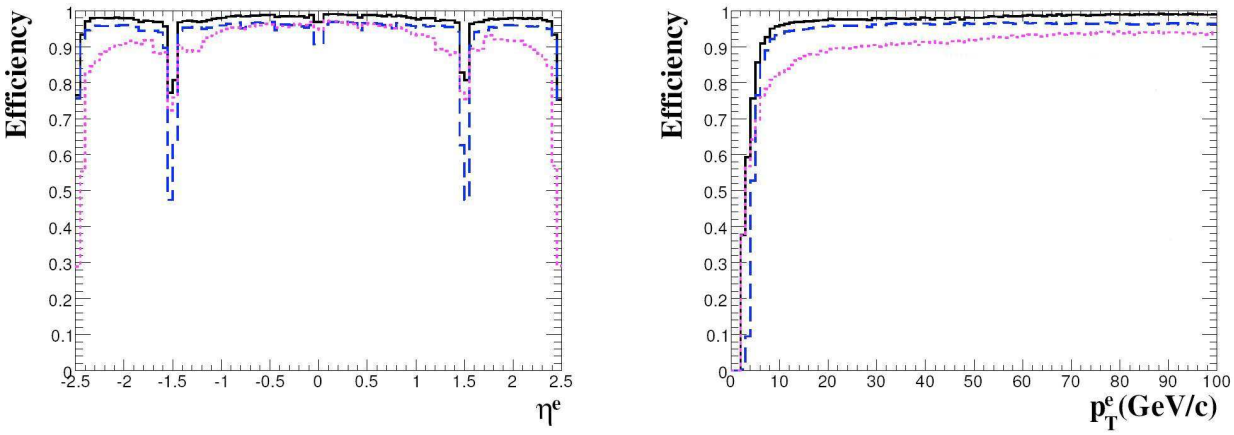

Figure 3.1: Electron seeding efficiency (solid line) as a function of the generated electron η (left) and generated electron p_T (right) for electrons in a simulated sample of $Z \rightarrow e^+e^-$ events [66]. The individual contributions from the supercluster-driven (dashed line) and from the tracker-driven (dotted line) seeding algorithms are also shown.

As it is seen from Figure 3.1, the electron seeding efficiency obtained from the tracker-driven approach is higher for the low-p_T electrons, as aimed for. Also two drops observed in the η distribution of the electron seeding efficiency refer to the transition region between the ECAL barrel and the ECAL endcaps. The lower efficiency obtained from the supercluster-driven approach compared to the efficiency derived from the tracker-driven approach is mainly due to an inefficiency in the reconstruction of superclusters in the transition region.

3.1.3 Electron Momentum Determination

The final electron momentum is obtained by combining the energy measurement provided by the electromagnetic calorimeter with the momentum measurement at the vertex provided by the tracker. If the two measurements are comparable, then a combination of both energy and momentum measurements is used as an estimate of the final electron momentum. Otherwise, if they disagree significantly, the energy measurement from the ECAL sub-detector is taken as the final estimation given that it exceeds 15 GeV. In case that the ECAL measures an energy below 15 GeV, then the momentum measured by the tracker is used and assigned to the electron. The idea behind this is that in particular at energies of around 15 GeV and below, the momentum estimation from the tracker is more precise than the energy measurement provided by

the ECAL.

In order to evaluate the performance of the electron reconstruction algorithm in CMS, one can compare the various properties of reconstructed electrons with the generated electrons. Therefore a matching in (η, ϕ) space is performed to assign a reconstructed electron to a generated one. The efficiency of finding a reconstructed electron close to a generated electron as a function of the p_T of the generated electron, is shown in Figure 3.2. The plot, made for electrons in a simulated $Z \rightarrow e^+e^-$ sample after applying some quality criteria, has been extracted from [66].

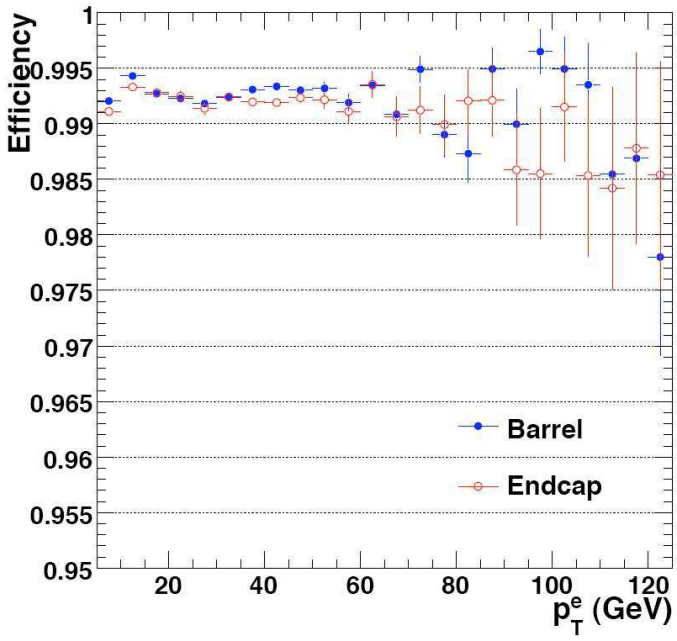

Figure 3.2: Electron reconstruction efficiency as a function of the p_T of the generated electron [66]. Electrons are from a simulated sample of $Z \rightarrow e^+e^-$ decays.

It is obvious from Figure 3.2 that the efficiency of reconstrucing electrons in the CMS detector is larger than 99%, which means in almost all the time an electron can be reconstructed very close to the generated electron.

3.2 Jet Reconstruction

Quarks and gluons appear in the final states of many processes of interest in the proton collisions observed by the CMS experiment. They are carrying color charge, hence due to the quark confinement principal they cannot exist in a free form. According to the theory, partons get hadronized to make some colorless hadrons which then are able to be detected. Thus the signature of a quark or gluon is the presence of many particles moving in the direction of the original parton. With a good approximation, the hadrons

which arise in the hadronisation process can be collected in a cone around the direction of the parton that characterizes a jet.

Different types of jets are reconstructed at CMS depending on the information of which sub-detectors is used [67, 68]. As only calorimeter jets are processed in this analysis, there is a dedicated section to describe the details of this kind of jets, followed by descriptions of the diverse jet clustering algorithms currently available at CMS.

Calorimeter jets, hereafter called CaloJets, only use the information of the CMS electromagnetic (ECAL) as well as the hadronic (HCAL) calorimeters. The building blocks for CaloJets are calotowers which are made of the HCAL towers and the corresponding ECAL crystals. The energy associated to a calotower is simply obtained by collecting the energy deposits in those components which pass the so-called "Scheme B" energy thresholds [69]. This helps to suppress the electronic noise and not to count them in the energy of the calotower. In order to further reduce the additional contribution from pile-up collisions, calotowers are required to have a transverse energy exceeding 0.5 GeV to be used for jet clustering algorithms.

3.2.1 Jet Clustering Algorithms

Several methods to define a jet exist [70–72]. Two of them with great importance in the CMS colaboration are described below. It will become clear that a first method, the iterative cone algorithm, is used in the online triggering program while a second one, the anti-k_T algorithm, is important as it is highly recommended for the offline analysis.

Iterative Cone Algorithm

This simple algorithm merges the energy deposits located in a cone around the most energetic calotowers. The steps taken to create the iterative cone jets are listed below.

- an E_T-ordered list of input objects is created;
- a cone of size R in (η, ϕ) space is cast around the seed, the object with the largest E_T;
- objects inside the cone are used to calculate the properties of a protojet

$$E_T = \Sigma_i E_{Ti}, \quad \eta = \frac{\Sigma_i(\eta_i \times E_{Ti})}{E_T}, \quad \phi = \frac{\Sigma_i(\phi_i \times E_{Ti})}{E_T};$$

- this protojet is used to seed a new protojet;
- the procedure is repeated until the energy and direction of the protojet does not change between iterations;

- the stable protojet becomes a jet and the corresponding objects which made the protojet, are removed from the list to look for other protojets.

This algorithm has a fast response and is therefore usually used in the triggering software of the online trigger system. R is a free parameter of the algorithm.

Anti-k_T Algorithm

The detailed description of making anti-k_T jets is provided here.

- starting from an input list of objects, the algorithm looks for the smallest distance among all d_i and d_{ij} whose definitions are

$$d_i \equiv \frac{1}{{E_T}_i^2},$$

$$d_{ij} \equiv \min\{\frac{1}{{E_T}_i^2}, \frac{1}{{E_T}_j^2}\} \times (\frac{R_{ij}}{R})^2,$$

and $R_{ij} = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}$;

- if the smallest is a d_i, particle i is removed from the list of particles and called a jet;

- if the smallest is a d_{ij}, particles i and j are recombined to form one single new particle according to

$$E_T = \Sigma_i E_{Ti}, \quad \eta = \frac{\Sigma_i(\eta_i \times E_{Ti})}{E_T}, \quad \phi = \frac{\Sigma_i(\phi_i \times E_{Ti})}{E_T};$$

- the procedure is repeated until no particles are left.

Also here, R is a free parameter of the algorithm.

This algorithm is currently highly recommended within the CMS experiment. The reason is generally that, with respect to the iterative cone, the anti-k_T algorithm is well defined in the sense that it does not have the defects such as Infrared and Collinear unsafety which means that the algorithm is stable with respect to the soft radiation and collinear decomposition, respectively. In other words the output of the anti-k_T jet finding algorithm does not change if any minor reconfiguration in the input list of the objects, such as adding a soft radiation, occurs. However, the iterative cone jet finding algorithm has a fast response and is currently used in the software part of the online triggering system.

In this study, only anti-k_T jets are processed for which the R parameter is set to be 0.5.

3.2.2 Jet Reconstruction Performance

In order to evaluate the performance of the jet reconstruction algorithms, the created CaloJets are compared to the particle level jets, the so-called GenJets, which are independent of the response of the CMS detector. GenJets are made by applying the same jet reconstruction algorithm on the stable particles which are coming out of the hadronisation process. The GenJets are not observable jets, because they can only be accessed from the simulated information. Equivalently, the actual detected jets are CaloJets which are GenJets right after they reach the detector. The jet matching efficiency which is defined as the fraction of the GenJets that match to the CaloJets within a ΔR cone of 0.5, is shown in Figure 3.3 for various jet clustering algorithms [69].

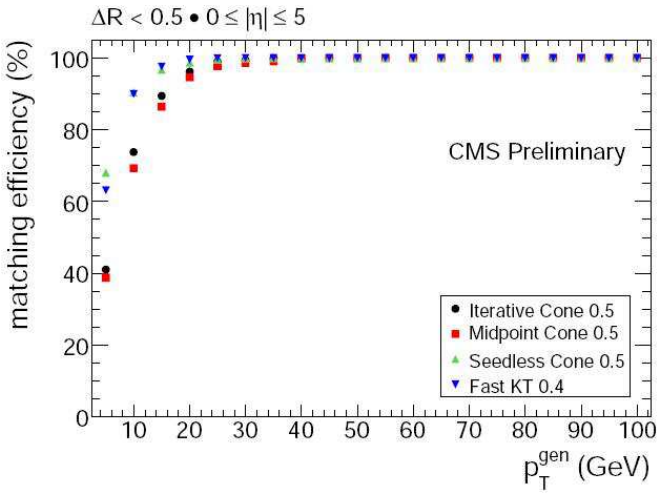

Figure 3.3: Efficiency of matching a reconstructed CaloJet to a GenJet as a function of p_T^{GenJet} for various jet clustering algorithms [69].

According to Figure 3.3, the probability of finding calorimeter jets that match to generated jets with, for example, $p_T = 30\ GeV$ is larger than 95%. The drop in the efficiency at low p_T is due to the presence of the magnetic field. Since low-p_T charged hadrons have a smaller bending radius, therefore they escape from the cone which is cast arount the jet and results in a wrong determination of the direction of the jet.

Another parameter that describes how well a jet clustering algorithm works, is the calorimeter response to jets $R(\eta, p_T^{GenJet})$ which is defined as

$$R(\eta, p_T^{GenJet}) = \frac{p_T^{CaloJet}}{p_T^{GenJet}}.$$

The plot of the jet response as a function of the p_T of the GenJet, can be found in Figure 3.4. It is seen that the response is not equal to one which means that, for example, for a CaloJet with true $p_T = 100\ GeV$, only about 60% of its energy can be collected. The effect of out-of-cone showering can explain partly why the non-unity

behaviour of the jet response function is observed. Due to that the low-p_T hadrons which are produced in the hadronization step are mostly affected by the magnetic field and have usually a small bending radius, hence they escape from the cascade of hadrons and do not contribute in the cone of the clustering algorithm and results in less collected energy compared to the energy of the initiating parton. In addition, the low-p_T hadrons interact with the nuclei of the material of the ECAL detector, even before they reach to the HCAL. This could also explain partially the low response of the detector to the calorimeter jets with low transverse energy.

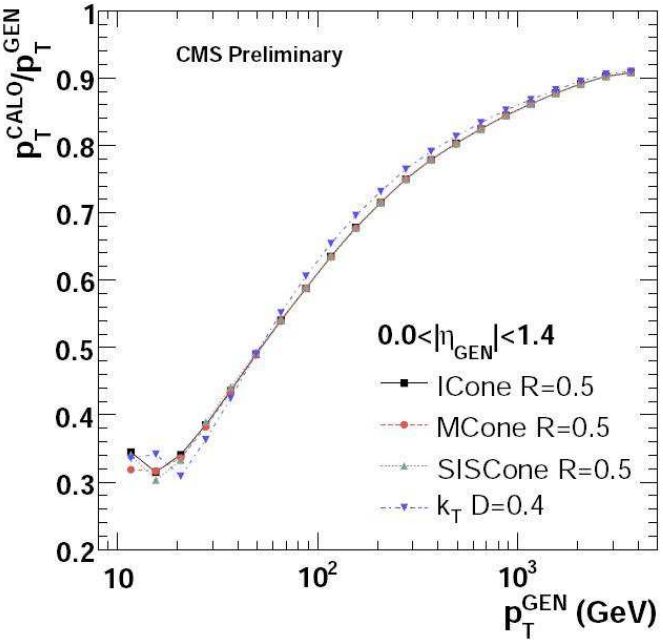

Figure 3.4: The detector response to calorimeter jets as a function of p_T^{GenJet} for various jet clustering algorithms [69].

In order to solve this issue and make the jet response equal to one, the jet calibration idea is proposed. The calibration of the jet energy, which is the subject of the next section, corrects the energy of a CaloJet back to the true energy of its matched GenJet.

3.2.3 Jet Energy Calibration

A factorized multi-level jet correction has been developed in CMS [73, 74]. The first required level of correction called Level 1 offset, is to apply noise and pile-up suppression before invoking the jet algorithm with either a set of cuts on the calotower energies or to apply a pile-up subtraction algorithm. The remaining required correction levels are twofold. A Level 2 (L2) relative η correction which corrects for variations with η relative to a control region and another Level 3 (L3) absolute p_T correction which corrects on average the CaloJet p_T back to GenJet p_T. Both L2 and L3 corrections are described below.

L3 Absolute p_T Correction

- Method with Simulation
  CaloJets from simulated QCD dijet events are considered in order to derive the simulation based calibration factors to be used to scale the CaloJet Lorentz vectors. A CaloJet is required to be located in the control region defined to be the barrel part of the HCAL with $|\eta| < 1.3$ and also it is required to match with a GenJet in $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.25$. For the various bins of GenJet p_T and for each pair of matched jets, the quantity $\Delta p_T = p_T^{CaloJet} - p_T^{GenJet}$ is calculated and the corresponding distribution is fitted with a Gaussian. A typical Δp_T distribution calculated for GenJets with 150 $GeV < p_T^{GenJet} < 200$ GeV, is shown in Figure 3.5.

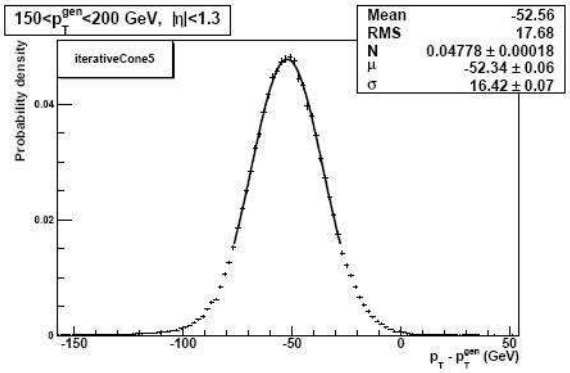

Figure 3.5: The distribution of Δp_T for jets reconstructed with the iterative cone algorithm and located in the control region for the specified p_T^{GenJet} bin [75].

It is seen from Figure 3.5 that the distribution of a generic Δp_T is peaked at negative values, as expected due to the underestimation of the p_T of the calorimeter jets. The calorimeter response to jets which is already defined in Section 3.2.2, can be obtained from the distribution of the Δp_T variable by

$$R(\eta, p_T^{GenJet}) = 1 + \frac{< \Delta p_T >}{p_T^{GenJet}},$$

where $< \Delta p_T >$ is the mean value of the fit result in each bin of the p_T of the GenJet. Finally, the absolute jet response is inverted numerically to obtain the absolute jet correction $C^{abs.}(\eta, p_T^{CaloJet})$, hence

$$C^{abs.}(\eta, p_T^{CaloJet}) = \frac{1}{R(\eta, p_T^{GenJet})}.$$

It should be noted that the absolute jet correction is expressed in terms of $p_T^{CaloJet}$. Both the absolute jet response and the absolute jet calibration functions are shown in Figure 3.6.

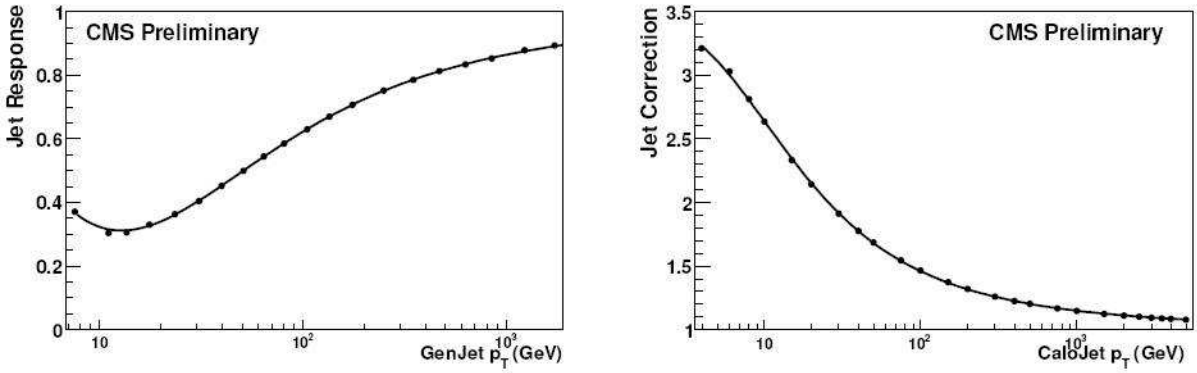

Figure 3.6: The jet response as a function of p_T^{GenJet} (left) and L3 correction factors as a function of $p_T^{CaloJet}$ (right) derived from the simulated information [73].

According to the absolute calibration curve shown in Figure 3.6, a calorimeter jet with a measured transverse momentum of about 100 GeV should be multiplied by a factor of 1.5 to represent the true transverse momentum, resulting in a jet with $p_T = 150\ GeV$. Also it can be understood that the high-p_T jets receive smaller calibration factors compared to the low-p_T jets which require larger correction factors.

- Data Driven Method
  In a data driven method, γ+jet or Z+jet events selected in the collision data, can be used to extract the jet correction factors. Assuming the conservation of the energy in the transverse plane, a p_T balance technique can be used to compare the reconstructed CaloJet p_T with the transverse momentum of the produced photon or Z boson. Since photon reconstruction only relies on the information of the ECAL which has a better precision on the energy resolution compared to the HCAL, the jet p_T measurement of the HCAL is corrected for the photon p_T measurement of the ECAL. The γ+jet events suffer from huge QCD backgrounds which can be suppressed by asking for isolated photons.
  Complementary a collection of Z+jet events of which the Z boson decays to a pair of muons, provides a clean signature of the reconstructed Z boson and is suitable to be used to derive the jet energy calibration factors. In this case, the HCAL sub-detector is compared to the muon chamber of the CMS experiment. The calibration curve for the jet energy which has been derived from simulation is compared in Figure 3.7 with one of the data driven methods, namely the Z+jet balance. A good agreement between the two procedures for obtaining the jet calibration factors can be observed.

L2 Relative η Correction

- Method with Simulation
  The purpose of the L2 correction is to correct jets at arbitrary η relative to a control region where the absolute calibration is easier. As a consequence of the

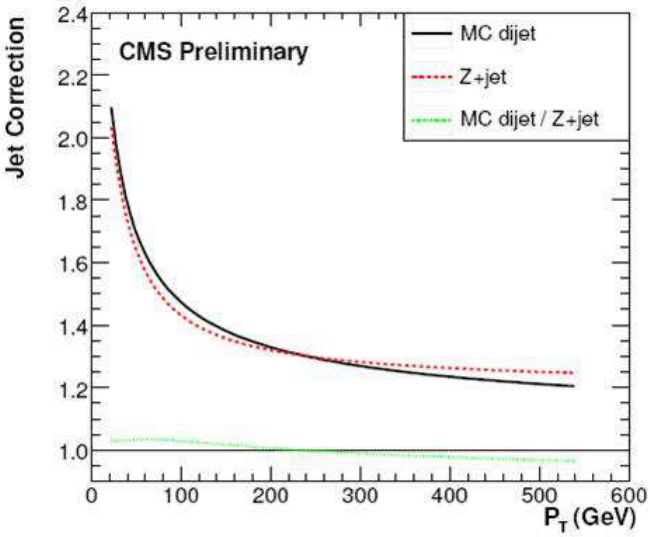

Figure 3.7: Comparison between the L3 calibration curve derived from the simulation and the Z+jet balance method applied on collision data, and the ratio of the two corrections [73].

L2 correction, the jet response becomes flat as a function of the pseudorapidity. In order to derive the L2 correction factors from the simulation, it is essential to calculate the η dependent L3 absolute correction, $C^{abs.}(\eta, p_T^{CaloJet})$, as introduced above. This is achieved by repeating exactly the same procedure described in the "L3 Absolute p_T Correction" part for multiple η regions.
The relative correction is defined as

$$C^{rel.}(\eta, p_T^{CaloJet}) = \frac{p_T^{control}}{p_T^{CaloJet}},$$

where $p_T^{control}$ and $p_T^{CaloJet}$ are the measured transverse momenta of a GenJet of the same p_T^{GenJet} in the control region and in an arbitrary η region, respectively. Starting from a reconstructed CaloJet with arbitrary η and with $p_T^{CaloJet}$, the absolute correction obtained for that specific η region can be used to find the true value of the CaloJet p_T, being p_T^{GenJet}

$$p_T^{GenJet} = p_T^{CaloJet} \times C^{abs.}(\eta, p_T^{CaloJet}).$$

Then with the use of the response of the detector in the control region, one can find the value of the reconstructed CaloJet p_T in the control region for the same p_T^{GenJet}

$$p_T^{control} = R(control, p_T^{GenJet}) \times p_T^{GenJet}.$$

Having determined the $p_T^{control}$, it is possible to calculate the relative jet correction as

$$C^{rel}(\eta, p_T^{CaloJet}) = \frac{p_T^{control}}{p_T^{CaloJet}} = \frac{R(control, p_T^{GenJet}) \times p_T^{GenJet}}{p_T^{CaloJet}}$$

$$= \frac{R(control, p_T^{GenJet})}{R(\eta, p_T^{GenJet})} = C^{abs.}(\eta, p_T^{CaloJet}) \times R(control, C^{abs.}(\eta, p_T^{CaloJet}) \times p_T^{CaloJet}).$$

Figure 3.8 shows the jet response as a function of pseudorapidity for CaloJets before and after applying L2 corrections. The response of the detector to the jets reconstructed in various parts of the detector becomes flat after applying L2 relative η correction. Also the same figure contains the corrected jet response considering both L2 and L3 correction steps.

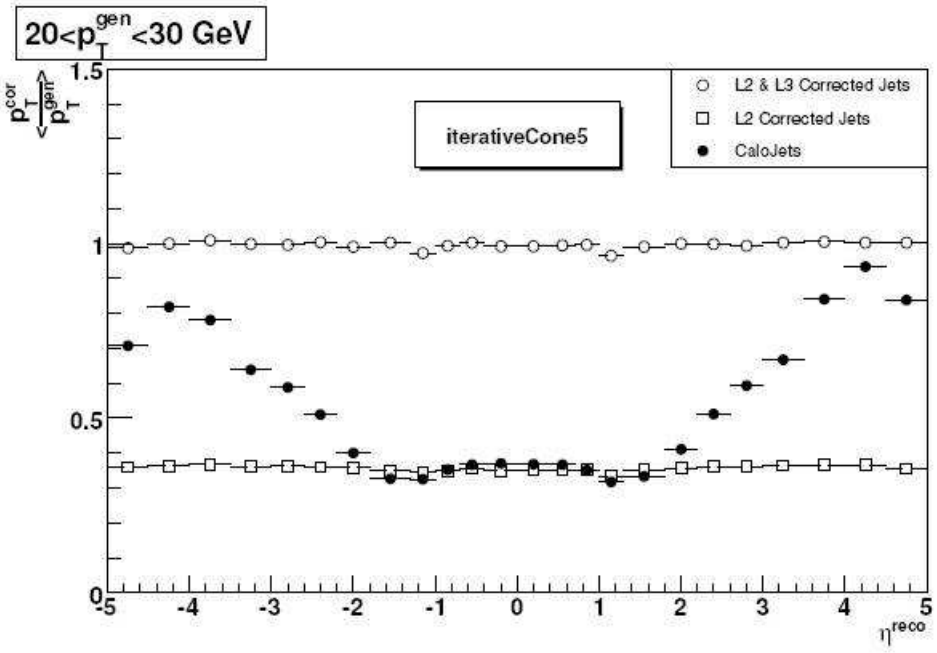

Figure 3.8: The detector response to CaloJets as a function of η without correction (filled circles), with L2 correction (open squares) and with the combined L2 & L3 corrections (open circles) [74].

- Data Driven Method
  QCD dijet events for which one of the reconstructed CaloJets is located outside the control region, defined to be the barrel part of the HCAL with $|\eta| < 1.3$, are used. In the limit when no other jet is present in the dijet topology, the p_T of the two leading jets should be the same. Any difference between the measurements of the p_T of the two jets is translated into a calibration factor which is needed to be applied on the CaloJet p_T reconstructed outside the barrel. This makes the calorimeter response to jets flat as a function of η and equal to the value obtained for the jets which are reconstructed in the control region.
  The resulting relative response function obtained from dijet balancing is compared with the one derived from the simulation study, which is shown in Figure 3.9. The agreement between these two confirms the validity of the proposed

data driven method.

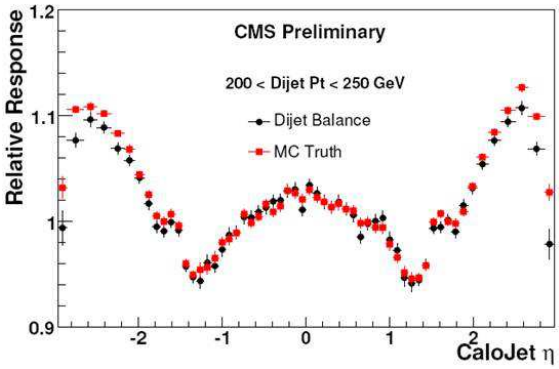

Figure 3.9: The relative jet response as a function of η from both the dijet balance method and the simulation for $200\ GeV < p_T^{dijet} < 300\ GeV$ [73].

3.2.4 Jet Energy Resolution

Now that jets are corrected, it is the time to check how successful the correction steps L2 and L3 are. Jet energy resolution is the uncertainty on the measured jet energy after applying the required corrections. Different methods for extracting the resolutions on the energy of the jets exist and are discussed below [76].

- Method with Simulation
  So-called Monte Carlo resolutions are obtained from simulated data by plotting the distributions of the ratio between the corrected CaloJet p_T and the corresponding GenJet p_T as a function of the p_T of the GenJet. The CaloJets are spatially matched to the GenJets with $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.25$. Each distribution is subsequently fitted with a Gaussian and the width of the fit is used as the resolution for that particular p_T bin.
  Figure 3.10 shows a generic distribution of the jet response after applying L2 and L3 corrections, overimposed with a fitted Gaussian function.

- Data Driven Method
  The data driven method for calculating the jet p_T resolutions is based on QCD dijet events. The parameter A, which is called "Asymmetry", is used to derive the jet resolutions and defined to be

$$A = \frac{(p_T^{jet1} - p_T^{jet2})}{(p_T^{jet1} + p_T^{jet2})},$$

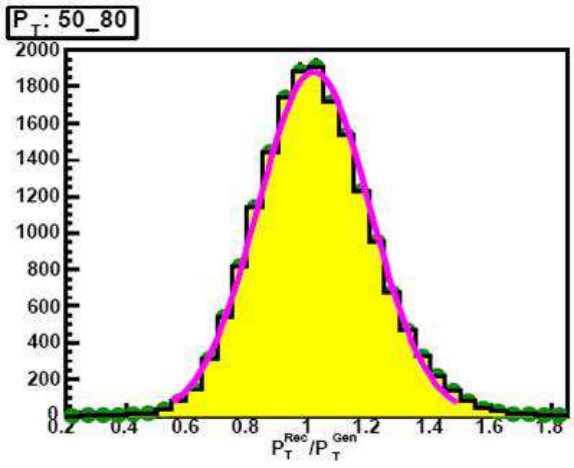

Figure 3.10: The jet response after applying L2 and L3 corrections for a particular p_T^{GenJet} bin specified in the plot. The width of the fitted function is used to extract the jet energy resolutions.

where p_T^{jet1} and p_T^{jet2} are the transverse momenta of the two leading jets in each QCD dijet event. The variance σ_A can be measured and simply written as

$$\sigma_A^2 = |\frac{\partial A}{\partial p_T^{jet1}}|^2 \sigma^2_{p_T^{jet1}} + |\frac{\partial A}{\partial p_T^{jet2}}|^2 \sigma^2_{p_T^{jet2}}.$$

Using the idea of the momentum conservation in the transverse plane of the dijet events, one can define $p_T \equiv p_T^{jet1} = p_T^{jet2}$. The two leading jets are further required to be in the same η region, so that any dependency on the resolution parameters originating from the differences in the pseudorapidity regions is removed and one can assume that $\sigma_{p_T} \equiv \sigma_{p_T}^{jet1} = \sigma_{p_T}^{jet2}$. Hence the fractional jet p_T resolution can be obtained from the width of th the Asymmetry parameter in the following way

$$(\frac{\sigma_{p_T}}{p_T}) = \sqrt{2}\sigma_A.$$

The resolutions derived from the Asymmetry method are only valid in the limit of exactly two jets in the QCD dijet events. In the case when other jets are present, the assumption of the equality of the transverse momenta of the two leading jets is no longer valid. Hence in order to eliminate the radiated jets which spoil the assumption of p_T-balancing for the two leading jets, a cut on the transverse momentum of the third leading jet is applied, $p_T^{jet3} < p_T^{cut}$. Therefore the resolution values are calculated for different threshold cuts in bins of $p_T^{avr} = \frac{(p_T^{jet1} + p_T^{jet2})}{2}$. As an example, the distribution of the Asymmetry variable for the calorimeter jets with $125 < p_T^{avr} < 160\ GeV$ and $p_T^{jet3} < 15\ GeV$, is shown in Figure 3.11.

The Asymmetry distributions for various threshold cuts on the p_T^{jet3} are fitted with a Gaussian. The jet resolutions for each bin of p_T^{avr} can then be plotted as a function

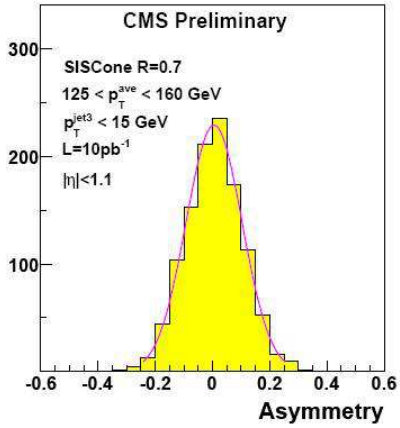

Figure 3.11: The distribution of the Asymmetry variable for calorimeter jets with $125 < p_T^{avr} < 160\ GeV$ and $|\eta| < 1.1$. The third leading jet is required to have a p_T less than 15 GeV [76].

of p_T^{cut} and fitted with a line. The true jet resolution value is obtaind by extrapolating the fitted line to $p_T^{cut} = 0$ which corresponds to a QCD event with exactly two jets. For the same p_T^{avr} as the one shown in Figure 3.11, the corresponding jet resolution is shown in Figure 3.12.

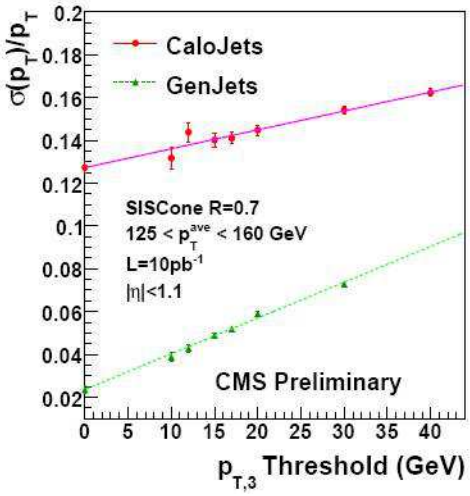

Figure 3.12: The fractional energy resolution as a function of the p_T^{jet3} threshold in the same range of p_T^{avr} and η as Figure 3.11 derived from Asymmetry method [76].

The validation of the determination of the jet energy resolutions is done by comparing the performance of both the simulation driven and data driven dijet Asymmetry methods and is shown in Figure 3.13.

There is a good agreement between the simulation based and the data driven methods to determine the relative resolution on the transverse momentum of the reconstructed jets, as can be observed from Figure 3.13. A first observation is that the

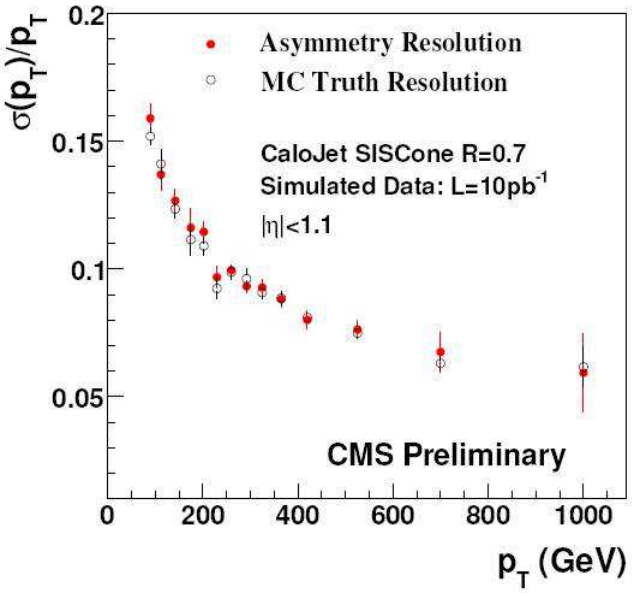

Figure 3.13: Comparisons of the jet energy resolutions from the simulation and Asymmetry method for calorimeter jets with $|\eta| < 1.1$ [76].

resolution decreases for the jets with higher transverse momenta. This means the more energetic jets are measured with a better precision compared to the less energetic jets. It can also be understood that a calorimeter jet with an energy of, for example, about 100 GeV, can be determined with an accuracy of about 15%.

3.2.5 Jet-Electron Cleaning

Since the same detector information like energy deposits in the calorimeters is used to reconstruct several physics objects, there is always a possibility to double count the physics objects. In the analyses that contain both electrons and jets in their final states, specific attention should be paid to the cleaning of the jet collection from electrons as electrons are always re-reconstructed as jets. The idea of the cleaning method is to look for electrons in the event and then remove any adjacent reconstructed jet. This is the default cleaning method that is also used in this analysis. Also there are some other studies which propose to start the cleaning even before jet reconstruction step [77]. This way the cleaning is done by looking at the more low-level objects, rather than the high-level physics objects such as jets, and removing any calotower which is reconstructed close to the electrons, followed by re-running the jet clustering algorithm over the cleaned collection of calotowers. The results between both of these proposed cleaning methods are comparable. This issue is studied in more detail in the next chapter.

3.3 MET Reconstruction

Neutrinos which appear in the final state of many physics processes, remain undetected due to their weak interactions with matter. They are the only Standard Model particles that escape detection without depositing any energy in the CMS detector. Since the initial proton beams are supposed to collide heads-on, the energy is conserved in the plane perpendicular to the beam line. The missing transverse energy, referred to as E_T^{miss}, is therefore understood as the missing amount of energy deduced from the energy balancing in the transverse plane.
The missing transverse energy is defined as the transverse part of the vectorial sum of all the energy deposits collected in calotowers

$$\vec{E}_T^{miss} = -\Sigma_n(E_n sin\theta_n cos\phi_n \vec{x} + E_n sin\theta_n sin\phi_n \vec{y}),$$

where θ and ϕ are the polar and azimuthal angles of the calotowers, respectively. By definition the uncorrected calotowers are used to calculate the E_T^{miss}. The performance of E_T^{miss} has been further improved when corrections due to leptons, such as muon, and jets are applied [78]. For example as muons are also weakly interacting particles, they do not deposit their energy in the calorimeter system. Hence for events that contain muons, the transverse momentum of the reconstructed muons should be subtracted from the above formula. Figure 3.14 shows the resolution of the missing transverse energy for inclusive $t\bar{t}$ events as a function of the reconstructed E_T^{miss}.

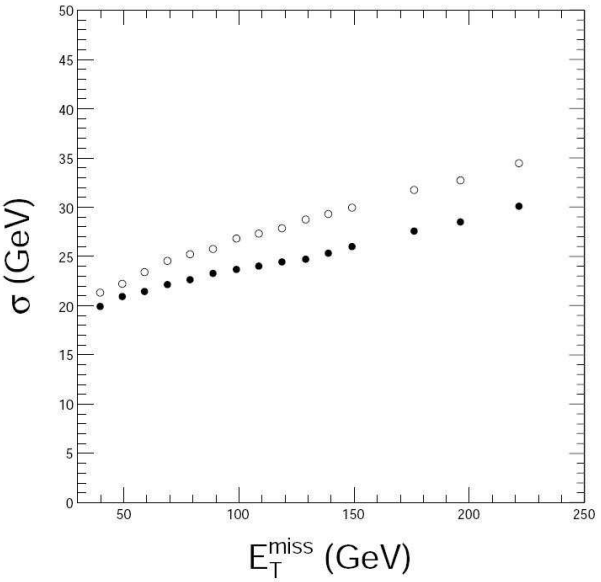

Figure 3.14: The resolution of the missing transverse energy before (open circles) and after jet corrections (filled circles) for inclusive $t\bar{t}$ events as a function of the reconstructed E_T^{miss} [44].

It is seen from Figure 3.14 that in $t\bar{t}$ events, the resolution on the missing transverse energy grows with the reconstructed E_T^{miss}. Hence in order to avoid large systematic

uncertainties that can enter into the analysis by cutting on E_T^{miss}, the reconstructed missing transverse energy is not processed in this study.

Chapter 4

Reconstructing and Selecting Physics Event

4.1 Monte Carlo Event Simulation

In the absence of experimental data and to model the available physics theories, the toy Monte Carlo data is produced. The first step in the Monte Carlo production chain is generating physics "events", i.e. sets of outgoing particles produced in the interactions between two incoming particles, followed by simulating the detector response based on the interactions of particles produced in the event with the material of the detector. Hence a very precise modeling of physics and a good understanding of the detector should result in simulated events similar to those from real collision.

The whole generation chain of an event can be factorized into several stages as illustrated in Figure 4.1.

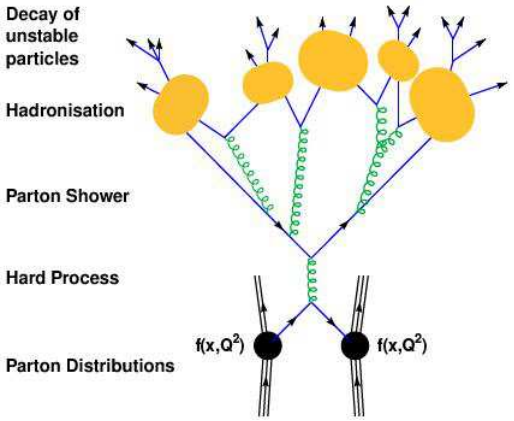

Figure 4.1: Illustration of the event generation chain in proton-proton collisions adapted from [79].

The first step which is the core of the generation procedure is to produce the hard scattering process which is a result of the interaction of two high-energy partons arising from the two colliding protons. The next step is to consider the possibility of branching of a parton to other partons or radiation of a gluon resulting in a cascade of partons referred to as parton showering. After producing a list of partons carrying color-charges, due to the quark confinement principle, they hadronize and combine in such a way to make colorless hadrons. The produced hadrons that are not stable particles, decay subsequently in the last step of the event generation procedure. The final particles interact with the detector material and can be reconstructed as described in Chapter 3.

Some of these generation steps such as the hard scattering can be described from first principles, while for others like the hadronization stage require phenomenological models to be modeled. The different generation steps are discussed in more detail in the following sections.

4.1.1 Hard Scattering Event

In a proton-proton collision, along with so many soft elastic interactions, hard inelastic scattering occurs which can include any new physics such as Higgs bosons or hints for super-symmetric phenomena. In order to describe the physics at high-energy hadron colliders, the knowledge of Quantum Chromodynamics (QCD) is needed [80]. According to the parton model, a proton is not an elementary particle and comprises quarks and gluons, generally called partons. When two protons collide heads on, actually the collision happens between their constituents. Partons within the proton are described by the parton distribution functions (pdfs) which are represented as a function of the momentum fraction carried by the parton inside the proton and also with the Q^2 energy scale at which the proton is probed. At high center of mass energies where the strong coupling constant α_s is small, perturbative calculations are feasible and the collision between the two incoming protons can be described accordingly. A hard scattering processes is schematically depicted in Figure 4.2.

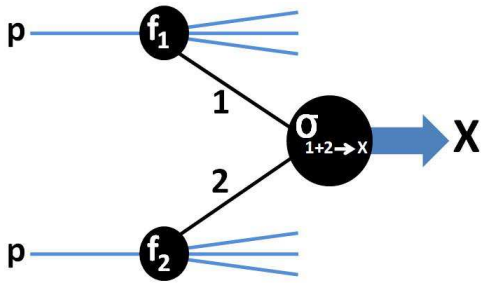

Figure 4.2: A typical hard scattering process in proton-proton collision.

For the hard scattering interactions, the factorization theorem which predicts the characteristics of the event, is used. As a result, the total hadronic cross section can

be obtained as

$$\sigma_{p+p\to X} = \int_0^1 dx_1 \int_0^1 dx_2 f_1^p(x_1, Q^2) f_2^p(x_2, Q^2) \sigma_{1+2\to X}.$$

Hence the evaluation of the total rate for a hard scattering interaction, reduces to the calculation of the sub-process cross section $\sigma_{1+2\to X}$ by identifying the leading-order partonic processes that contribute to $1 + 2 \to X$. For example, for the top quark pair production the Feynman diagrams contributing to the final state $X = t\bar{t}$ at leading order of perturbation theory are shown in Figure 4.3.

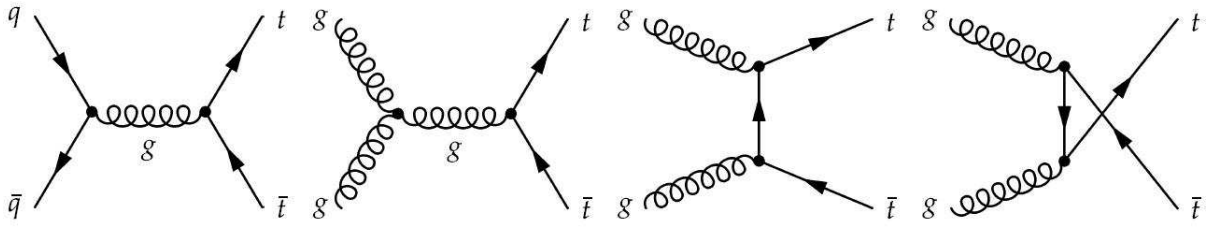

Figure 4.3: Leading order Feynman diagrams for the production of top quark pair via quark-antiquark annihilation and gluon-gluon fusion.

The distribution of the various pdfs, extracted from deep-inelastic scattering data, are provided by different collaborations such as CTEQ [81] and can be found in Figure 4.4. They are obtained at $Q = 100GeV$ which is about the order of the energy needed to produce a pair of heavy particles like top quark. It is seen that for high center of mass energies corresponding to small distances, gluons are more energetic comparing to the other quarks. Therefore it can be deduced that at the LHC the production of $t\bar{t}$ via gluon-gluon fusion is the dominant process compared to the Tevatron where quark-antiquark annihilation is the main channel for top pair production.

4.1.2 Parton-Parton Interaction and Event Generation

According to the factorization theorem described in the previous section, the calculation of the partonic sub-process cross section $\sigma_{1+2\to X}$ which is then convoluted with the appropriate pdfs is the major task in evaluating the cross section of the hard scattering event. The partonic cross section can be further factorized as follows

$$d\sigma_{1+2\to X} = \frac{1}{F} \times |\mathcal{M}|^2 \times dcos\theta d\phi,$$

where $\mathcal{M}$ is the invariant amplitude of the interaction which depends on the phase space variables such as θ or ϕ and F is proportional to the center of mass energy squared. The differential cross section $\frac{d}{dcos\theta d\phi}\sigma_{1+2\to X}(\theta, \phi)$ provides the probability density functions which are then fed into a generator to generate events accordingly. Monte Carlo techniques are used in order to generate events randomly according to their

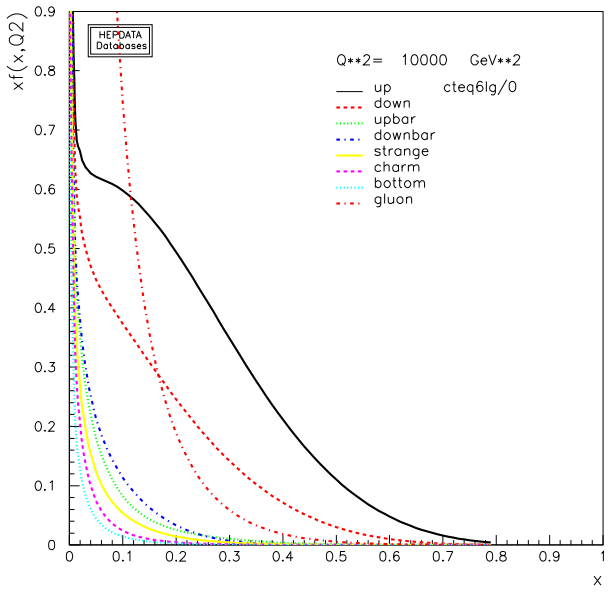

Figure 4.4: Parton distribution functions $xf(x)$ for quarks, antiquarks and gluons in the proton at $Q = 100GeV$. These distributions are obtained from a fit to deep inelastic scattering data performed by the CTEQ collaboration (CTEQ6L1) [82].

differential cross section. Matrix Element event generators are the main software tools in the procedure of generating events as they are responsible for producing the hard scattering event based on Monte Carlo methods. They provide the leading Feynman diagrams contributing to the hard process and calculate the invariant amplitude $\mathcal{M}$ for that hard interaction. Although the contributions from higher order corrections up to Next-to-Leading Order (NLO) in perturbative calculations can also be included in the Matrix Element methods, only the effects of the dominant terms in the invariant amplitude are usually taken into account. For example, NLO corrections to the hard elementary scattering amplitude is included in a Matrix Element by considering the emission of a hard gluon from the initial or final partonic legs in the Feynman diagrms while ignoring the contributions including a loop with a virtual gluon. In this sense, singularities are not present in the Matrix Element generators. `MadGraph` [83] and `ALPGEN` [84] are two examples of Matrix Element generators which are widely used in the CMS collaboration.

With a Matrix Element generator, at the end one would have events with particles at the final state which were produced based on the Feynman rules presented in the Standard Model of particle physics. The partons in either initial or final state of an event entail the probability of radiating a gluon $q \rightarrow qg$, referred to as Initial State Radiation (ISR) or Final State Radiation (FSR) respectively. The emited gluon is subsequently decomposed into either a pair of quark-antiquark $g \rightarrow q\bar{q}$ or a pair of gluons $g \rightarrow gg$. Further the radiated partons entail the possibility of additional radiations resulting in a shower of partons moving along the direction of the shower initiator parton. The splitting of a parton to other partons is controled by the DGLAP (Dokshitzer-Gribov-Lipatov-Altarelli-Parisi) equations [85–87]. The dedicated softwares which perform the

parton showering are called Parton Showers. In a Parton Shower approach, the emited gluon can be radiated collinear with respect to the original parton or be very soft which would yield mathematical singularities. These kind of singularities are handled by applying a lower cut-off on the appropriate parameters. Some of the multi-purpose event generators which besides being a physics event generator, include Parton Shower programs are `PYTHIA` [88] and `HERWIG` [89].

In order to complement the event generation chain, the output of a Matrix Element generator is interfaced with a Parton Shower program to implement the splitting effects on the produced partons of the generated event. Since Parton Shower algorithms can also produce hard partons in the final state, it is important to resolve the ambiguities that can arise in the final state configurations. For example, a given event with N+1 jets in the final state can be obtained by either a Matrix Element generator with N+1 partons in the final state or a Matrix Element generator with N partons in the final state where an additional parton is produced during the parton showering. Therefore some matching algorithms have been developed to perform the matching between jets of hadrons produced in a Parton Shower program and hard partons produced in a Matrix Element generator so that a probable double counting of events is avoided. The CKKW algorithm [90, 91] and the MLM algorithm [92, 93] are two important examples of matching schemes used in the event generation procedure.

4.1.3 Detector Simulation

The generated Monte Carlo simulated events provide physically observable particles only if they undergo a detector simulation. Equivalently a generated particle is observed when it interacts with matter. The simulation of the CMS detector is based on the GEANT4 package which simulates all possible electromagnetic and hadronic interactions of particles with matter within a magnetic field. For example, in the case of a generated electron, the possible interactions result in ionization or radiation. The emitted photons which present themselves by releasing heat and energy, are collected in the read-out. This would yield the so called experimental "hits" which are the deposited energy of the particles while traversing the detector material and constitue the building blocks for the roconstruction of the physics objects. It is worth to mention that the same reconstruction algorithms are run on either Monte Carlo simulated events or events coming from a real proton-proton collision so that one could make a fair comparison between theory and experiment.

4.2 The Event Topology

With either Monte Carlo simulated samples or experimental data, physics objects can be reconstructed. Electronic signals collected in the read-out of the detector are translated into position parameters along the particles trajectory and are used to reconstruct the physics objects and their kinematic properties as described in detail in the previous chapter. After the object reconstruction, it is the time to reconstruct the physics event of interest which is called the signal event hereafter. Any other kind of event with the same observed final state as the signal but not originating from the $pp \rightarrow t\bar{t}$ process, is

generally called background and should be removed in order not to enter the analysis and spoil the physics results.

The description of the configuration of the full signal event in this study is the subject of the next section.

4.2.1 The Signal Signature in the Detector

Top quark decays almost exclusively to a W boson and a b quark which subsequently hadronizes and produces a jet in the detector

$$t(\bar{t}) \rightarrow W^{+}(W^{-}) + b(\bar{b}).$$

Hence the decay products of a pair of top quarks only depend on the various W boson decay channels as listed in Table 4.1 [94, 95].

W decay mode	Branching Ratio
$W \rightarrow l\nu_e;\ l = e, \mu, \tau.$	$\sim 33\%$
$W \rightarrow q\bar{q};\ q\bar{q} = ud, c\bar{s}.$	$\sim 66\%$

Table 4.1: Various W decay channels and their branching ratio.

If both of the Ws coming out of the top quarks decay into leptons and their corresponding neutrinos, it is called a full leptonic decay mode of $t\bar{t}$ system. If only one of the Ws decays leptonically and the other goes to hadrons, this makes a semi leptonic decay mode of a top quark pair. Otherwise, a full hadronic $t\bar{t}$ decay mode corresponds to the situation when both of the W bosons decay hadronically. The branching ratios of the different decay modes of $t\bar{t}$ event are summarized in Table 4.2.

	$t\bar{t}$ decay mode	Branching Ratio
Full leptonic	$t\bar{t} \rightarrow W^{+}bW^{-}\bar{b} \rightarrow l^{+}l^{-}b\bar{b} + E_T^{miss}$	$\sim 11\%$
Semi leptonic	$t\bar{t} \rightarrow W^{+}bW^{-}\bar{b} \rightarrow lq\bar{q}b\bar{b} + E_T^{miss}$	$\sim 45\%$
Full Hadronic	$t\bar{t} \rightarrow W^{+}bW^{-}\bar{b} \rightarrow q\bar{q}q\bar{q}b\bar{b}$	$\sim 44\%$

Table 4.2: Various $t\bar{t}$ decay channels and their branching ratio.

In this analysis, the semi leptonic decay channel of a $t\bar{t}$ system with the lepton being an electron is considered as signal and a typical signal event is depicted in Figure 4.5.

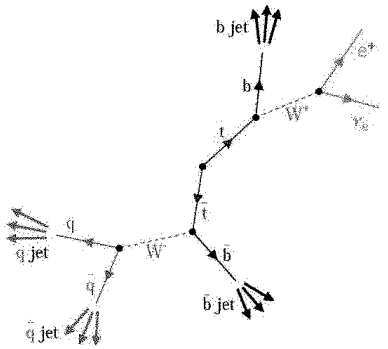

Figure 4.5: Observed topology of a generic e+jets decay of a $t\bar{t}$ system.

The final state of a e+jets $t\bar{t}$ decay, contains an isolated electron, at least four jets as a result of the hadronization of the four quarks of which two of them originate from heavy partons and missing transvers energy due to the presence of a neutrino among the decay products.

4.3 Simulated Signal and Background Monte Carlo Samples

The simulation of all signal and background events was performed using the `MadGraph` generator except for QCD samples which were produced using `PYTHIA`. The MLM matching scheme was used to perform the matching between the partons generated by `MadGraph` and the parton showering from `PYTHIA`. In addition to the production of top quark pairs via the strong interaction, single top quarks which are produced via electroweak interactions have similar signatures as the signal events and are considered as backgrounds. Also W and Z bosons in association with extra jets, where the vector boson decays leptonically, can mimic the final state of semi leptonic $t\bar{t}$ events. Due to the large cross section of the W+jets process as well as the inclusive production of a prompt electron in the final state, these events constitute the main backgrounds to the signal.

The other main background are QCD multijet events for which due to the huge cross section, it is hard to produce sufficient amount of statistics. Also because of the fact that normally no prompt electron is produced in a generic QCD event, QCD events are suppressed and one would encounter a lack of QCD statistics in the analysis. Therefore in order to obtain reliable results, there are some filters applied at generator level namely BCtoE and EMEnriched resulting in events which are suitable for the analyses containing electrons. The first filter tries to save those events containing electrons with at least 10 GeV of transverse energy that are produced in the tracker acceptance. These electrons are asked to originate from a b or c quark so that events which are able to pass this filter are enriched with electrons appearing in the decay

products of hadrons containing heavy quarks. The second filter aims to gather those events with isolated electrons at generator level. It looks for single electrons with at least 20 GeV of transverse energy within the tracker acceptance that are isolated in both the tracker and calorimeter sub-detectors. Both filters are made in such a way that the QCD events surviving each filter should be used together. For simplicity QCD events are produced in different bins of $\hat{p}_T$, which is defined as the center of mass energy of the two colliding partons. The detailed information of the Monte Carlo simulated sets which are processed in this study, are given in Table 4.3.

Process	Name of Monte Carlo Sample	σ[pb]	#Events
$t\bar{t}$			
	/TTbarJets-madgraph	157.5	1483404
W+jets ($W \to l\nu$)			
	/WJets-madgraph	31314	10068895
Z+jets ($Z \to l^+l^-$)			
	/ZJets-madgraph	3048	1084921
t			
	/SingleTop_tChannel-madgraph	20.93	528593
	/SingleTop_tWChannel-madgraph	10.6	459589
	/SingleTop_sChannel-madgraph	1.4	402055
QCD			
	/QCD_BCtoE_Pt20to30	108330	2491921
	/QCD_BCtoE_Pt30to80	138762	2475597
	/QCD_BCtoE_Pt80to170	9422.4	1119546
	/QCD_EMEnriched_Pt20to30	1719150	33355445
	/QCD_EMEnriched_Pt30to80	3498700	39479587
	/QCD_EMEnriched_Pt80to170	134088	5494911

Table 4.3: Simulated signal and background samples that are processed in this study. All samples have the ending "/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO" which has been dropped in the table. The mentioned cross sections correspond to the Next-to-Leading Order (NLO) values except for QCD samples where the quoted cross sections have been calculated at Leading Order (LO) [96]. The numbers are taken from a common reference page of the top quark analysis group in CMS [97].

4.4 Event Selection

Together with signal events, at a proton-proton collider, there are many other kind of events produced of which some mimic the signal. For example, QCD multijet events where a jet fakes an electron can produce a similar final state as e+jets $t\bar{t}$ events. The same reasoning can be made for a W+jets event when the W decays leptonically. Also $Z \rightarrow e^+e^-$ events of which one electron escapes detection would be another possible background. Hence, in order to suppress all these background events, one can benifit from the kinematics of the signal and cut on the observables that differ between the signal and its backgrounds. A detailed procedure of selecting e+jets $t\bar{t}$ events among all other backgrounds is described in the next section.

4.4.1 Selection Cuts

Expected Number of Events at the Center of Mass Energy of 7TeV

At 7 TeV center of mass energy, the cross sections for signal and other relevant backgrounds are listed in Table 4.3. According to

$$N = \sigma L$$

where L is the integrated luminosity and σ is the cross section for a specific kind of process, N is the number of events expected after accumulating the data equivalent to L. The current study is based on $L = 100\ pb^{-1}$ and the derived results are valid at this regime of luminosity.

Pre-Selection

In order to reduce the size of the various simulated samples, a set of pre-selection cuts is applied. All events are required to have at least one reconstructed electron and four reconstructed jets. Reconstructed electron and jets are asked to have a p_T exceeding respectively 15 GeV and 20 GeV and to be within the tracker acceptance of $|\eta| < 2.4$. The p_T and η distributions of all reconstructed electrons and jets for those events surviving the pre-selection including all signal and backgrounds are shown in Figure 4.6.

As it can be seen from Figure 4.6, the distribution of the pseudorapidity of the electrons drop down in two parts which are representing the cracks between the ECAL barrel and the ECAL endcap. Also it can be understood that the jets in the semi electronic $t\bar{t}$ events, are mainly reconstructed in the central part of the detector.

Primary Vertex

In order to make sure that the event which is under processing, comes out of a hard scattering and has not originated from pile-up interactions, some quality cuts are applied on the reconstructed vertex of the event. Since one of the characteristics of pile-up collision is a displacement primary vertex with respect to the nominal interaction vertex, one can ask each event to have a reconstructed primary vertex very close to the

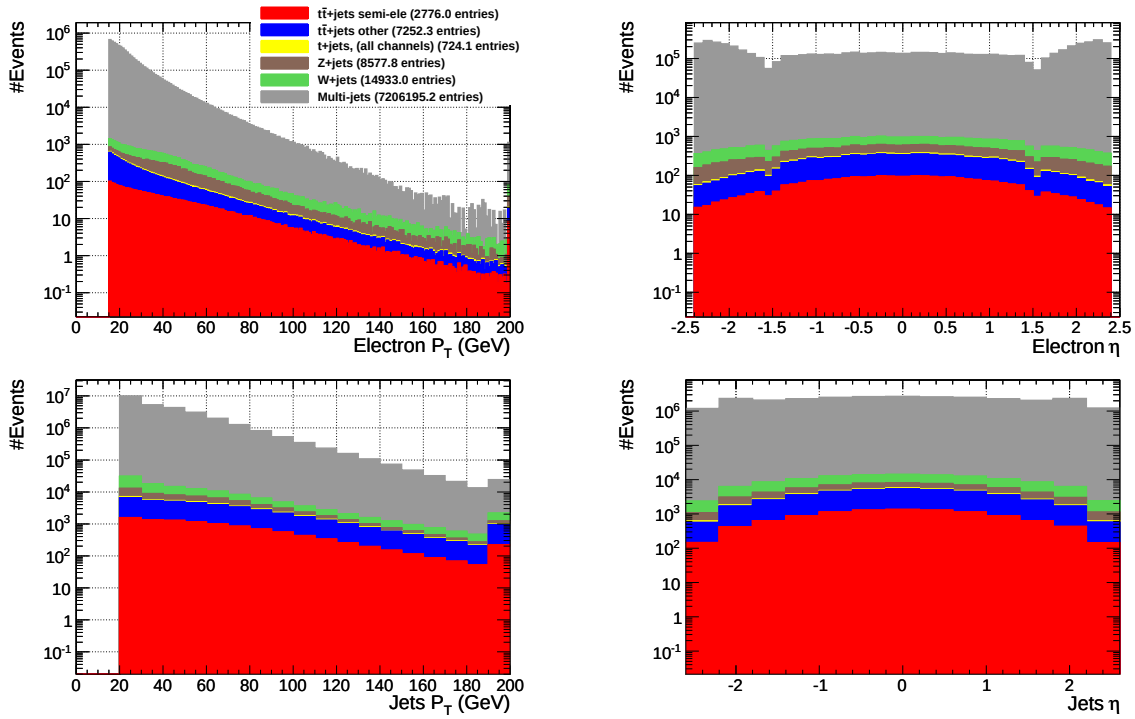

Figure 4.6: The p_T and η distributions of all reconstructed electrons and jets for the events surviving the pre-selection step.

interaction region. In this analysis, each event is required to have a good reconstructed primary vertex which is defined with $ndof > 4$, where the number of degrees of freedom is a measure of the number of tracks used for the construction of the primary vertex. Also a good primary vertex fulfills $|z| < 24\ cm$ and $\rho < 2\ cm$, where $\rho \equiv \sqrt{x^2 + y^2}$ with x and y the coordinates of the reconstructed primary vertex in the transverse plain. The multiplicity of good primary vertices for events which have already passed the pre-selection cuts is shown in Figure 4.7.

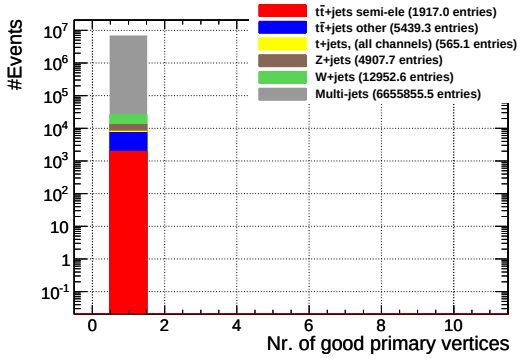

Figure 4.7: Distribution of the good primary vertices in an event after applying pre-selection cuts.

Exactly One Electron

Since one of the signal signatures is the presence of exactly one isolated electron in the final state, cutting on this property of the event will reduce the huge amount of backgrounds specially QCD multijet events where no prompt electron is usually produced. The selected electron is required to pass some more tighter cuts than the ones already applied at the pre-selection step. A detailed set of cuts that the selected electron is asked to pass, is listed below.

- The selected electron for each event, if there exists one, is required to pass a tighter cut on p_T namely $p_T > 30\ GeV$. Also, it is asked not to be reconstructed in the transition region between the ECAL barrel and the ECAL endcap. As a result, if the η of the supercluster of a candidate electron fulfills $1.4442 < |\eta^{supercluster}| < 1.5660$, then it is failed to be labeled as selected.

- Another powerfull observable that is widely used to distinguish between prompt electrons coming out of the hard scattering and those which are produced in jets together with lots of hadrons, is the isolation variable. The isolation value, as the name says, is a measure of how much an electron is isolated with respect to the other activity in the event and defined as the sum of the energy deposits in a cone with a specific radius around the electron. Instead of an absolute isolation value, a relative isolation variable is mostly used which is defined as

$$RelIso = \frac{caloIso^{ECAL+HCAL} + trackIso}{E_T^e},$$

  where $caloIso^{ECAL+HCAL}$ and $trackIso$ refer to the absolute observed transverse momentum or energy obtained in a cone of 0.3 around the electron axis in the calorimeter and the tracker, respectively. The distribution of the relative isolation variable of all reconstructed electrons for those events that already passed the primary vertex cut, is plotted in Figure 4.8. The distribution of the relative isolation variable of the electron with the highest p_T in the event is also shown in Figure 4.9. An electron is selected if its relative isolation value is below 0.1.

- In order to be selected, the electron candidate is asked to pass the identification cuts. The parameters used for electron identification aim to make a difference between the shape of the showering of a reconstructed electron compared to a jet. They are briefly reviewed below. The cut values are chosen so that an identification efficiency equal to 70% can be reached [98].

  - h/e which is defined as the ratio of the energy deposited in the HCAL tower just behind the electromagnetic seed cluster over the energy of the seed cluster. The selected electron is required to have an h/e value less than 0.025.

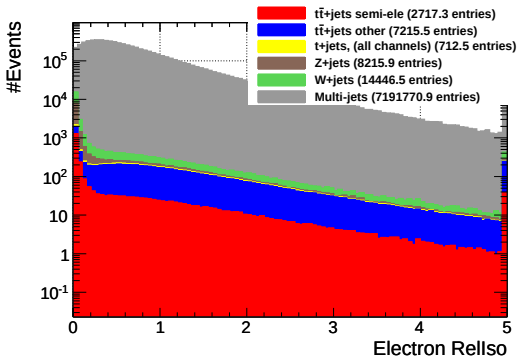

Figure 4.8: Relative isolation variable of all reconstructed electrons for the events passed primary vertex cut.

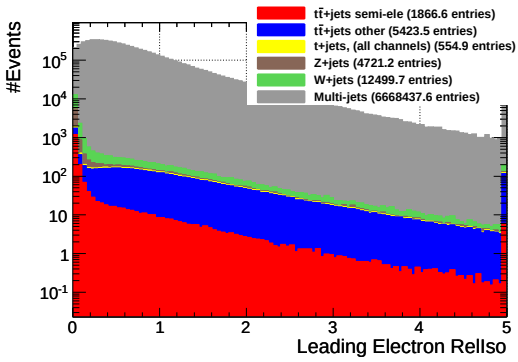

Figure 4.9: Relative isolation variable of the electron with the maximum p_T for the events passed primary vertex cut.

- $\Delta\eta_{in}$ which is defined as the η difference between the supercluster and the corresponding track extrapolated to the ECAL starting from the interaction vertex. Depending on the position of the reconstructed electron, a barrel electron is asked to have a $\Delta\eta_{in}$ less than 0.004 while for the endcap electron this value is set to be 0.005.

- $\Delta\phi_{in}$ which is defined as the ϕ difference between the supercluster and the corresponding track extrapolated to the ECAL starting from the interaction vertex. Depending on the position of the reconstructed electron, a barrel electron is asked to have a $\Delta\phi_{in}$ less than 0.03 while for the endcap electron this value is set to be 0.02.

- $\sigma_{i\eta i\eta}$ is a measure of the shower width in a direction which is very largely unaffected by showering of electrons in the tracker material and so has been a key variable in electron identification. Depending on the position of the reconstructed electron, a barrel electron is asked to have a $\sigma_{i\eta i\eta}$ less than 0.01 while for the endcap electron this value is set to be 0.03.

- Finally, the selected electron is required to have $d_0 < 0.02\ cm$ where d_0 is defined as the distance of the reconstructed track of the electron in the transverse plain with respect to the beam spot. This variable is useful in differentiating between prompt electrons and those which are produced in jets. The electrons appearing in jets usually have larger d_0 values. Hence cutting on this variable would reject most of the QCD events.

A summary of the cut values applied on the various identification variables of the selected electron is given in Table 4.4.

	h/e		$\Delta\eta_{in}$		$\Delta\phi_{in}$		$\sigma_{i\eta i\eta}$	
	barrel	endcap	barrel	endcap	barrel	endcap	barrel	endcap
cut value	0.025	0.025	0.004	0.005	0.03	0.02	0.01	0.03

Table 4.4: Required cut values on the identification variables of the selected electron.

The distributions of the shower shape variables together with the d_0 parameter of all reconstructed electrons in those events that already passed the primary vertex cut, can be found in Figure 4.10.

As it is seen from Figure 4.10, the distribution of the $\Delta\eta_{in}$ variable is sharper compared to the $\Delta\phi_{in}$ distribution, representing the radiated electrons which bend in the ϕ-direction due to the magnetic field in the z-direction.

Also the multiplicity of the selected electrons per event which already passed the primary vertex cut, is shown in Figure 4.11.

It can be seen from Figure 4.11 that most of the QCD multijet events do not contain a selected electron while two selected electrons can be found in Z+jets events or $t\bar{t} \rightarrow$ other, as expected.

Those events with $N_e^{sel} = 1$, corresponding to the situation that exactly one selected electron is present per event, are accepted for further analysis.

Second Electron Veto

Some of the backgrounds such as Z+jets events where $Z \rightarrow e^+e^-$ or backgrounds coming from full leptonic decay of $t\bar{t}$ events of which both of leptons are electrons, contain more than one electron in their final states. Hence vetoing of the presence of a second electron can suppress these kind of backgrounds. For the second electron, there are some looser cuts applied so that the background rejection efficiency is increased. The p_T cut is lowered to 20 GeV compared to the selected electron. The cut thresholds on the different identification parameters are also loosen and are listed in Table 4.5. No isolation requirement neither d_0 cut is applied on the second electron.

The multiplicity of the second loose electron per event, excluding the selected electron, is shown in Figure 4.12.

The event is rejected if it has at least one second loose electron as defined above.

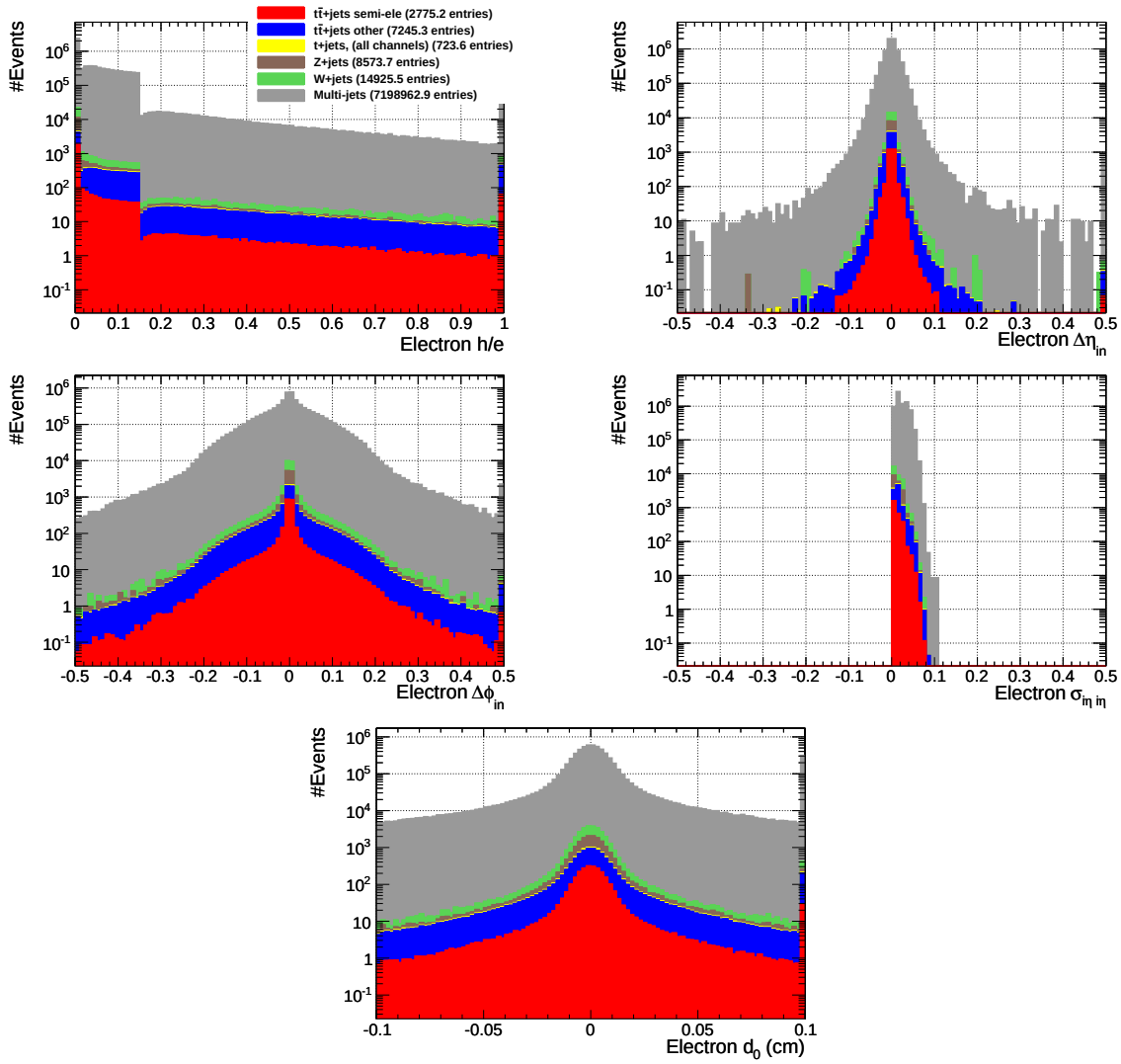

Figure 4.10: Shower shape variables and d_0 distribution of all reconstructed electrons in the events passing the primary vertex cut.

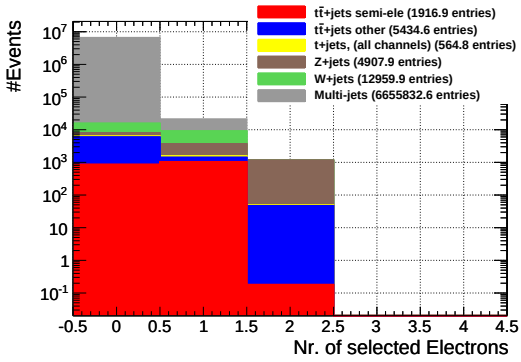

Figure 4.11: Multiplicity of the selected electrons per event which passed the primary vertex cut.

	h/e		$\Delta\eta_{in}$		$\Delta\phi_{in}$		$\sigma_{i\eta i\eta}$	
	barrel	endcap	barrel	endcap	barrel	endcap	barrel	endcap
cut value	0.15	0.07	0.007	0.01	0.8	0.7	0.01	0.03

Table 4.5: Required cut values on the identification variables of the second electron.

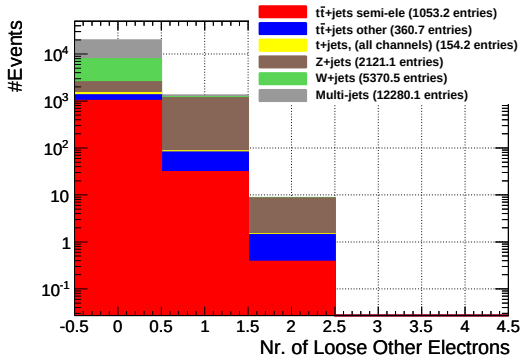

Figure 4.12: Multiplicity of secondary loose electrons per event which has already had exactly one selected electron.

Muon Veto

The background events from μ+jets $t\bar{t}$ decays contain one muon in their final states. Also a muon can be reconstructed in the final state of a W+jets event where the W boson decays to a muon and corresponding neutrino. Hence vetoing muons in the final state of the event would reject most of the backgrounds containing a reconstructed muon. The distribution of the muon muliplicity for those events that already passed the second electron veto cut, is shown in Figure 4.13. The reconstructed muons are asked to have a p_T exceeding 20 GeV and $|\eta|$ less than 2.1. Also they are required to fulfill a relative isolation value, which is defined as the sum of the energy deposits in a cone of 0.3 in the tracker and the calorimeter divided by the muon transverse momentum, below 0.05. It can be seen that most of the $t\bar{t}$ events other than signal occupy the second bin of the distribution corresponding to those events with one reconstructed muon in the final state. Hence these kind of backgrounds are suppressed by vetoing the existence of a reconstructed muon in the event.

Only those events appearing in the first bin of the plot shown in Figure 4.13, corresponding to the events with $N_{\mu}^{sel} = 0$, are processed in the rest of analysis.

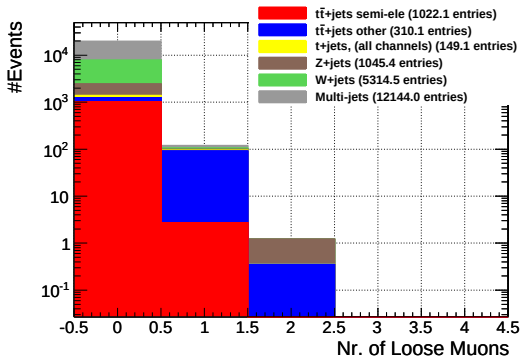

Figure 4.13: Muon multiplicity per event which passed the Second Electron Veto cut.

At Least Four Jets

At the pre-selection step, all events are required to have at least four reconstructed jets in their final states. By tightening the thresholds on the jet kinematics and also by applying some identification criteria, the probability of selecting the four jets originating from the quarks coming out of the hard scattering process becomes higher. The event is required to have at least four selected jets with p_T greater than 30 GeV. Also jets, in order to be labeled as selected, are asked to pass identification cuts that are briefly described below.

- f_{EM} which is called the electromagnetic energy fraction and is defined as the fraction of the jet energy measured in the ECAL. This parameter is useful in differentiating between electrons and jets since jets with a high value of f_{EM} probably are electrons with energy deposits also used in the jet reconstruction algorithm. The upper limit of this variable is set to be 0.9. Also in order to suppress those jets which are actually pure electronic noise, a lower limit equal to 0.05 is put on this parameter.

- n90Hits is defined as the minimal number of reconstructed hits containing 90% of the jet energy and is another observable that is used for the jet identification. The selected jets are required to have a n90Hits variable exceeding 4.

- f_{HPD} is defined as the fraction of energy in the hottest HPD readout. The selected jets are asked to fulfill $f_{HPD} < 0.98$.

The distributions of the various identification variables of all reconstructed jets for those events that already passed the Muon Veto cut, are shown in Figure 4.14.

Looking at the distribution of the electromagnetic energy fraction, there is a clear peak at $\sim$1 representing the electrons which are re-reconstructed as jets. Another plot that shows the jet collection includes electrons and therefore needs to be cleaned, is the distance between the selected electron and the closest jet among all reconstructed

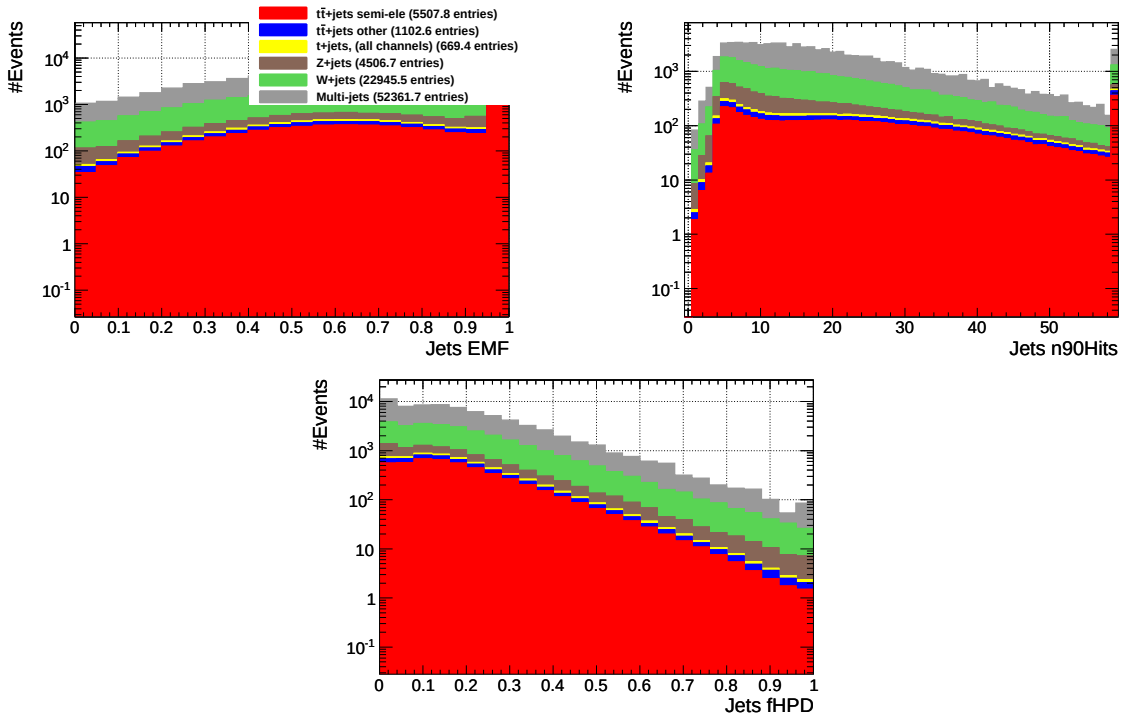

Figure 4.14: Jet identification variables of all reconstructed jets for the events which pass the Muon Veto cut.

jets and is shown in the left plot of Figure 4.15. Hence an extra criterion is applied on each selected jet requiring it not to be located very close to the selected electron. All selected jets are asked to be reconstructed in a $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ cone of 0.3 from the selected electron. For comparison, the same plot of the distance between the selected electron and the closest jet after cleaning is also shown in the right plot of Figure 4.15.

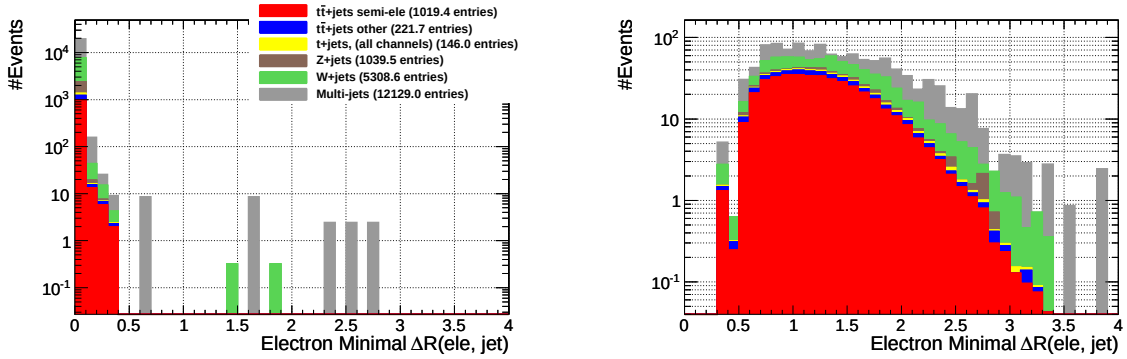

Figure 4.15: Minimum ΔR distance between selected electron and the closest reconstructed jet before (left) and after (right) cleaning. Left plot is filled for the events passing the Muon Veto cut while the events with at least four selected jets populate the right plot.

The multiplicity of the selected jets per event which already passed the Muon Veto cut, is shown in Figure 4.16. Those events with more than or equal to four selected jets, are accepted for further analysis.

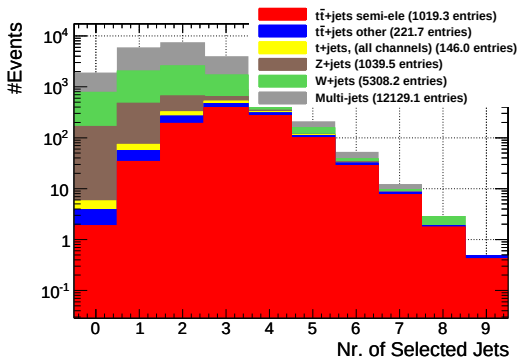

Figure 4.16: Multiplicity of the selected jets per event which pass the Muon Veto cut.

At Least One B-Tagged Jet

As there are two b jets produced in the final state of the signal topology, it can be used as a discriminator between signal and backgrounds. In order to identify b quark jets, various b-tagging algorithms have been defined within the CMS collaboration. A simple one is called the "Track Counting High Efficiency" algorithm [99] which exploits the long lifetime of the b mesons produced in the hadronization process of the b quarks. Because of this property, the tracks which are associated to a b quark jet, are usually displaced with respect to the primary vertex.
The algorithm to tag b jets starts by selecting good quality tracks and associating them to a jet by a matching procedure. In a first step, the track impact parameter IP is computed by the algorithm, which is defined as the distance between the primary vertex and the linearised track in the point of minimum distance between the track and the jet axis. This is followed by the calculation of the track impact parameter significance IP_S, which is defined as the impact parameter divided by the resolution of the impact parameter σ_{IP}, hence $IP_S = \frac{IP}{\sigma_{IP}}$. In the next step, the tracks associated to a jet are ordered in decreasing IP_S. Finally, the "Track Counting High Efficiency" algorithm returns the IP_S assigned to the second track in the list of ordered tracks as the b-tagging discriminant. The distribution of the b-tag discriminator value returned by this algorithms, is shown in Figure 4.17. The events with at least four selected jets are taken into account.

Due to the fact that the efficiency of b-tagging is not high, and in order not to loose too many signal events, the event is asked to have at least one b-tagged jet among the four selected jets. A jet is defined as a b jet if the tagging value returned by the "Track Counting High Efficiency" algorithm exceeds 4. The multiplicity of b-tagged jets among the four leading jets for those events that have at least four selected jets is

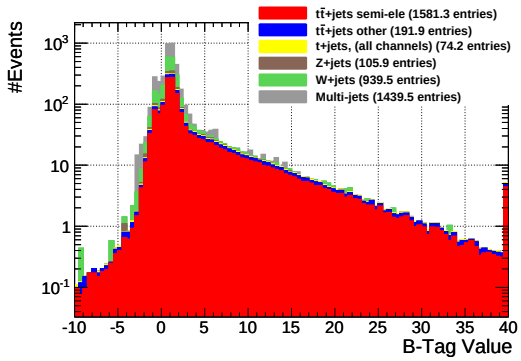

Figure 4.17: The distribution of the b-tagging discriminator of the four leading jets, calculated by the "Track Counting High Efficiency" algorithm, for the events which have at least four selected jets.

shown in Figure 4.18.

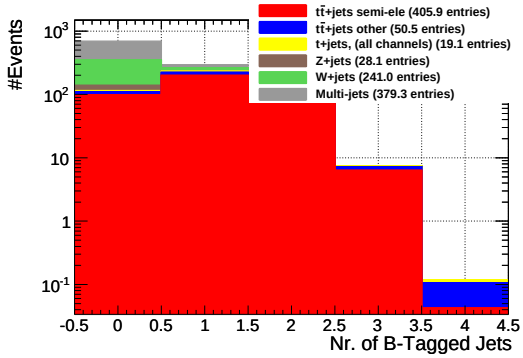

Figure 4.18: Multiplicity of the b-tagged jets, which is defined as a jet with a b-tagging value returned by the "Track Counting High Efficiency" algorithm exceeding 4, for the events which have at least four selected jets.

Most of the W+jets backgrounds as well as the QCD multijet events do not contain a b-tagged jet and are suppressed when requiring the event to have at least one b jet candidate.
A summary of all selection cuts and the number of events surviving each step for various simulated signal and background samples is listed in Table 4.6.

Having applied the above-mentioned selection cuts, one would get a signal over background ratio of about 2.5. Hence the final sample of selected events is dominated by signal events and the purity of selecting the signal events can be further improved by applying additional cuts which are described in more detail in the next chapter.

	$t\bar{t}$ $\rightarrow e + jets$	$t\bar{t}$ $\rightarrow$ other	W+jets $\rightarrow l + jets$	Z+jets $\rightarrow l^+l^- + jets$	t $\rightarrow l + jets$	QCD multijets
Initial	2333.0	13417.0	3131400.0	304800.0	3293.0	5.6E+8
Pre-Selection	1917.3	5449.0	12961.9	4909.2	565.0	6681710.1
PV	1916.5	5446.5	12958.5	4907.6	564.8	6680208.7
One Electron	1053.0	360.6	5369.9	2121.1	154.2	12280.2
2^{nd} Electron Veto	1021.9	310.0	5313.9	1045.4	149.1	12135.3
Muon Veto	1019.2	221.7	5308.0	1039.5	146.0	12120.3
Four Jets	405.9	50.5	241.0	28.1	19.1	379.2
One B Jet	307.8	39.1	30.2	4.5	13.5	35.6
Selection Efficiency (%)	13.2	0.3	0.0	0.0	0.4	0.0

Table 4.6: Table of event selection cuts. The numbers are equivalent to an integrated luminosity of 100 pb^{-1}. Also shown in the last row of the table, is the efficiency of the selection cuts which is defined as the survived number of events divided by the initial number of events before applying any cuts. The selection cuts mainly keep the e+jets $t\bar{t}$ events while rejecting the diverse backgrounds.

Chapter 5

Estimating of the Jet Energy Scale Calibration Factor

In the previous chapter, the selection cuts, which are applied to enhance the purity of the e+jets signal events in the final selected sample, were introduced. At the end when the survived event sample is enriched in e+jets $t\bar{t}$ events, one can try to reconstruct the W boson and the top quark four-vectors by combining the individual objects and reconstruct backwards the decay chain. We divide the $t\bar{t}$ system into two separate parts depending on the decay products of the W boson, namely hadronic and leptonic,

$$t^h \rightarrow W^h b^h \rightarrow q\bar{q}b^h,$$

$$t^l \rightarrow W^l b^l \rightarrow e\nu_e b^l,$$

where indexes "h" and "l" refer to the hadronic and the leptonic parts, respectively. Since the four-vector of ν_e is not reconstructed, the four-vectors of the leptonic W boson and subsequently the leptonic top quark can not be made from their corresponding decay products. While the leptonic part would remain non-reconstructed, the hadronic part can be fully reconstructed. In the hadronic part, the light quarks q and $\bar{q}$, hadronize and produce the light jets. Given that the light jets are known among the collection of the reconstructed jets, the corresponding four-vectors are summed and make the four-vector of the hadronic W boson. Hence in order to reconstruct the hadronic W boson, one needs to determine which jets arise from the W boson decay. Consequently, in order to reconstruct the hadronic top quark, one has to determine which jet is the hadronic b jet. Although the leptonic top quark would not be fully reconstructed, in order to benefit from the kinematics of the leptonic b quark, one may be interested in associating a jet to the leptonic b quark, as well. The task of assigning reconstructed jets to partons arising from $t\bar{t} \rightarrow e\nu_e q\bar{q}b\bar{b}$, is a key topic which is explained in more detail in Section 5.1.

Having associated the jets to the partons, one can measure the invariant masses of the hadronic W boson and the hadronic top quark. In Section 5.1.6, it is shown that the measured invariant masses of the W boson as well as of the top quark, deviate from their nominal values that are used in the generation step of the simulation. This observation motivates to correct the energy of the jets so that the measured masses of both the W boson and the top quark are peaked at their generated values. Applying

the mass constraints of both the W boson and the top quark on $t^h \to W^h b \to q\bar{q}b^h$ and requiring the reconstructed jets to fulfill the constraints, while allowing the measured energies to change within their resolutions, would result in obtaining jets with modified energy. The method of applying mass constraints, which makes use of the Lagrange Multipliers, is called a Kinematic Fit and is introduced in Section 5.2.
In addition to the baseline event selection cuts described in the previous chapter, some extra cuts are applied on the selected events which are introduced in Section 5.3. The method to estimate the residual jet energy calibration factors is described in Section 5.4 and the statistical properties of these estimators are discussed in Section 5.5. Finally, the method is applied on the proton-proton collision data which has been collected by the CMS experiment during the physics runs in 2010. The plots which compare the simulation versus data are shown in Section 5.6.

5.1 Multi-Variate Analysis Method (MVA)

In the final state of the e+jets $t\bar{t}$ decay, $t\bar{t} \to e\bar{\nu}_e q\bar{q}b\bar{b}$, there are four quarks produced which should result in at least four reconstructed jets in the detector. If no tagging tools are used to distinguish between light jets, originating from the quarks which appear in the decay products of the hadronic W boson, and b jets, originating from the b quarks which appear in the decay products of both the hadronic and the leptonic top quarks, one can not differentiate between them. Therefore in order to label the jets and assign them to the four quarks, there are $4! = 24$ different ways to do the assignment given that only four reconstructed jets are taken into account. Since the order of the jets which are assigned to the light quarks is not important when reconstructing the W boson momentum, they can be used interchangably. Hence the number of different ways to assign four reconstructed jets to the four quarks is divided by 2 and reduces to 12. The quarks which are produced in the decay products of e+jets $t\bar{t}$ event, initiate from heavy particles such as W boson and top quark. Hence the four leading jets that are the jets with the maximum p_T among the reconstructed jet collection, could be good candidates to represent the four quarks and thus are used to do the assignment. It should be noted that there are usually more than four jets reconstructed in the event due to radiation, hence the four leading jets are not always the correct jets as will be discussed in Section 5.1.1. Another important note to be emphasized is that the electron fakes a jet but is removed from the list of reconstructed jets.
To summarize the above discussion, there exist 12 different combinations in order to associate four leading jets to the four quarks arrising from the e+jets $t\bar{t}$ decay. Among 12 possible combinations, only one combination represents the correct assignment. Since the determination of the correct combination of jets is essential for the rest of analysis, specific efforts should be made in order to select the correct combination.
In order to select the correct jet-parton combination among 12 combinations, when no simulation information is available, it is possible to take randomly one combination. The probability of selecting the correct jet combination randomly is $\frac{1}{12} \sim 8\%$ as only one among twelve comibnations is the right candidate in each trial. The probability of selecting the right jet combination is further improved by considering the kinematics of some variables that can differentiate between the right and wrong jet combinations.

This is the subject of the methods, generally called Multi-Variate Analyses (MVA), which combine the information of different variables and extract a value that can discriminate between the right and wrong jet combinations.
There are many MVA methods implemented in order to separate signal, which is in this case a right jet combination, from backgrounds, which are the other eleven wrong jet combinations [100]. The Likelihood ratio method is used in this analysis in order to label the jets and select the correct jet combination.

5.1.1 Jet-Parton Matching Algorithm

When analyzing simulated samples for which the information from the generator level can be accessed, a Monte Carlo truth jet-parton matching can be introduced. Even in case of access to the full information of the partons produced in the final state, including their four-vectors, no unambiguous definition of matching is present. There are many algorithms to associate the reconstructed jets to the generated partons [101]. In general they match jets to partons based on the spatial distance between them. The one, which is used in this analysis, is called `ptOrderedMinDist` and explained briefly here.
The algorithm starts by making a p_T-ordered list of partons produced in the final state. Then it looks for the closest reconstructed jet to the leading parton by matching them in (η, ϕ) space. A reconstructed jet is assigned to the leading parton if

$$\Delta R = \sqrt{(\eta^{parton} - \eta^{jet})^2 + (\phi^{parton} - \phi^{jet})^2} < 0.3.$$

In case such a jet is found, the jet is matched to that parton and is subsequently removed from the list of the reconstructed jets. The algorithm continues by taking the second leading parton in the event. In this procedure, if there is no jet found close to the parton, the parton remaines unmatched.
Applying the above mentioned jet-parton matching algorithm, it is found that in only 26% of the selected e+jets $t\bar{t}$ events, a correct jet-parton matching exists. This relatively low fraction of selected events, for which the four leading jets are matched to the four quarks arrise in $t\bar{t} \rightarrow e\nu_e q\bar{q}b\bar{b}$, can be explained as follows. In most of the e+jets signal events, one can find energetic jets originating from Initial State Radiation (ISR) which finally appear in the list of four leading jets. The appearance of the ISR jets in the list of four leading jets consequently spoils the procedure of matching of the hard-scatter partons to the four leading jets. This physics result can be understood by looking at the plot in Figure 5.1.
According to Figure 5.1, which shows the multiplicity of the generated ISR jets with $p_T > 30\ GeV$ per selected signal event, the fraction of the selected e+jets $t\bar{t}$ events that contain at least one ISR with a p_T exceeding 30 GeV would reach to 74%.
Also it can be calculated how often an ISR jet is reconstructed among the four leading jets. As understood from Figure 5.2, which shows the number of ISR jets that can be matched to the four leading jets per selected signal event, in 56% of all selected e+jets $t\bar{t}$ events, there is at least one ISR that is matched to a jet among the four leading jets.

Therefore, the presence of ISR jets among four leading reconstructed jets in signal events can partially explain why in a large fraction of the selected events, the four

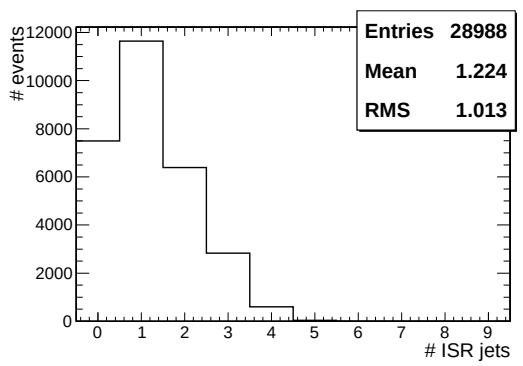

Figure 5.1: The number of ISR jets with $p_T > 30\ GeV$ per selected signal event.

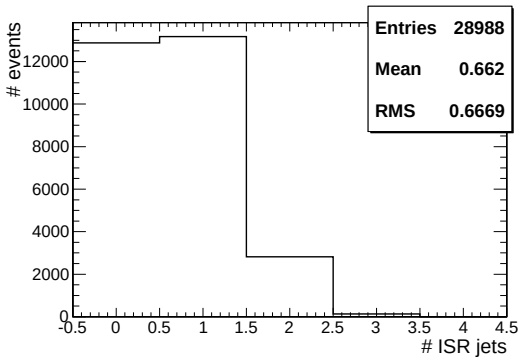

Figure 5.2: The number of ISR jets that can be matched to the four leading jets per selected signal event.

leading jets can not be matched to the four quarks in the final state of $t\bar{t}$ system. Another reason that can explain why the partons produced in the hard scattering $t\bar{t}$ process, do not match to the four leading reconstructed jets, is due to the acceptance requirement. Since reconstructed jets have already passed the cuts on η and are not allowed to be located in the forward region, the fraction of events where at least one parton is produced out of acceptance region, would not have matched jets. Final State Radiation (FSR) can spoil the procedure of matching of partons to reconstructed jets, too. Since FSR can split the initial parton which consequently yields two separate jets, therefore the jets might be reconstructed far from the direction of the original parton and can not be found in the matching algorithm.

5.1.2 Likelihood Ratio Method

The normalized distributions of the various discriminating variables, which make the Probability Density Functions (PDFs) of these variables, are used to define the signal and background likelihoods, $\mathcal{L}_S$ and $\mathcal{L}_B$ respectively. A likelihood function is defined as the product of the PDFs $P_k(x_k)$ of all input variables x_k, which constitute n_{var} variables and can be expressed for signal and background as

$$\mathcal{L}_S = \prod_{k=1}^{n_{var}} P_{S,k}(x_k),$$

and

$$\mathcal{L}_B = \prod_{k=1}^{n_{var}} P_{B,k}(x_k),$$

respectively. For each combination among twelve possible combinations, the Likelihood ratio $y_{\mathcal{L}}$ is defined as the signal likelihood divided by the sum of the signal and background likelihoods

$$y_{\mathcal{L}} = \frac{\mathcal{L}_S}{\mathcal{L}_S + \mathcal{L}_B}.$$

The ratio $y_{\mathcal{L}}$ is then calculated for every jet combination in each event. Per event, the jet combination whose $y_{\mathcal{L}}$ is the largest, is chosen and returned by the MVA method. Almost all the MVA methods which include a Likelihood Ratio method need to be trained before applying the method on data. In the training phase, one introduces the signal and background PDFs to the MVA method. Then the MVA method trains itself and learns what kind of behaviours can be extracted from both the signal and backgrounds PDFs. In the application phase, using the information which has been collected in the training phase, the MVA method returns a value per event and categorizes that event as signal or background. In case of a jet combination study, where one signal against eleven backgrounds is present per event, the signal is defined as the jet combination with the highest value returned by the MVA method, although in some cases the MVA method is not able to return the "true" jet combination, which corresponds to the correct jet combination that is matched with the hard-scatter partons.

5.1.3 Likelihood Concept in Bayesian Statistics

According to the Bayes' theorem, posterior probability $p(Y|X)$, is related to prior probability $p(Y)$, with the following equation

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)},$$

where $p(X|Y)$ is the conditional probability and is also referred to as the Likelihood. In the above equation, $p(X|Y)$ is the probability distribution of the parameter X, which is usually a continuous variable of the event that belongs to the class Y and $p(Y|X)$ is then defined as the probability of assigning a new observed event to the class Y given that the value X is measured for that particular event. Therefore in order to obtain the posterior probability $p(Y|X)$, in addition to the prior probability $p(Y)$, $p(X|Y)$ should also be known. In the language of multi-variate techniques, $p(X|Y)$ is obtained in the training phase. During training, the classes Y with their properties X are introduced to the trainer and subsequently the corresponding PDFs are extracted.
In order to clarify a bit more how a generic multi variate method works, a simplified example based on Bayes' theorem is explained here. Consider a space with one variable X that can take only two values, namely "$X = 1$" or "$X = 2$". In this space, events are categorized in either signal or background classes, hence "$Y = S$" or "$Y = B$". An event is said to be measured when the X value of that particular event is determined. Assume 10 such events are selected and fed to an MVA method. The results of training over 10 events, are summarized in Figure 5.3.

The information, containing the conditional as well as the prior probabilities, that can be derived from Figure 5.3, is listed below.

$$p(1|S) = \frac{3}{5},\ p(1|B) = \frac{1}{5},$$

$$p(2|S) = \frac{2}{5},\ p(2|B) = \frac{4}{5},$$

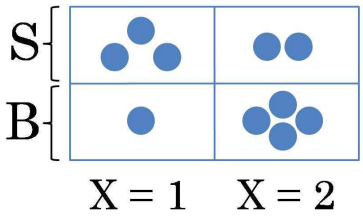

Figure 5.3: A simple example to show the basic concepts of the MVA method. In this example, the circles, which are representing events, are observed in two different bins, being either "1" or "2" and can belong to one of "S" or "B" classes.

$$p(S) = \frac{5}{10},$$
$$p(B) = \frac{5}{10}.$$

From the above elements, $p(1)$ and $p(2)$ can also be extracted, according to the "sum rule" of the probability theory, as follows

$$p(1) = p(1|S)p(S) + p(1|B)p(B) = \frac{2}{5},$$

$$p(2) = p(2|S)p(S) + p(2|B)p(B) = \frac{3}{5}.$$

Now that all the input information is complete, one can calculate the posterior probabilities as expressed below

$$p(S|1) = \frac{p(1|S)p(S)}{p(1)} = \frac{9}{12},$$

$$p(S|2) = \frac{p(2|S)p(S)}{p(2)} = \frac{4}{12},$$

$$p(B|1) = \frac{p(1|B)p(B)}{p(1)} = \frac{3}{12},$$

$$p(B|2) = \frac{p(2|B)p(B)}{p(2)} = \frac{8}{12}.$$

The above numbers can be interpreted as follows. Any new event, whose measured X value yields to $X = 1$, would be assigned to class S since the probability of being a type "S" is three times more than being a type "B" event when a value equals 1 is measured. With the same reasoning, any new event is grouped to the class B given that the meaurement of the X property of that event results in $X = 2$.
The above example shows the basic idea of assigning a new observed event to a specific class in a one-dimensional phase space. In most cases, usually more than one input variable is used. As a result, the problem of labeling an event as signal or background, would not be that simple. Defining a single variable out of many input variables can provide a possible solution. Hence the Likelihood function $\mathcal{L}(\vec{x})$, which combines the

information of all input variables $\vec{x} = (x_1 \dots x_n)$ as explained in Section 5.1.2, is introduced in an n-dimensional phase space.
By definition, $\mathcal{L}$ can take values in a range of [0,1]. Therefore, in the language of Likelihood Ratio, an event is assigned as signal if the measurement of its X property would happen in the bin whose Likelihood Ratio value is greater than 0.5. Then the event is more signal-like compared to background-like. Otherwise, it is labeled as background. In case of finding the correct jet combination, where eleven backgrounds, being wrong combinations, versus one signal, being the true combination, exist, the chosen combination is defined as the combination whose Likelihood Ratio value is the maximum among the other combinations.

5.1.4 Input Variables for Training

Various observable variables can be used to train the Likelihood Ratio method. Also a combination of two or three variables is allowed. Among so many candidate variables which can be used in the training, the most discriminating variables that differentiate maximally between signal and backgrond, are desired. The discriminating power of a generic variable can be defined using different methods. As an example, the "separation" S, which can be a measure of discriminating power between signal and background, is defined as

$$S = \frac{1}{2} \int \frac{(P_S(X) - P_B(X))^2}{(P_S(X) + P_B(X))} dX,$$

where $P_S(X)$ and $P_B(X)$ are the PDFs of the observable X for signal and background, respectively. By definition, for identical signal and background shapes, the separation is zero. If the signal and background PDFs do not overlap, then the separation takes a value equal to one. Therefore higher values of the separation, indicate the variable has a higher probability to be a good discriminant candidate.
Since the four vector of the neutrino is not known, hence the leptonic W boson and consequently the leptonic top quark can not be fully reconstructed. Therefore the input variables which can be chosen to train the MVA method with, should be independent from the reconstructed neutrino. This makes the input candidates to be limited to a smaller collection. Different kinematics of the objects such as transverse momentum p_T, pseudorapidity η, polar angle θ and azimuthal angle ϕ can be used to compare the candidates in $t\bar{t}$ system. In this analysis, the Θ variable, which is defined as the space angle between two vectors in a three-dimensional phase space is used which is expressed as

$$\Theta(\vec{v}_1, \vec{v}_2) = \cos^{-1}\left(\frac{\vec{v}_1.\vec{v}_2}{|\vec{v}_1||\vec{v}_2|}\right) = \cos^{-1}(\sin\theta_1 \sin\theta_2(\cos(\phi_1 - \phi_2)) + \cos\theta_1 \cos\theta_2).$$

The two-component variables in the $t\bar{t}$ system, excluding those variables related to the kinematics of neutrino, which are considered in this analysis, are listed below.

- $\Theta(t^h, W^h)$ which is the space angle between hadronic top quark and hadronic W boson,

- $\Theta(t^h, b^h)$ which is the space angle between hadronic top quark and hadronic b,
- $\Theta(t^h, b^l)$ which is the space angle between hadronic top quark and leptonic b,
- $\Theta(t^h, e)$ which is the space angle between hadronic top quark and electron,
- $\Theta(W^h, b^h)$ which is the space angle between hadronic W boson and hadronic b,
- $\Theta(W^h, b^l)$ which is the space angle between hadronic W boson and leptonic b,
- $\Theta(W^h, e)$ which is the space angle between hadronic W boson and electron,
- $\Theta(b^h, b^l)$ which is the space angle between hadronic b and leptonic b,
- $\Theta(b^h, e)$ which is the space angle between hadronic b and electron,
- $\Theta(b^l, e)$ which is the space angle between leptonic b and electron,
- $\Theta(q, \bar{q})$ which is the space angle between light quarks appearing in the decay of hadronic W boson.

Most of the above mentioned variables are not independent and contain the same information. This can be seen by looking at their linear correlations which are shown in Table 5.1. Since the masses of the W boson and the top quark are going to be used as constraints in this analysis, as will be explained in the following sections, it is also required that the selected variables are not highly correlated with the W boson and top quark masses. The last two rows of Table 5.1 contain the correlation factors among all considered space angle variables with the masses of the W boson and top quark.

The space angle variables are required to have a mutual correlation factor among themselves less than 40%. They are also asked to be correlated with the W boson and the top quark masses less than 20%. Since the separation powers of each individual variable are of the same order, taking into account the requirement on the correlation factor, only 5 variables among the above mentioned space angle variables are selected to be used in the training of the MVA method. In addition, there are two more variables which are selected for training. A first one, which is defined as the transverse momentum of the hadronic top quark candidate relative to the sum of the transverse momenta of all hadronic top quark candidates $\frac{p_T^{t^h}}{\Sigma p_T}$. Since with the four leading jets, there are four different ways to combine three of them to reconstruct a top quark candidate, then the sum in the denominator is performed over the transverse momenta

	$\Theta(t^h, W^h)$	$\Theta(t^h, b^h)$	$\Theta(t^h, b^l)$	$\Theta(t^h, e)$	$\Theta(W^h, b^h)$	$\Theta(W^h, b^l)$	$\Theta(W^h, e)$	$\Theta(b^h, b^l)$	$\Theta(b^h, e)$	$\Theta(b^l, e)$	$\Theta(q, \bar{q})$
$\Theta(t^h, W^h)$	+100.0	-	-	-	-	-	-	-	-	-	-
$\Theta(t^h, b^h)$	-10.0	+100.0	-	-	-	-	-	-	-	-	-
$\Theta(t^h, b^l)$	+5.6	+6.9	+100.0	-	-	-	-	-	-	-	-
$\Theta(t^h, e)$	+5.2	+5.8	+40.8	+100.0	-	-	-	-	-	-	-
$\Theta(W^h, b^h)$	+32.4	+67.9	+7.9	+8.7	+100.0	-	-	-	-	-	-
$\Theta(W^h, b^l)$	+2.0	+7.3	+60.1	+36.5	+8.0	+100.0	-	-	-	-	-
$\Theta(W^h, e)$	+0.5	+4.7	+24.0	+61.1	+4.4	+35.2	+100.0	-	-	-	-
$\Theta(b^h, b^l)$	+3.6	-4.4	+43.2	+32.1	-1.4	+32.7	+33.6	+100.0	-	-	-
$\Theta(b^h, e)$	+3.8	-3.7	+18.3	+43.5	-0.1	+25.6	+31.5	+31.1	+100.0	-	-
$\Theta(b^l, e)$	+15.0	+11.8	+1.2	+0.5	+15.9	+1.3	-1.3	-1.0	+3.1	+100.0	-
$\Theta(q, \bar{q})$	+16.8	-1.8	+2.2	+4.5	+23.6	+0.7	+1.3	+4.3	+5.8	+4.4	+100.0
m_W	-0.7	+2.0	+0.9	-0.1	+1.4	-1.3	-0.5	+3.2	+1.2	-2.5	+21.3
m_{top}	+3.6	+2.0	+1.7	+2.1	+7.1	-0.4	+1.6	+0.9	+1.2	-1.5	+10.7

Table 5.1: The mutual correlation factors of some space angle variables and their correlation coefficients with the masses of the W boson and the top quark. All numbers are quoted in %.

of the four possible top quark candidates. A second variable is derived from the b-tagging discriminant (β) of each of the two b jet candidates which exist in the final state. The variable which is used in the training is defined as the b-tagging discriminator of the hadronic b jet candidate multiplied by the b-tagging discriminator of the leptonic b jet candidate, hence $\beta^{b^h} \times \beta^{b^l}$.
The final selected variables which are used to train the Likelihood Ratio method, are listed in Table 5.2. The mutual correlation factors between the selected variables and the top quark and the W boson masses are also quoted. The separation power of each individual variable is calculated and given in the last row of Table 5.2.

	$\Theta(t^h, W^h)$	$\Theta(t^h, b^h)$	$\Theta(t^h, b^l)$	$\Theta(t^h, e)$	$\Theta(b^l, e)$	$\frac{p_T^{t^h}}{\Sigma p_T}$	$\beta^{b^h} \times \beta^{b^l}$
m_W	-0.7	+2.0	+0.9	-0.1	-2.5	+0.4	-0.7
m_{top}	+3.6	+2.0	+1.7	+2.1	-1.5	+2.5	+0.9
Separation Power	0.08	0.06	0.20	0.09	0.30	0.10	0.40

Table 5.2: The correlation coefficients of the variables, which are selected to train the MVA method, with the masses of the W boson and the top quark. The separation power of each variable is also shown in the last row. All numbers are quoted in %.

The distribution of the selected variables, which are shown separately for the correct and the wrong combinations, can be found in Figure 5.4 and Figure 5.5. All selected events in which a correct jet-parton combination exists, are taken into account.

5.1.5 The MVA Performance

After training the Likelihood Ratio method using the input variables listed in Table 5.2, one can check the performance of the method by looking at some control plots. For example, the fraction of e+jets $t\bar{t}$ events for which the chosen jet combination returned by the Likelihood Ratio method corresponds to the true jet-quark combination, could be a possible measure to check how well the MVA method works.
Figure 5.6 shows the Likelihood Ratio response, which is defined to be the maximum value among twelve possible values, per event. Only those selected e+jets events for which the hadronic jets are matched to the hadronic quarks in $t \to Wb \to q\bar{q}b$, contribute to this plot. In association with the Likelihood Ratio response, the chosen jet-parton combination is also returned by the Likelihood Ratio method, which was chosen either rightly or wrongly. The response for the events with either right or wrong jet combination chosen by the MVA is shown in Figure 5.6. Here, "Good" jet combination, as labelled in the plot, is defined as the event where the correct combination is chosen as the one where the hadronic jets are matched to the hadronic quarks $t \to Wb \to q\bar{q}b$. Therefore even if the MVA is not able to assign the fourth jet among the four leading jets to the leptonic b quark arrising from the leptonic top quark, the

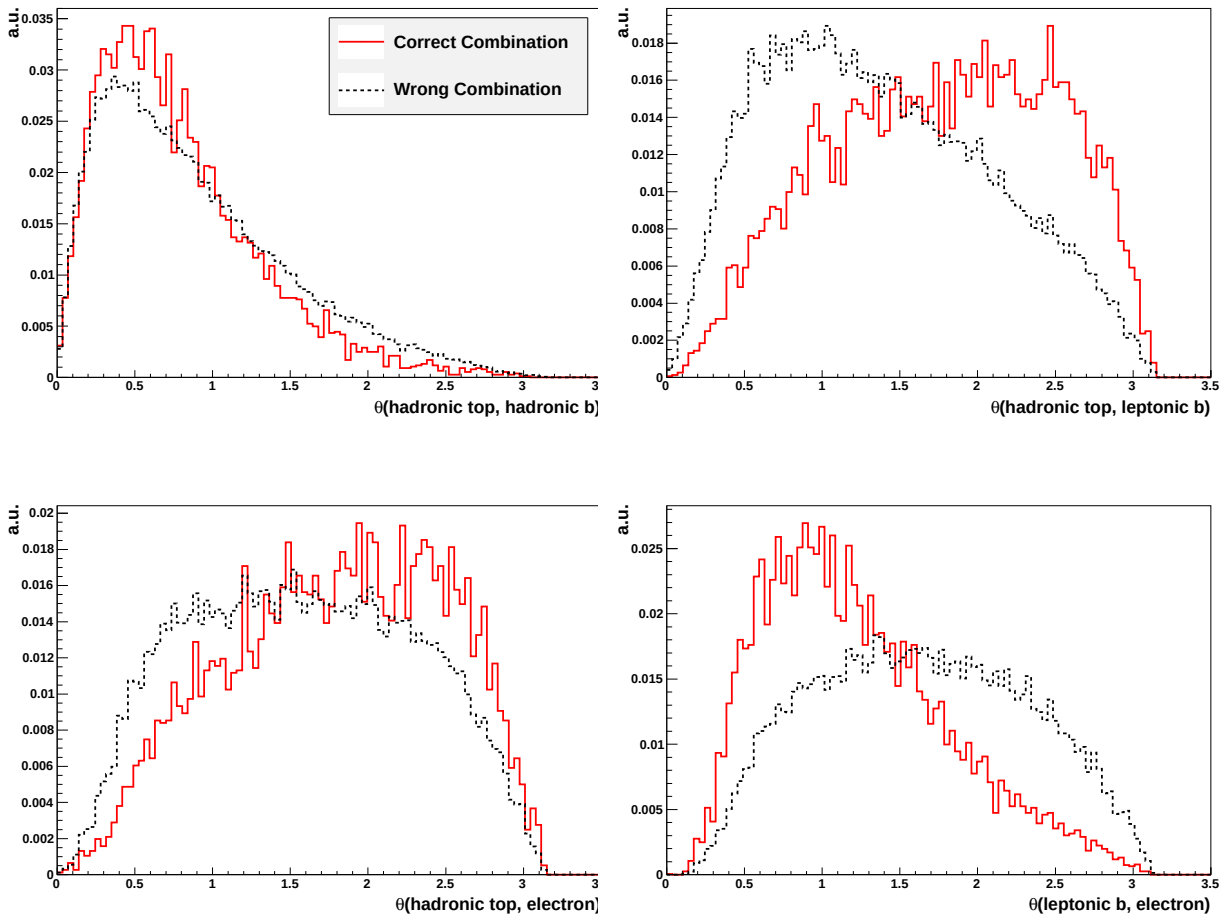

Figure 5.4: The distributions of the selected input variables to train the MVA method.

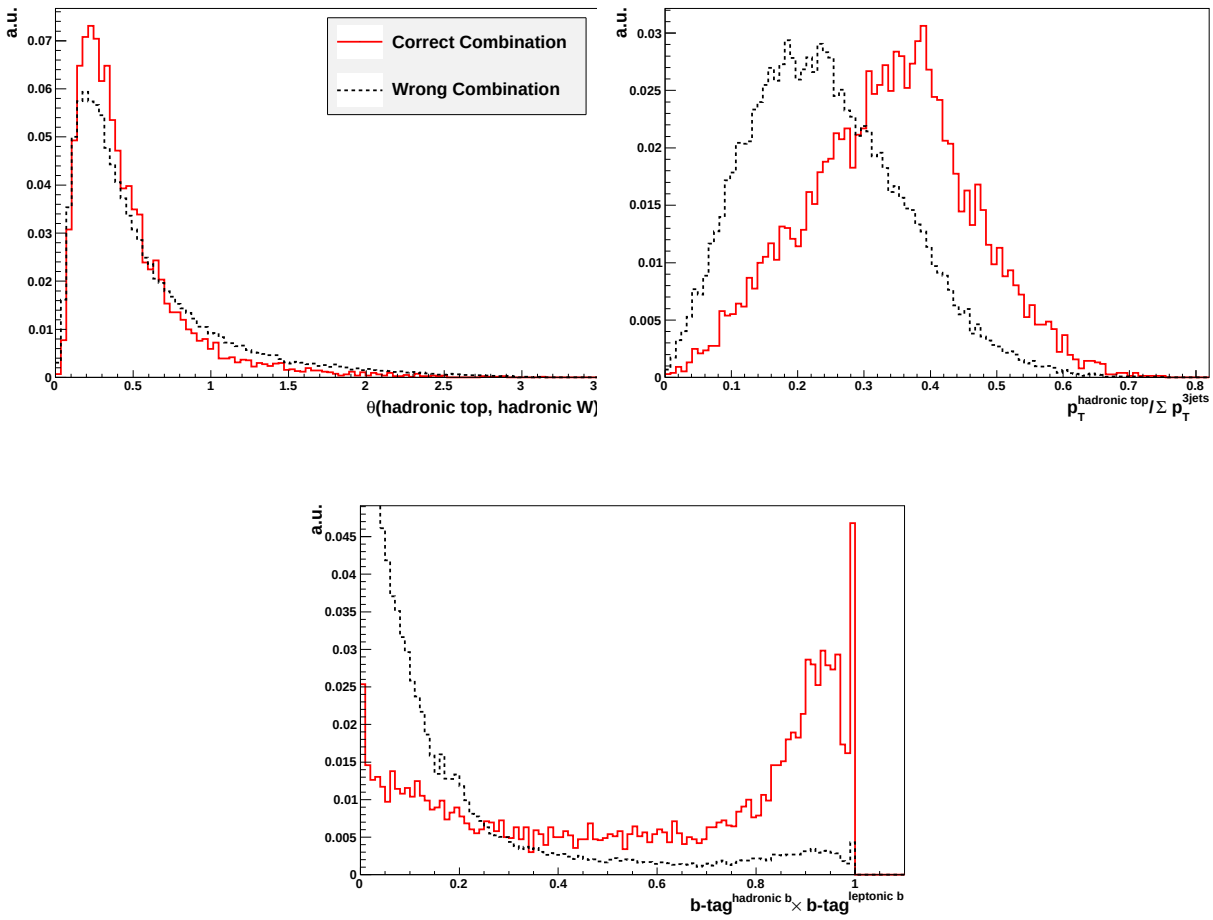

Figure 5.5: The distributions of the selected input variables to train the MVA method.

chosen jet-parton combination returned by the MVA method would still be taken as correct. Since there is no interest in the reconstruction of the leptonic part of $t\bar{t}$ system for this analysis, there is no force applied on the MVA method to associate a jet correctly to the leptonic b quark in the final state. The "Bad" jet combination then is the combination for which the MVA cannot associate, correctly, at least one hadronic quark in the $t \to Wb \to q\bar{q}b$ decay topolology, to a reconstructed jet in the four leading jets found in the event. Although in some cases the chosen jet-parton combination by the MVA, is the true combination, in about half of the cases the Likelihood Ratio method is not able to return the true jet-parton combination.

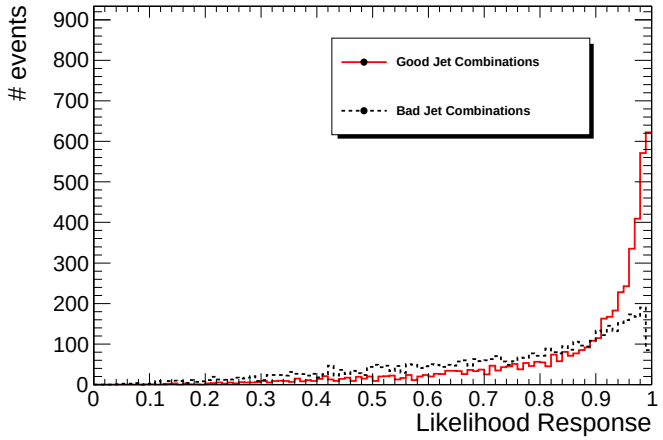

Figure 5.6: The maximum among twelve possible values returned by the Likelihood Ratio method per event. Only those selected e+jets events for which the hadronic jets are matched to the hadronic quarks in $t \to Wb \to q\bar{q}b$, are taken into account.

Considering those selected events with a correct jet-parton combination, the plot which contains the purity versus efficiency for the different cut values on the Likelihood response, is shown in Figure 5.7. Efficiency is defined as the fraction of the selected e+jets events with a right jet combination that pass the cut on the Likelihood Ratio response. Purity is defined to be the number of the selected e+jets events with a right jet combination that pass the cut on the Likelihood response relative to the multiplicity of all selected e+jets events, including either a right or wrong jet combination, that pass the cut on the Likelihood Ratio response.

According to the Monte Carlo truth, in only 34% of the selected e+jets $t\bar{t}$ events, the three reconstructed jets among the four leading jets are matched, truly, to the hadronic quarks in $t \to Wb \to q\bar{q}b$. As it is understood from Figure 5.7, the Likelihood Ratio method is able to successfully find the true combination in 51% of those events with a true jet-parton matching, given that no cut is applied on the Likelihood Ratio response. Comparing to a random selection of the true combination that yields a probability of about 8% of selecting the right jet combination, the great improvement which is obtained when using the MVA method is obvious.

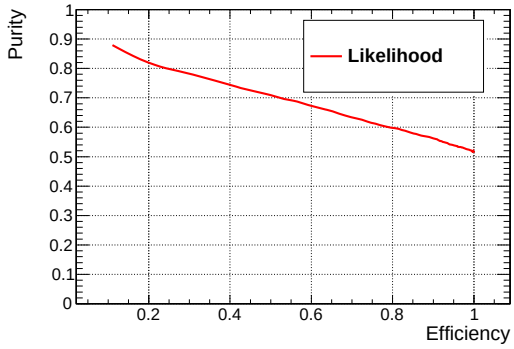

Figure 5.7: The purity versus efficiency plot when cutting on the MVA output for the selected e+jets $t\bar{t}$ events with a correct jet-parton combination.

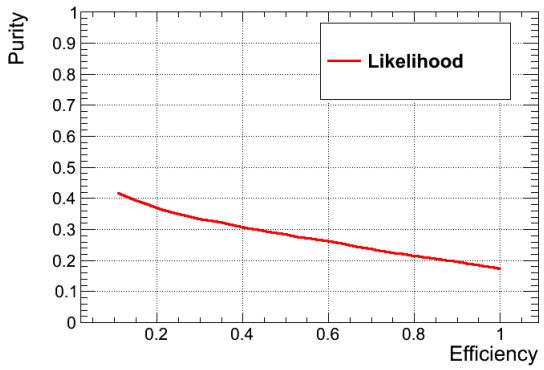

Figure 5.8: The purity versus efficiency plot when cutting on the MVA output for all selected e+jets $t\bar{t}$ events.

Also shown in Figure 5.8 is the distribution of the purity versus efficiency when cutting on the Likelihood response which is plotted for all selected e+jets $t\bar{t}$ events, including those with no correct jet-parton combination. Considering these kind of events, the efficiency of selecting the correct combination with the use of the Likelihood method, when no cut is applied on the Likelihood response, reduces to about 17% which is obtained as $34\% \times 51\% \sim 17\%$.

5.1.6 Reconstructing W Boson and Top Quark Masses

Now that the reconstructed jets are associated to the hadronic quarks which are produced in the decay products of the hadronic top quark of an e+jets $t\bar{t}$ event, $t \to Wb \to q\bar{q}b$, one can reconstruct the W boson four-vector by summing the four-vectors of the two light jets. The four-vector of the top quark can be subsequently reconstructed by adding the four-vector of the hadronic b jet to the reconstructed hadronic W boson. Furthermore, one can measure the invariant masses of the W boson and top quark from their reconstructed four-vectors. In Figure 5.9, the masses of the W boson and the top quark are shown for those events where the MVA successfully returned the true jet-parton combination. It is observed that the reconstructed W boson mass is peaked around 89 GeV while the parameter which has been used for the W boson mass when simulating the events was chosen to be 80.4 GeV. Also the mean value of the distribution of the top quark mass is shifted from the value used in the simulation procedure, being 172.5 GeV, to a reconstructed value which is found to be about 181 GeV. One should note that the events which are used to fill the mass distributions with, are those events where the MVA successfully chose the true jet combination. If all selected events were taken into account, tails would appear on both histograms.
This observation leads to the fact that, the jet energies which are used to build the W boson and top quark four-vectors are over-estimated which yields the obtained higher invariant mass values compared to the input masses. Therefore this motivates to build

a method to correct the jet energies so that the known measured values of the W boson and the top quark masses can be re-obtained.

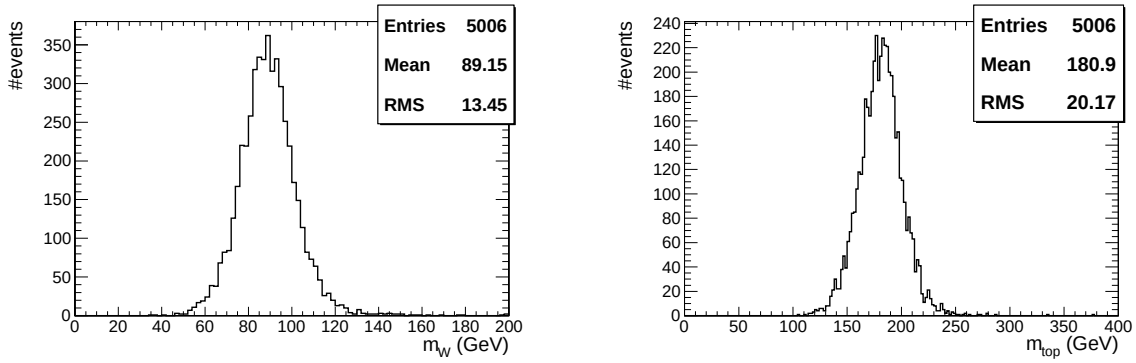

Figure 5.9: The distributions of the reconstructed W boson mass (left) and the top quark mass (right) for the selected events where a true jet-parton matching is returned by the MVA method.

5.2 Kinematic Fit

For a given event, four-vectors of the reconstructed objects are measured. The same event may contain particles, such as neutrinos, for which a direct measurement is not possible. With the assumption of certain hypotheses, the measured as well as the unmeasured quantities are supposed to fulfill the kinematic constraints which are usually deduced from the event hypothesis. Since there are uncertainties on the measured quantities, the constraints are not exactly satisfied. The constraints can then be used to slightly change the measured values within their uncertainties. The concept of an event-by-event least square fitting together with the application of Langrange Multipliers can be utilized to insure that the measured as well as the unmeasured quantities fulfill the kinematic constraints. This procedure is referred to as a kinematic fit.
The kinematic fit method has been frequently used in various physics analyses performed in different collaborations [15]. Since at an electron-positron collider, the initial energy and momentum of the incoming particles are well known, it is possible to apply constraints like energy and momentum conservation via a kinematic fit method to extract fitted values for the unmeasured quantities in the final state. Also the kinematic fit technique can be used for calibration purposes. For example, the electromagnetic calorimeter of the BaBar experiment is calibrated by means of the kinematic fit technique [102]. Using the radiative Bhabha events $e^-e^+ \rightarrow e^-e^+\gamma$, given that the momenta of the incoming and outgoing electrons and positrons, as well as the photon's angular position are measured, the photon energy can be obtained via a kinematic fit. This fit results in an absolute measurement of the photon energy which then can be compared to the measured photon energy to obtain calibration constants.
At hadron colliders like the LHC, it is not possible to apply energy-momentum constraints since the initial kinematic properties of the colliding partons arrising from

protons can not be known. However, the usage of the kinematic fit method is still feasible by applying the mass constraints in the specific event topologies. For example, in the processes where a W boson is produced and subsequently decays to a pair of quarks $W \rightarrow q\bar{q}$, applying the mass constraint of the W boson and forcing the four-momenta of its decay products to fulfill the mass equation, being

$$m_W^2 = (E_{jet_1} + E_{jet_2})^2 - (\vec{P}_{jet_1} + \vec{P}_{jet_2})^2,$$

would improve significantly the resolution on the reconstructed jets. In the current study, the mass constraints of both the W boson and top quark are applied in the hadronic branch of e+jets $t\bar{t}$ system, namely $t \rightarrow Wb \rightarrow q\bar{q}b$. The reconstructed objects in the final state, being the light jets arrising from the W boson decay and the b quark jet arrising from the top quark decay, are asked to fulfill the mass constraints which are expressed as

$$m_W^2 = (E_{l_1} + E_{l_2})^2 - (\vec{P}_{l_1} + \vec{P}_{l_2})^2,$$

$$m_{top}^2 = (E_{l_1} + E_{l_2} + E_b)^2 - (\vec{P}_{l_1} + \vec{P}_{l_2} + \vec{P}_b)^2,$$

where index l_i refers to one of the two light jets originating from light quarks, namely up, down, charm and strange. Also index b in the above formula, refers to the b jet arrising from the heavy quark, namely the bottom quark.
The kinematic fit results are used to estimate the residual jet energy correction factors for both light as well as b quark jets as will be described in detail in Section 5.4. The mathematical concept of the kinematic fit technique is reviewed in the next section.

5.2.1 General Mathematical Concept

A physics problem consists of measured quantities, $\vec{y} = (y_1, y_2, \ldots, y_n)$, and unmeasured values, $\vec{a} = (a_1, a_2, \ldots, a_p)$. Also a certain hypothesis originating from the physics principal such as conservation of the energy and momentum may be introduced. Therefore the observed events, containing n measured and p unmeasured values, are constrained to fulfill the hypothesis. Due to the presence of uncertainties on the measured values, usually the hypothesis is not respected in the physics problem which is under consideration. The hypothesis, which is expressed via constraint equations $\vec{f} = (f_1, f_2, \ldots, f_m)$, is satisfied for only the true parameters, namely $\bar{y}$ and $\bar{a}$ as written below.

$$\begin{aligned} f_1(\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_p, \bar{y}_1, \bar{y}_2, \ldots, \bar{y}_n) &= 0, \\ f_2(\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_p, \bar{y}_1, \bar{y}_2, \ldots, \bar{y}_n) &= 0, \\ &\vdots \\ f_m(\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_p, \bar{y}_1, \bar{y}_2, \ldots, \bar{y}_n) &= 0. \end{aligned}$$

In order to obtain the correction to the measured values, $\Delta\vec{y}$, which yields the constraints to be fulfilled for the corrected measurements, the extrema of the above equations should be found. Hence the Lagrange Multipliers method is used to determine the true values, $\vec{y}' = \vec{y} + \Delta\vec{y}$, for which the constraints are fulfilled. At the same

time the calculated corrections to the measured values should be minimized which is goverened with the use of a χ^2 term. Together with the Lagrange term, one could write a combined likelihood expression, $L(\vec{y},\vec{a},\vec{\lambda})$, as follows

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

where $\vec{\lambda}$ are the Lagrange Multipliers and $S(\vec{y})$ is the χ^2 term and is expressed explicitly as

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

where $\mathcal{V}$ is the covariance matrix of the measured parameters.
In order to find a local minimum for the likelihood equation, one has to minimize $S(\vec{y})$ under the constraints $f_k(\vec{y},\vec{a}) = 0$. Due to the non-linearity of the physics constraints which are used, one has to solve the constraints iteratively.
Since the constraints are fulfilled in the limit of true values, $(\vec{y}',\vec{a}')$, then $f_k(\vec{y},\vec{a})$ can be expanded arount a desired value, for example could be the value obtained during the previous iteration represented by $(\vec{y}^*,\vec{a}^*)$, as follows

$$f_k(\vec{y}',\vec{a}') \approx f_k(\vec{y}^*,\vec{a}^*) + \Sigma_{j=1}^{p}\frac{\partial f_k}{\partial a_j}.(\Delta a_j - \Delta a_j^*) + \Sigma_{i=1}^{n}\frac{\partial f_k}{\partial y_i}.(\Delta y_i - \Delta y_i^*) \approx 0,$$

where $\vec{X}$, $\vec{X}^*$ and $\vec{X}'$, with X could be either y or a, are respectively the start value, the value after the previous iteration and the value after the current iteration. Also the $\Delta\vec{X}$ and $\Delta\vec{X}^*$ are defined as $\Delta\vec{X} = \vec{X}' - \vec{X}$ and $\Delta\vec{X}^* = \vec{X}^* - \vec{X}$. One can write the kinematic constraints in vector notation

$$\vec{f}^* + \mathcal{A}(\Delta\vec{a} - \Delta\vec{a}^*) + \mathcal{B}(\Delta\vec{y} - \Delta\vec{y}^*) \approx 0,$$

or equivalently

$$\mathcal{A}\Delta\vec{a} + \mathcal{B}\Delta\vec{y} - \vec{c} = 0,$$

with

$$\vec{c} = \mathcal{A}\Delta\vec{a}^* + \mathcal{B}\Delta\vec{y}^* - \vec{f}^*; \quad \vec{f}^* = \begin{pmatrix} f_1(\vec{a}^*,\vec{y}^*) \\ f_2(\vec{a}^*,\vec{y}^*) \\ \vdots \\ f_m(\vec{a}^*,\vec{y}^*) \end{pmatrix}$$

In the vector notaion, the matrices $\mathcal{A}$ and $\mathcal{B}$ are defined as $\mathcal{A} = \frac{\partial\vec{f}}{\partial\vec{a}}$ and $\mathcal{B} = \frac{\partial\vec{f}}{\partial\vec{y}}$ for which their components are expressed explicitly below.

$$\mathcal{A} = \begin{pmatrix} \frac{\partial f_1}{\partial a_1} & \frac{\partial f_1}{\partial a_2} & \cdots & \frac{\partial f_1}{\partial a_p} \\ \frac{\partial f_2}{\partial a_1} & \frac{\partial f_2}{\partial a_2} & \cdots & \frac{\partial f_2}{\partial a_p} \\ \vdots & & & \\ \frac{\partial f_m}{\partial a_1} & \frac{\partial f_m}{\partial a_2} & \cdots & \frac{\partial f_m}{\partial a_p} \end{pmatrix}$$

$$\mathcal{B} = \begin{pmatrix} \frac{\partial f_1}{\partial y_1} & \frac{\partial f_1}{\partial y_2} & \cdots & \frac{\partial f_1}{\partial y_n} \\ \frac{\partial f_2}{\partial y_1} & \frac{\partial f_2}{\partial y_2} & \cdots & \frac{\partial f_2}{\partial y_n} \\ \vdots & & & \\ \frac{\partial f_m}{\partial y_1} & \frac{\partial f_m}{\partial y_2} & \cdots & \frac{\partial f_m}{\partial y_n} \end{pmatrix}$$

All of the partial derviatives of the constraints, in each step of the iteration, need to be obtained at the value of the parameters at the previous iteration X^*. Since the constraints depend directly on the four-vector components $\vec{P} = (\vec{p}, E)$, the chain rule is used to obtain the partial derivatives $\frac{\partial \vec{f}}{\partial \vec{y}} = \frac{\partial \vec{f}}{\partial \vec{P}} . \frac{\partial \vec{P}}{\partial \vec{y}}$. As four-momenta of the particles can be parametrized in different ways, the definition of the $\frac{\partial \vec{P}}{\partial \vec{y}}$ is not unique. Various parametrization for the reconstructed objects exist which are described in more detail in the next section.
In the new notation, one has to minimize the likelihood expression which is written as

$$L = \Delta\vec{y}^T \mathcal{V}^{-1} \Delta\vec{y} + 2\lambda^T (\mathcal{A}\Delta\vec{a} + \mathcal{B}\Delta\vec{y} - \vec{c}).$$

The conditions for a local minimum are obtained, after differentiating the above expression with respect to $\vec{y}$, $\vec{a}$ and $\vec{\lambda}$, as listed below

$$\mathcal{V}^{-1}\Delta\vec{y} + \mathcal{B}^T\vec{\lambda} = 0,$$

$$\mathcal{A}^T\vec{\lambda} = 0,$$

$$\mathcal{B}\Delta\vec{y} + \mathcal{A}\Delta\vec{a} = c.$$

Collecting the above three equations in a compact form with partitioned matrices

$$\begin{pmatrix} \mathcal{V}^{-1} & 0 & \mathcal{B}^T \\ 0 & 0 & \mathcal{A}^T \\ \mathcal{B} & \mathcal{A} & 0 \end{pmatrix} \begin{pmatrix} \Delta\vec{y} \\ \Delta\vec{a} \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ c \end{pmatrix}$$

This matrix equation will be solved iteratively for the unknown values $\Delta\vec{y}$, $\Delta\vec{a}$ and $\Delta\vec{\lambda}$.
The iteration procedure is repeated until some predefined convergence criteria are fulfilled. The first one requires that the change in χ^2 expression between the current iteration, n, and the previous iteration, n-1, is smaller than a given value ϵ_S, and is written as

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

where *ndf* is the difference between the number of constraints and the number of unmeasured quantities, being

$$ndf = m - p.$$

The second convergency criterion which is checked in each iteration, requires that the constraints are fulfilled better than a given value ϵ_F, and is expressed as

$$F = \Sigma_{k=1}^{m} f_k^{(n)}(\vec{y}, \vec{a}) < \epsilon_F.$$

A more complete description of the kinematic fit technique can be found in [103].

5.2.2 Four-Vector Parametrization

Several parametrizations in the kinematic fit package are implemented, two of which are commonly used in hadron colliders and are introduced briefly here. A first one, called `TFitParticleEtThetaPhi`, uses E_T, θ and ϕ variables for parametrization. The momentum vector and the energy can be written in this parametrization as

$$\vec{p}_f = \begin{pmatrix} E_T \cos\phi \\ E_T \sin\phi \\ E_T \cot\theta \end{pmatrix},$$

$$E = \frac{E_T}{\sin\theta}.$$

In a second parametrization, called `TFitParticleEtEtaPhi`, where E_T, η and ϕ variables are used as parameters, the four-vector can be expressed as

$$\vec{p}_f = \begin{pmatrix} E_T \cos\phi \\ E_T \sin\phi \\ E_T \sinh\eta \end{pmatrix},$$

$$E = E_T \cosh\eta.$$

In the two parametrizations explained above, the objects are considered as massless. In this study, `TFitParticleEtThetaPhi` parametrization is used with an additional requirement. During the fit procedure, it is crucial to keep the ratio of the measured energy of a jet to its measured momentum constant. Hence, in the kinematic fit method which is used in the current study, this additional requirement is applied which is expressed as follows

$$\frac{E_{fitted}}{|\vec{p}_{fitted}|} = \frac{E_{measured}}{|\vec{p}_{measured}|}.$$

5.2.3 Resolutions on the Jet Kinematics

The uncertainties on the measured parameters enter the kinematic fit package via the covariance matrix, as mentioned before. In this analysis where the measured objects are reconstructed jets, the resolutions on the jet properties are required to be calculated. These resolutions are obtained from simulated e+jets $t\bar{t}$ events in which a correct jet-parton combination, according to the generator level information, is found. According to the jet-parton matching algorithm described in Section 5.1.1, the reconstructed jets are matched to the partons if ΔR difference between them is less than 0.3. The reconstructed jets are grouped in either light jets, which are matched to the partons arising from the decay of the W boson, or b jets, which are matched to the b flavored partons originating from the decay of the top quarks. Since the obtained resolutions are found to be dependent on the transverse momentum and the pseudorapidity of the reconstructed jets, jets are further categorized in various bins of p_T and η. In this analysis, eleven bins of transverse momentum and twelve bins of pseudorapidity are considered. For those jets which populate a certain p_T-bin and η-bin, the resolutions are calculated as explained in the following. The resolution on the X property of a

jet is computed as the width of the Gaussian function fitted on the distribution of the difference between the value of the X parameter of the reconstructed jet and the value of the X parameter of the matched parton to that jet. The resolutions on the various kinematics of the jets, namely transverse energy E_T, polar angle θ and azimuthal angle ϕ as a function of the transverse momentum of the jets are shown in Figure 5.10 and Figure 5.11, for the light quark and the b quark jets, respectively. In each figure, the left plots represent the typical jets reconstructed in the barrel while the right plots are obtained for typical jets reconstructed in a narrow region in the endcaps. The obtained resolution plots are fitted with functions which are specified on the plot. For example, the distribution of the resolution on the transverse energy of the jets is fitted with a function which is parametrized as $a + b\sqrt{p_T} + cp_T$. The free parameters of the function are determined from the fit and quoted on each plot.

From Figure 5.10, it can be seen that the absolute resolution on the transverse energy for jets with higher transverse momenta is worse, while the relative resolution is worse for the jets with lower transverse momenta, as discussed in Section 3.2.4. Also it can be deduced that the resolution on the angular distributions is better determined compared to the resolution on the transverse energy. This is because of the matching requirement applied between the reconstructed jets and the partons in $(\eta \times \phi)$ space. The resolutions on the spatial coordinates of the reconstructed jets are decreased for jets with higher transverse momenta, because the jets with higher transverse momenta are less affected by the magnetic field. The polar angle of reconstructed jets can be determined with higher precision compared to the azimuthal angle. This is due to the bending of charged particle in the direction of ϕ. Comparing with the plots in Figure 5.11, it can be understood that the resolution on the transverse energy of b quark jets is worse compared to the resolution on the transverse energy of light quark jets. This is due to the presence of muons or neutrinos in the decay products of b quark jets which yields to cluster less energy for b quark jets as muons and neutrinos are weakly interacting particles. There is no big difference observed when comparing the resolutions on the spatial kinematics of b quark and non-b quark jets.

5.3 Extra Event Selection Cuts

So far, events have been required to pass the basic event selection cuts that are summarized in Table 4.6. The selected events are subsequently fed to the MVA package. The output of the MVA method is a matching between the reconstructed jets in the final state of the e+jets $t\bar{t}$ event and the partons appearing in the hadronic branch of the $t\bar{t}$ system $t \to Wb \to q\bar{q}b$, as described in Section 5.1. The chosen jet-parton combination returned by the MVA method, can be either right or wrong. As already mentioned, the MVA method which has been trained using the variables listed in Section 5.1.4, is not able to choose the correct jet-parton matching in about 50% of the cases. This is partly due to the ISR/FSR effects which makes it possible to get the radiation jets among the four leading jets and then at least one of the four leading jets can not be matched to the hard-scatter partons. The e+jets $t\bar{t}$ events, for which a wrong jet combination is recognized by the MVA or the event itself is badly reconstructed, contribute to the

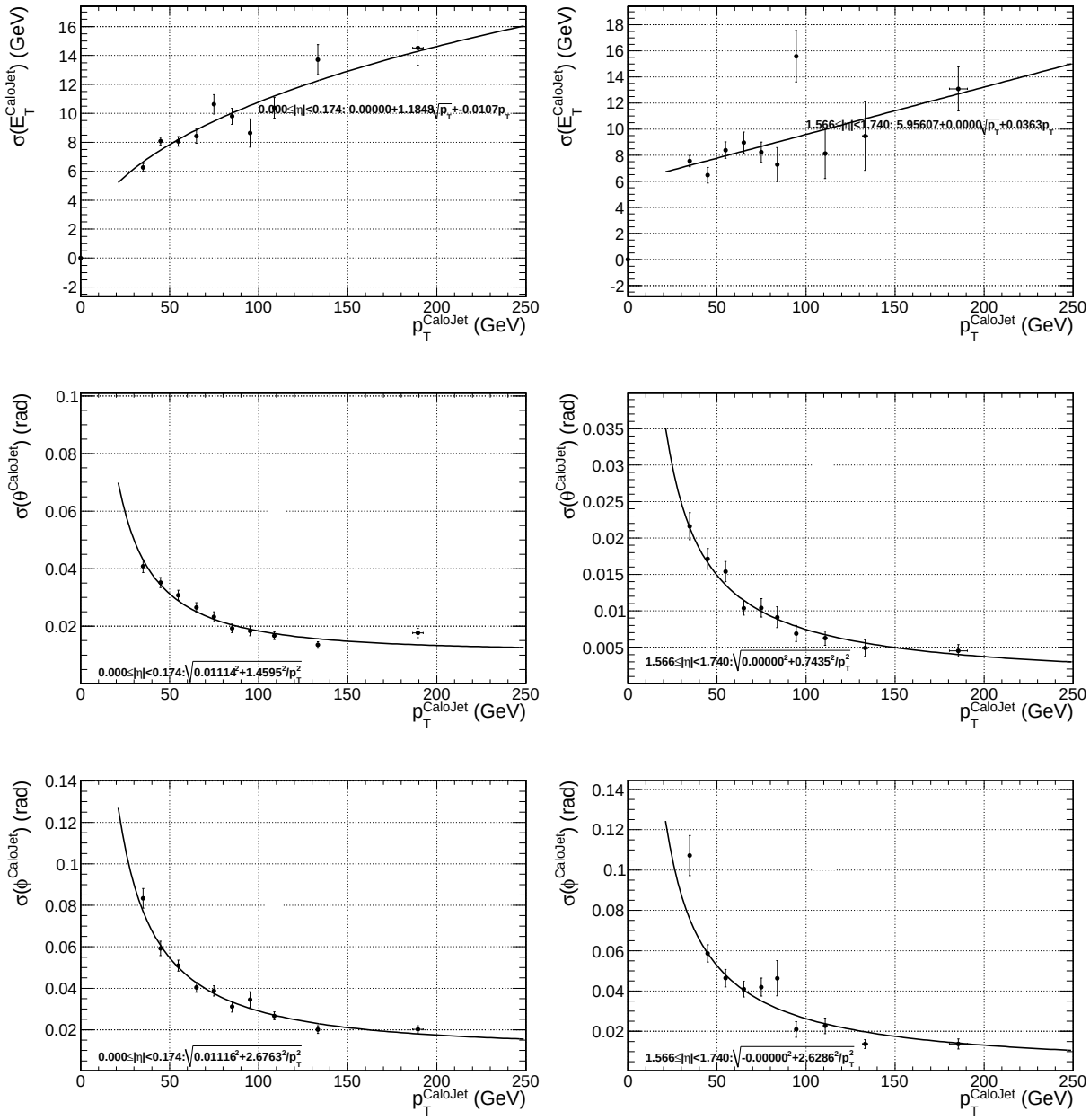

Figure 5.10: The distributions of the resolutions on the E_T (top row), θ (middle row) and ϕ (bottom row) of the light quark jets shown for two typical bins of pseudorapidity of the jets, being $0 < |\eta| < 0.174$ (left column) and $1.566 < |\eta| < 1.740$ (right column). Also shown is the fit function imposed on each distribution.

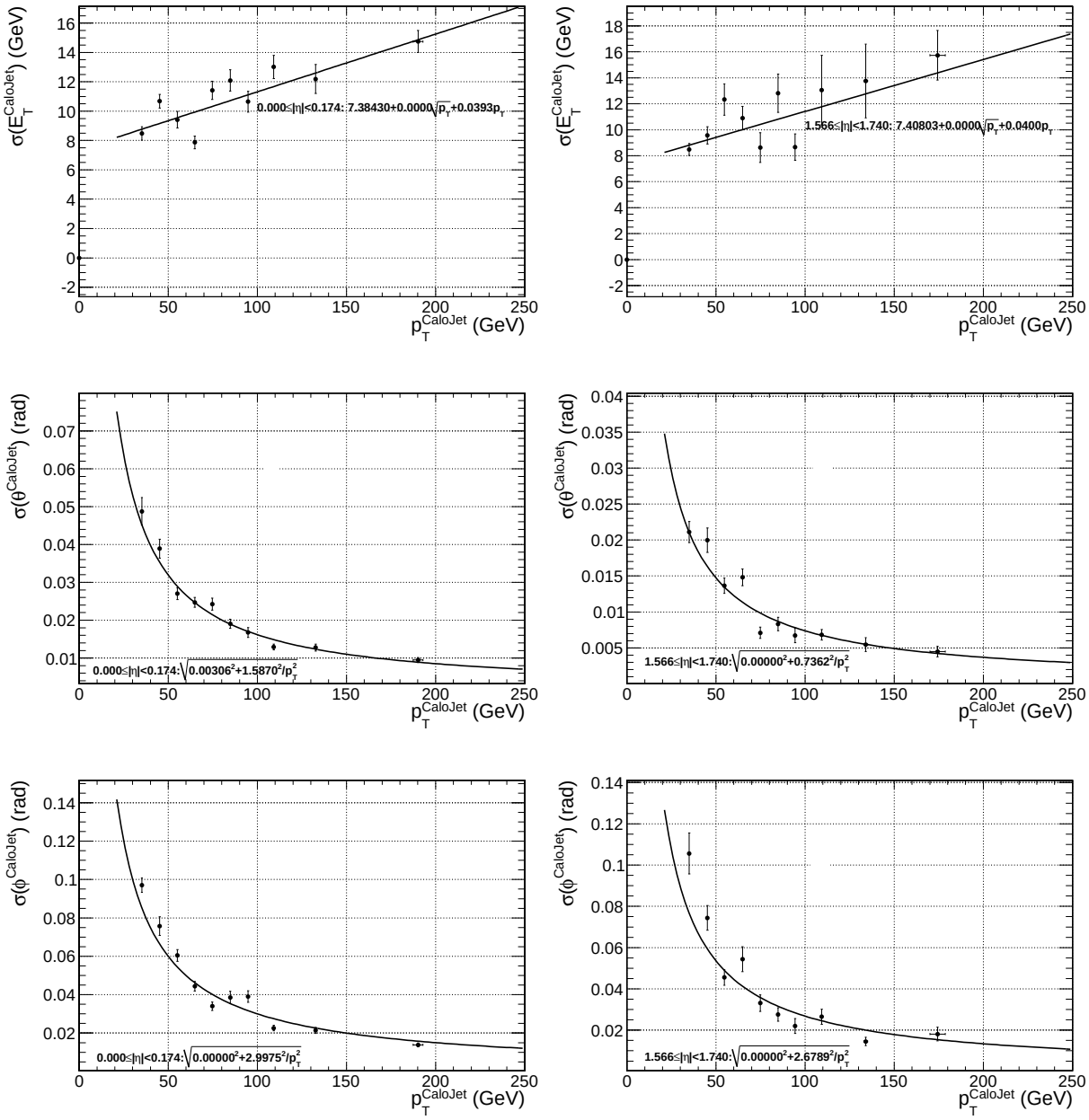

Figure 5.11: The distributions of the resolutions on the E_T (top row), θ (middle row) and ϕ (bottom row) of the b quark jets shown for two typical bins of pseudorapidity of the jets, being $0 < |\eta| < 0.174$ (left column) and $1.566 < |\eta| < 1.740$ (right column). Also shown is the fit function imposed on each distribution.

combinatorial backgrounds. In order to reduce the contribution of the combinatorial backgrounds, additional event selection cuts are applied.
In the topology of the e+jets $t\bar{t}$ event, there are four quarks produced in the final state which subsequently contribute to at least four reconstructed jets in the detector. In addition to the four jets which represent the quarks generated in the hard scattering process, any other jet which is reconstructed in the event is the consequence of the radiation of a parton in the initial or final state, generally known as ISR or FSR effect, respectively. As already discussed in Section 5.1.1, most of the e+jets $t\bar{t}$ events, contain at least an ISR jet which is among the four leading reconstructed jets. Then these kind of events, where the four leading jets can not be matched to the four hard-scatter partons, would contribute to the combinatorial backgrounds. Since there is no true jet-parton matching in such events, the MVA method would always return a wrong jet combination. Therefore, any event with at least one extra jet is vetoed and is not processed in the rest of analysis. In addition to the four leading jets in the selected events, any other jet with $p_T > 30\ GeV$ within the tracker acceptance is defined as a radiation jet and the events containing at least one radiation are rejected.

In order to further reduce the multiplicity of badly reconstructed events in the final selected sample, the events are constrained to fulfill some kinematic requirements. All events which pass the radiation veto cut, are fed to the kinematic fit package. The three reconstructed jets in the e+jets $t\bar{t}$ event, which are now associated to the partons in the hadronic branch $t \to Wb \to q\bar{q}b$ by means of the MVA method, are forced to fulfill the mass constraints as described in Section 5.2. The masses of the W boson and top quark used in the fit package are taken to be equal to those used in the simulation step, being 80.4 GeV and 172.5 GeV respectively. The energy of the three jets associated to the hadronic top quark are shifted in certain steps. A three-dimensional space, which is spanned by the changes in the energies of the three reconstructed jets $(\Delta E_{l_1}, \Delta E_{l_2}, \Delta E_b)$, is obtained, where ΔE_{l_1} and ΔE_{l_2} represent the correction factors on the light jets from the hadronic W boson decay and ΔE_b corresponds to the changes in the energy of the b jet arrising from the hadronic top quark decay. In each point in the three-dimensional space, the kinematic fit is performed. The fit is applied on a grid in this space in steps of 2% within a window of ±40% around the reconstructed energies of both light as well as b jets, individually. As a result, the energy of a generic jet with, for example, E=100 GeV is shifted from E=60 GeV to E=140 GeV in steps of 2 GeV and in each step the fit is performed, hence 41 times per jet per event. Since the light jets arrising from the W boson are indistinguishable, their correction factors can be assumed to be equal $\Delta E_{l_1} = \Delta E_{l_2} = \Delta E_l$. Therefore the kinematic fit is performed in a two-dimensional space which is obtained by shifting in the energy of the light jets, for which the shifts are taken to be identical, and in the energy of the b jet $(\Delta E_l, \Delta E_b)$. As a result, the kinematic fit is performed 41×41 times per event.
In each point of the two-dimensional grid, the kinematic fit returns a probability $P_{KinFit}(\Delta E_l, \Delta E_b)$, which represents how likely, for that particular point of the grid, the hypothesis of originating the three jets in e+jets $t\bar{t}$ event from the decay of the top quark in the topology $t \to Wb \to q\bar{q}b$, is correct. At the same time, the fit probability states how much the mass hypotheses are correct for the three jets in e+jets $t\bar{t}$ event

from the decay of the top quark in the topology $t \to Wb \to q\bar{q}b$. For a typical event, the distribution of the $P_{KinFit}(\Delta E_l, \Delta E_b)$ is shown in Figure 5.12.

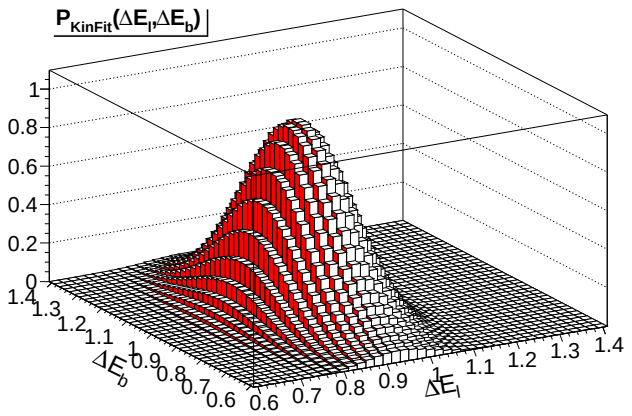

Figure 5.12: The distribution of the $P_{KinFit}(\Delta E_l, \Delta E_b)$ for an e+jets $t\bar{t}$ event.

The central point of the distribution of the $P_{KinFit}(\Delta E_l, \Delta E_b)$, which is expressed by $P_{KinFit}(\Delta E_l = 1, \Delta E_b = 1)$, corresponds to the situation when no correction is applied. It is understood from Figure 5.12 that the maximum of the fit, for that particular event, does not happen at the center of the grid, which shows the need for the residual jet energy corrections. In order to justify the need for the residual jet energy corrections, one has to increase the statistics and may look at the distribution of the values of the $P_{KinFit}(\Delta E_l = 1, \Delta E_b = 1)$, which is shown in Figure 5.13.

There is a clear peak at zero which can be partially interpreted as follows. The hypothesis that the three jets in e+jets $t\bar{t}$ event originate from the top quark is not fulfilled and the imposed mass constraints are not converged, when no correction is applied. This observation justify the nead for the residual jet energy correction. Also the events, where the MVA method is not able to return the correct jet-parton combination, contribute to the peak at zero as will be discussed in what follows.
For a given event, the point with the maximum probability in the distribution of the $P_{KinFit}(\Delta E_l, \Delta E_b)$, which is expressed by $P^{max}_{KinFit}(\Delta E_l, \Delta E_b)$, shows the point where the imposed constraints are maximally fulfilled. The distribution of the $P^{max}_{KinFit}(\Delta E_l, \Delta E_b)$ value for all events that already pass the radiation veto cut, is shown in Figure 5.14.

It is seen from Figure 5.14 that for most of the events, a maximum probability exceeding 0.9 is returned by the kinematic fit method, representing those events for which the maximum happens somewhere in the region that is scanned. Some of the events are peaked around zero which represent those events where no maximum is found in the window of ±40% around the measured values of the jet energies or events that the fit is not converged for the given jet-parton combination. This can be understood by looking at the two-dimensional histogram which is made by plotting the distribution

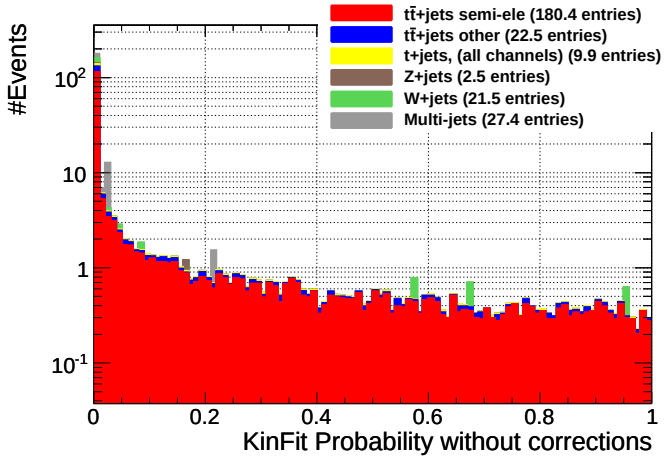

Figure 5.13: The distribution of the fit probability $P_{KinFit}(\Delta E_l = 1, \Delta E_b = 1)$ when no correction is applied. All events surviving the radiation veto cut, are taken into account.

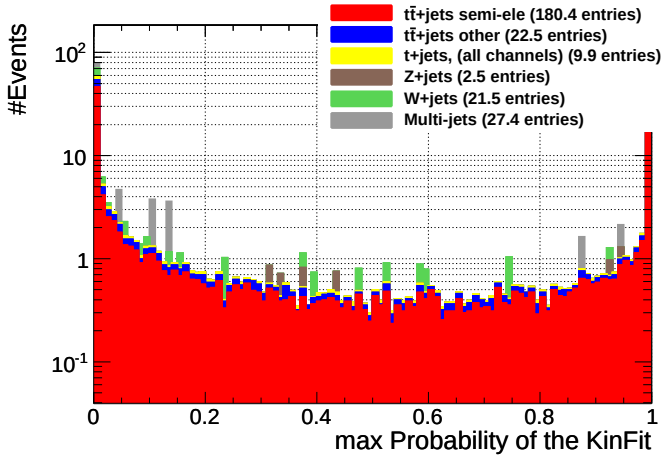

Figure 5.14: The distribution of the maximum fit probability $P^{max}_{KinFit}(\Delta E_l, \Delta E_b)$ among the fit values which are obtained in the scanned range around the uncorrected energies. All events survived the radiation veto cut, are taken into account.

of $P_{KinFit}(\Delta E_l = 1, \Delta E_b = 1)$ versus $P^{max}_{KinFit}(\Delta E_l, \Delta E_b)$ as shown in Figure 5.15.

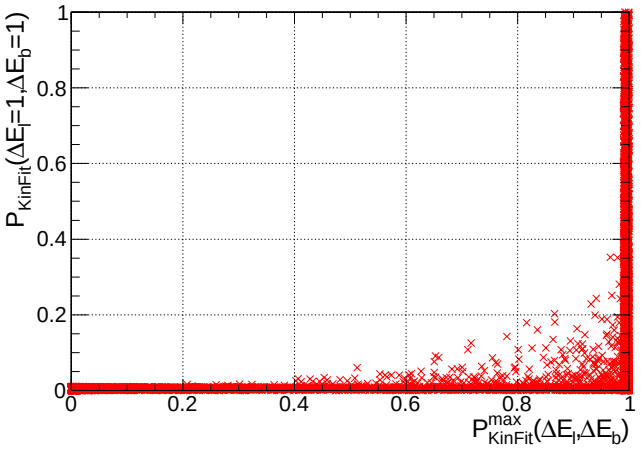

Figure 5.15: The distribution of $P_{KinFit}(\Delta E_l = 1, \Delta E_b = 1)$ versus $P^{max}_{KinFit}(\Delta E_l, \Delta E_b)$. Only e+jets $t\bar{t}$ events survived the radiation veto cut, are taken into account.

It is clear from Figure 5.15 that for the events with a maximum fit probability below 0.9, the probability returned by the fit method when no correction is applied, is populated around zero. This means that the jets are assigned wrongly to the hard scatter partons by the MVA method or the event is not correctly reconstructed due to the radiations.
Each selected event is required to have a P^{max}_{KinFit} exceeding 0.98. This cut aims to keep only those events for which the hypothesis is correct and the maximum occurs somewhere in the space which is being looked at around the uncorrected energies of the jets. No cut is applied on the $P_{KinFit}(\Delta E_l = 1, \Delta E_b = 1)$, since cutting on this parameter yields to reject those events where a maximum is found in the window of the residual correction and they have to be used for the final estimation of the jet corrections. In addition to the radiation veto cut and cutting on the $P^{max}_{KinFit}(\Delta E_l, \Delta E_b)$ variable, a third one which is the last in the sequence of extra event selection cuts, is introduced. There are some events for which the fit procedure does not converge in some points in the space of $P_{KinFit}(\Delta E_l, \Delta E_b)$. It is desired to have a fit probability which is smoothly distributed. Therefore, the events which contain at least one point in the grid where the kinematic fit does not converge, do not contribute in the rest of analysis.
The additional analysis specific cuts, which are applied on top of the basic event selection criteria as already introduced in Table 4.6, are summarized in Table 5.3 and the number of events passing each step are also mentioned individually for signal and relevant backgrounds.

It can be understood from Table 5.3 that the $t\bar{t} \rightarrow$ other events constitute the dominant background to e+jets $t\bar{t}$ signal events. Since the simulated sample of the QCD multijet events are statistically limited, the uncertainty on the 9.4 events is large corresponding to about 8.6 events. Therefore, the $t\bar{t} \rightarrow$ other is still the dominant background to the analysis.

	$t\bar{t}$ $\rightarrow e + jets$	$t\bar{t}$ $\rightarrow$ other	W+jets $\rightarrow l + jets$	Z+jets $\rightarrow l^+l^- + jets$	t $\rightarrow l + jets$	QCD multijets
Selection Table 4.6	307.8	39.1	30.2	4.5	13.5	35.6
Radiation Veto	180.4	22.5	21.4	2.5	9.9	27.4
P^{max}_{KinFit}	75.2	6.3	3.4	0.3	2.5	9.4
Gap Veto	66.8	5.5	2.8	0.3	2.2	9.4

Table 5.3: Table of extra event selection cuts. The numbers are equivalent to an integrated luminosity of 100 pb^{-1}.

Having applied the extra event selection cuts given in Table 5.3, one could reach a signal over background ratio of about 3.3 which yields an even more purified final sample of selected events compared to the baseline event selection cuts. It is worth to note that in a fraction of selected events, surviving the baseline cuts, which corresponds to 17%, the MVA method is able to return the correct jet-parton combination while after applying the additional event selection cuts, this fraction is enhanced to 52%. This means that the extra event selection cuts are able to reject the badly reconstructed events yields to suppress the combinatorial backgrounds.
Those events which are survived the additional selection requirements, are accepted for the final analysis.

5.4 Extracting Jet Energy Scale Calibration Factors

The estimation of the residual jet energy calibration factors is based on the output of the kinematic fit. For each selected event, the kinematic fit technique returns a two-dimensional fit probability $P_{KinFit}(\Delta E_l, \Delta E_b)$ in the dimensions of the residual light and b jet energy corrections, as explained in the previous section. The jet energy correction factors are then determinded by maximizing the fit probabilty returned by the fit procedure as described below.
The kinematic fit method is applied on the three reconstructed jets in e+jets $t\bar{t}$ event. The jets, which were associated to the three quarks originating from the hadronic branch of the top quark decay $t \rightarrow Wb \rightarrow q\bar{q}b$ by means of the MVA method, are forced to fulfill the mass constraints imposed on the system. The four-vectors of the three reconstructed jets, which are fed to the fit package, are parametrized using the `TFitParticleEtThetaPhi` parametrization for which the ratio of the energy over momentum $\frac{E}{|p|}$ is kept constant and equal to the measured value during the fit procedure, as already explained in Section 5.2.1. The parameters of the fit are chosen as follows. The parameter ϵ_S is set to be 5×10^{-5}. Also it is chosen to have $\epsilon_F = 10^{-4}$. The iteration procedure is repeated 30 times. If the fit is not converged after 30 times, the fitting is stopped and a value equal to zero is assigned to the probability of the fit for

that point in the two-dimensional space of $(\Delta E_l, \Delta E_b)$. These events are not selected as mentioned before.
The correction factors on the light jet energy ΔE_l, and on the b quark jet energy ΔE_b, can be interpreted equivalently as the residual shifts on the light jet energy $\Delta\varepsilon_l$, and on the b quark jet energy $\Delta\varepsilon_b$, by means of

$$\Delta E_l \equiv (1 + \Delta\varepsilon_l),$$

and

$$\Delta E_b \equiv (1 + \Delta\varepsilon_b).$$

Therefore, in each point of the two-dimensional grid $(\Delta E_{li}, \Delta E_{bj})$, the shifts on the light and the b quark jet energy are applied by means of multiplying the following factors

$$(1 + \Delta\varepsilon_{li}),$$

and

$$(1 + \Delta\varepsilon_{bj}),$$

by the initial measured light jet energy, called $E_{l_{init}}^{meas}$, and the initial measured b jet energy, called $E_{b_{init}}^{meas}$, respectively. In the multiplication factors, $0 \leq i, j < 41$ and the start points are taken to be

$$\Delta\varepsilon_{l(i=0)} = -0.4,$$

and

$$\Delta\varepsilon_{b(j=0)} = -0.4.$$

Hence, the region which is scanned by the fit procedure includes the following points

$$0.60E_{l_{init}}^{meas}, 0.62E_{l_{init}}^{meas}, \ldots (1 + \Delta\varepsilon_{li})E_{l_{init}}^{meas} \ldots, 1.38E_{l_{init}}^{meas}, 1.40E_{l_{init}}^{meas};$$

$$0.60E_{b_{init}}^{meas}, 0.62E_{b_{init}}^{meas}, \ldots (1 + \Delta\varepsilon_{bj})E_{b_{init}}^{meas} \ldots, 1.38E_{b_{init}}^{meas}, 1.40E_{b_{init}}^{meas}.$$

In each point of the scanned region, the kinematic fit package returns a probability which is obtained from the χ^2 of the fit. Therefore, instead of a two-dimensional distribution of the fit probability $P_{KinFit}(\Delta E_l, \Delta E_b)$, one can have a two-dimensional distribution of the χ^2 values $\chi^2(\Delta E_l, \Delta E_b)$. In each point of the two-dimensional grid, the χ^2 values of all selected events are summed which leads to obtain a two-dimensional parabola in the dimensions of the residual light and b jet energy corrections. The distribution of the $\chi^2(\Delta E_l, \Delta E_b)$ which is obtained by the sum of the χ^2 disribution of the individual selected event, is shown in Figure 5.16.

The estimated residual light and b jet energy correction factors are determined at the point with the minimum χ^2 fit probability, being the most central point in the plot shown in Figure 5.16.
The two-dimensional χ^2 distribution is projected in the directions of ΔE_l and ΔE_b to obtain two separate one-dimensional distributions which are subsequently used to estimate the final residual corrections. The projection into the ΔE_l axis is done using those events that participate in minimizing the χ^2 distribution in the direction of ΔE_b. Also, the events which contribute to the bin of the minimum χ^2 distribution in the

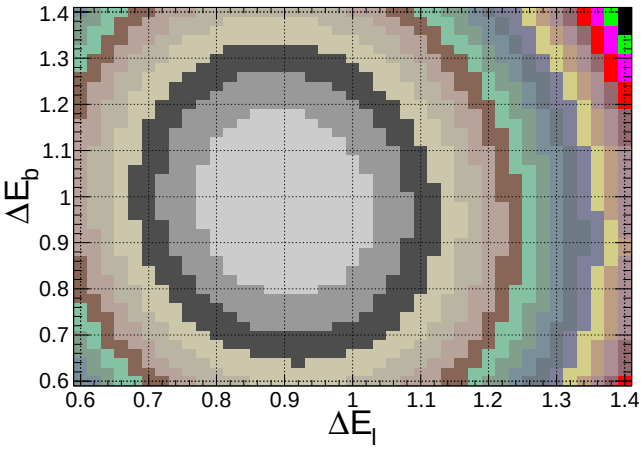

Figure 5.16: The distribution of the $\chi^2(\Delta E_l, \Delta E_b)$ which is obtained by the sum of the χ^2 distributions for all selected events.

direction of ΔE_l, are taken into account to make the projected distribution in the direction of ΔE_b axis.
Each of the two projected distributions is fitted with a second-degree polynomial around the minimum as shown in Figure 5.17.

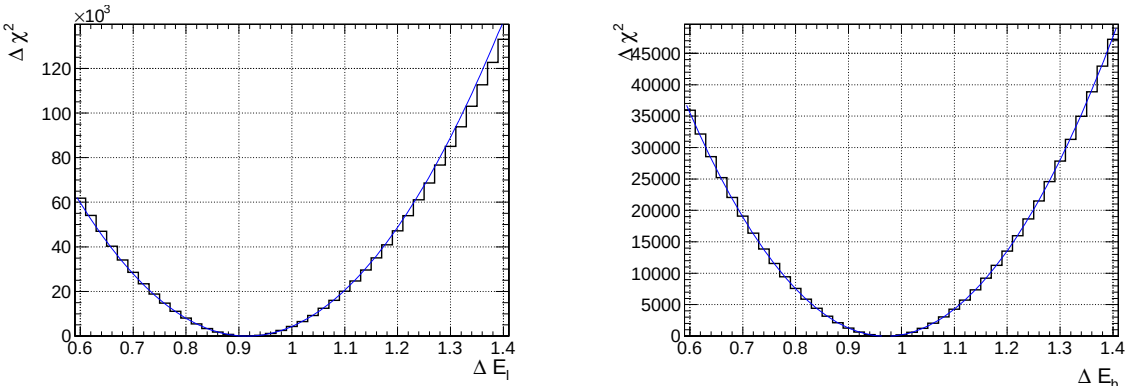

Figure 5.17: The projected distribution of the $\chi^2(\Delta E_l, \Delta E_b)$ into the (left) ΔE_l and (right) ΔE_b axis which are fitted with a second-degree polynomial. The results are obtained for 100 pb^{-1} of simulated proton collisions.

The parameters of the fitted parabolas are used to extract the estimated light as well as the b jet energy corrections, expressed as ${\Delta\varepsilon_l}^{est}$ and ${\Delta\varepsilon_b}^{est}$, respectively. The minimum of the parabolas refer to the final estimation of the jet energy corrections which are found to be

$$ {\Delta\varepsilon_l}^{est} = -10.12 \pm 1.11\%, $$

and

$$ {\Delta\varepsilon_b}^{est} = -1.31 \pm 1.72\%, $$

where the quoted uncertainties are the statistical uncertainties and correspond to the deviation of 1σ around the minimum.
The 5σ contour around the minimum of the $\chi^2(\Delta E_l, \Delta E_b)$ is obtained and shown in Figure 5.18. There is no significant correlation between light and b jet energy calibration factors observed.

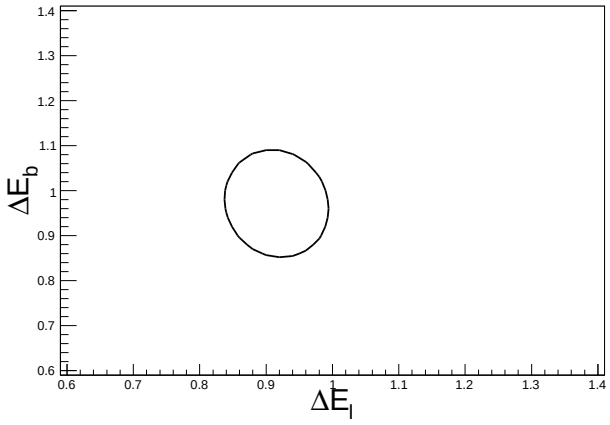

Figure 5.18: The 5σ contour around the minimum of the $\chi^2(\Delta E_l, \Delta E_b)$. The plot is obtained for an integrated luminosity of 100 pb^{-1}.

The quoted statistical uncertainties on the estimated jet energy corrections are calculated using a number of events corresponding to an integrated luminosity of about 9400 pb^{-1} but are rescaled to an integrated luminosity of 100 pb^{-1}. Then the number of simulated events are indeed more than what is expected in 100 pb^{-1} of accumulated data. Therefore, the statistical uncertainties are actually less than what have been quoted. The statistical uncertainties corresponding to an integrated luminosity of about 9400 pb^{-1} are evaluated and reported as the uncertainties on the estimated light and b jet energy corrections as follows

$$\Delta\varepsilon_l{}^{est} = -10.12 \pm 0.13\%,$$

and

$$\Delta\varepsilon_b{}^{est} = -1.31 \pm 0.20\%.$$

The various sources for the systematical uncertainties on the estimated results are discussed in the Section 5.4.2.
In a study based on the simulated data, the residual jet energy corrections can be determinded using the Monte Carlo truth information, too. A comparison between the expected residual corrections from Monte Carlo truth information and the estimated residual corrections derived from the kinematic fit procedure, can then demonstrate how well the estimation method works. If the four-vectors of the partons can be accessed from the simulation information, the reconstructed jets can be matched to the partons

using a jet-parton matching algorithm, as described in Section 5.1.1. Therefore for a pair of matched jet-parton, the expected residual correction can be obtained by

$$E_{jet}(1 + \Delta\varepsilon_{jet}^{exp}) = E_{parton}.$$

Hence, in the topology of the hadronic top quark decay $t \rightarrow Wb \rightarrow q\bar{q}b$, one can calculate the expected corrections for the light and b jets, as

$$\Delta\varepsilon_{l}^{exp} = \frac{E_{light\ quark} - E_{light\ jet}}{E_{light\ jet}},$$

and

$$\Delta\varepsilon_{b}^{exp} = \frac{E_{b\ quark} - E_{b\ jet}}{E_{b\ jet}}.$$

The distributions of the expected light as well as the b jet energy corrections, which are fitted with a Gaussian function in a range of ± 1.5 times the Root Mean Square (RMS) of the distribution around the mean, are shown in Figure 5.19. They are obtained using those selected events for which a true jet-parton matching exists.

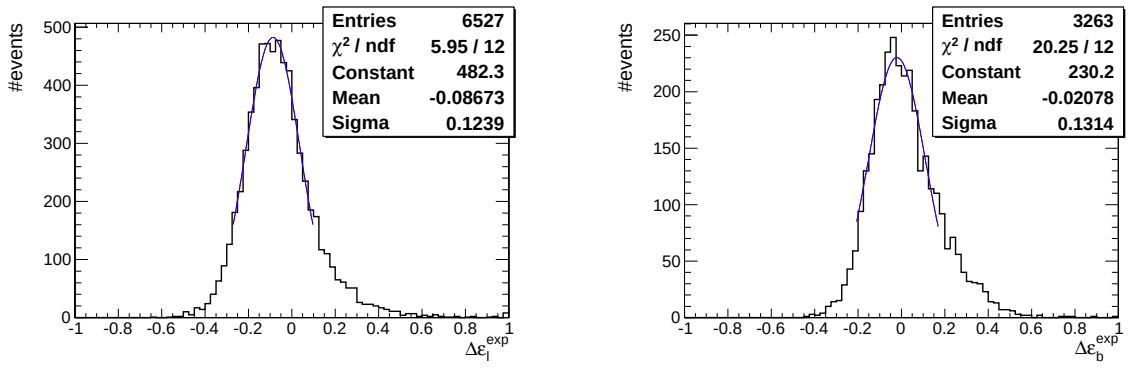

Figure 5.19: The distributions of the expected residual jet energy corrections for (left) light and (right) b jets. They are fitted with a Gaussian in the range of ± 1.5 times the RMS around the mean.

For each of the plots shown in Figure 5.19, the expectation value of the fitted Gaussian and its uncertainty are taken as the expected residual jet energy corrections which are found to be

$$\Delta\varepsilon_{l}{}^{exp} = -8.67 \pm 0.23\%,$$

and

$$\Delta\varepsilon_{b}{}^{exp} = -2.08 \pm 0.36\%.$$

The fit results strongly depend on the range which is used to perform the Gaussian fit. For example, the expected residual corrections, obtained for various ranges of the fit, are listed in Table 5.4.

As seen from Table 5.4, the expectation values of the fitted Gaussian functions for the light quark jets varies from -8.83% to -7.43% when increasing the fit range from

	±1.0 RMS	±1.5 RMS	±2.0 RMS	±3.0 RMS
$\Delta\varepsilon_l^{exp}(\%)$	-8.83±0.31	-8.67±0.23	-8.17±0.19	-7.43±0.18
$\Delta\varepsilon_b^{exp}(\%)$	-2.93±0.42	-2.08±0.36	-0.88±0.29	0.47±0.30

Table 5.4: The expected residual corrections for light and b jets which are derived from the Gaussian fit over the various ranges of the fit.

±1RMS to ±3RMS, which results in obtaining a small variation of about 1.4%. For the case of b jets, this variation is larger and reaches to about 3.4% when enlarging the fit range. This is due to the fact that, the distribution of the expected b jet residual correction, can not be fitted in the tail with a Gaussian function and the expected value of the fitted Gaussian would change when increasing the fit range. This observation of the non-Gaussian effects in the tail, which is specific to the b jet distribution, can be explained as follows. Since the b jets can decay to a lepton and the corresponding neutrino with a probability of 20%, the energy of the b jet is underestimated due to the escaping neutrino. Hence, using the definition of the expected b jet energy correction, one can have

$$\Delta\varepsilon_b^{exp} = \frac{E_{b\ quark}}{E_{b\ jet}} - 1 > 0,$$

for those events where the b jet energy is underestimated. Therefore, these kind of events contribute to the tail in the distribution of the expected b jet energy correction as shown in Figure 5.19. Therefore, in order to avoid the non-Gaussian effects in the tail, the fit range is chosen to be ±1.5 times the RMS around the mean of the histogram. However, the dependency of the fit results on the considered range of the Gaussian fit can be taken into account as systematical uncertainty on the expected values for the light and b jet energy corrections. Hence, for each of the expected light and b jet distributions, the maximum variation among the various fit ranges compared to the chosen fit range, being ±1.5RMS, is taken as the systematical uncertainty for that particular distribution. The resulting systematical uncertainties for the expected light as well as b jet energy corrections, are listed in Table 5.5.

	Sys. Unc.
$\delta^{sys.}(\Delta\varepsilon_l^{exp})(\%)$	±1.2
$\delta^{sys.}(\Delta\varepsilon_b^{exp})(\%)$	±2.5

Table 5.5: The systematical uncertainties on the expected light and b jet energy corrections.

In order to compare the expected versus the estimated results, a new parameter, called the bias, is introduced which is defined as the difference between $\Delta\varepsilon_{l/b}^{est}$ and

$\Delta\varepsilon_{l/b}^{exp}$. The estimation of the residual jet energy corrections would be well done if the bias would be equal to zero. The estimated as well as the expected results of the residual jet energy corrections are collected and the corresponding bias for the light and the b jets is calculated which can be found in Table 5.6.

$\Delta\varepsilon_l^{exp}(\%)$	$\Delta\varepsilon_l^{est}(\%)$	$\Delta\varepsilon_l^{est} - \Delta\varepsilon_l^{exp}(\%)$
-8.67±0.23	-10.12±0.13	-1.45±0.26
$\Delta\varepsilon_b^{exp}(\%)$	$\Delta\varepsilon_b^{est}(\%)$	$\Delta\varepsilon_b^{est} - \Delta\varepsilon_b^{exp}(\%)$
-2.08±0.36	-1.31±0.20	0.77±0.41

Table 5.6: The final estimated residual jet energy corrections obtained for 100 pb^{-1}. The corresponding expected residual corrections and the possible bias are also shown.

Taking into account the systematical uncertainties on the expected light and b jet corrections, mentioned in Table 5.5, would result to obtain a bias on, for example, the estimated b jets to be compatible with zero. Other systematical uncertainties on the estimated residual light and b jet energy corrections are calculated in Section 5.4.2.

5.4.1 Performance of the Method

As discussed in Section 5.1.6, the distributions of the reconstructed W boson and top quark masses in the topology of the e+jets $t\bar{t}$ events are shifted from the values that were used in the simulation which consequently motivates the need for residual jet energy corrections. Since in the hadronic branch of the e+jets $t\bar{t}$ event, $t \to Wb \to q\bar{q}b$, both light jets, arising in the decay of the hadronic W boson, and b jets, originating from the decay of the hadronic top quark, are present. Then by imposing the mass constraints of both the W boson and the top quark by means of a kinematic fit approach, one could estimate the residual light as well as the b jet energy corrections as described in Section 5.2. The aim of the method was to correct the energies of the reconstructed jets in the hadronic branch of the e+jets $t\bar{t}$ topology so that the resulting reconstructed W boson and top quark masses are peaked around the nominal parameters which have been used to generate the event.
The plots shown in Figure 5.20 are the invariant masses of the corrected four-vector of the W boson and top quark which are reconstructed from the corrected four-vectors of the reconstructed jets in the final state of e+jets $t\bar{t}$ events. The events for which a true jet-parton combination found by the MVA method, are used to fill the distributions.

Compared to the plots obtained in Figure 5.9, it is obvious that the corrected mass distributions are peaked around their nominal values as it was aimed. The reconstructed mass values of the W boson and the top quark can be further improved by applying a second iteration which is explained in the following. The estimated residual

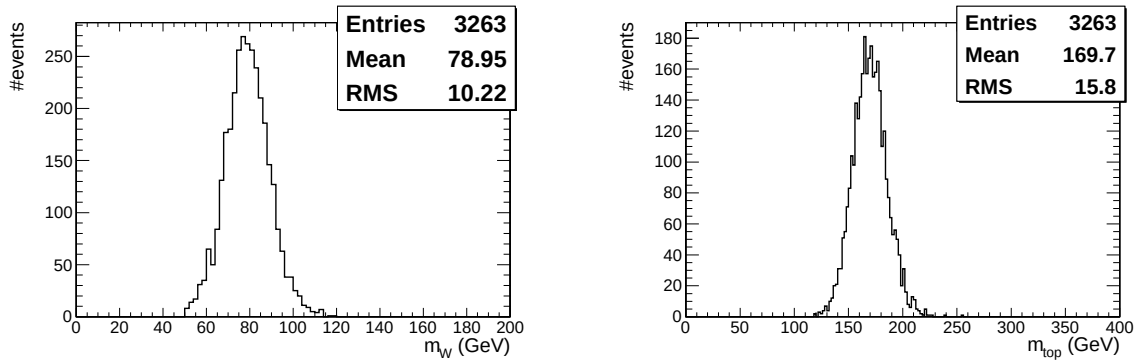

Figure 5.20: The distributions of the reconstructed W boson mass (left) and the top quark mass (right) after applying the estimated corrections on the reconstructed jets. The e+jets $t\bar{t}$ events that pass the extra selection cuts are taken into account.

jet energy corrections which are estimated by the method in a first iteration are applied on the energy of the reconstructed jets and the method is performed for a second time. Ideally the estimated residual jet energy corrections in the second iteration would be equal to zero if the bias in the first iteration was exactly zero. The sum of the estimated jet energy corrections obtained from the first and the second iterations will be a better estimate of the residual jet energy corrections. Another thing which can be done to obtain a more precise mass distribution for either the W boson or the top quark, is to apply correction factors which varies in terms of p_T of the jets. The method is applied on the events which are grouped according to the p_T of the jets to obtain an estimated correction for the jets contributing to that particular bin of p_T. The obtained results can then be applied on the reconstructed jets belonging to that p_T-bin. In addition to the p_T, other kinematic variables such as η can also be used. Applying this kind of differential estimation of the jet energy corrections is expected to result in a mass distribution of both the W boson and the top quark with an improved resolution.

5.4.2 Systematic Uncertainties

Various sources can affect the final estimated results of the jet energy scale correction factors. Some of the most important sources that can influence the results, are discussed below.

The Value of m_{top} as Input Variable for the Kinematic Fit

The kinematic fit package requires the event topology to fulfill the mass constraints. The W boson and the top quark mass values that are used in the package may vary within their uncertainties. While the mass of the W boson is measured with a high precesion of $\sigma_{m_W} = 0.03\%$, the mass of the top quark is measured with an accuracy of $\sigma_{m_{top}} = 0.52\%$, as discussed in 1.2. To be conservative, a larger uncertainty equivalent to $2.5\sigma_{m_{top}}$ is considered on the value of the top quark mass, hence 1.3%. This results in an uncertainty equals to ± 2.3 GeV on the top quark mass value of 172.5 GeV which

has been used in the kinematic fit procedure. Therefore, the analysis is performed with different top quark mass values, being the larger value 172.5+2.3 GeV and the smaller value 172.5-2.3 GeV, as a constraint in the kinematic fit package. The results are shown for the expected and estimated residual corrections in Table 5.7.

	$\Delta\varepsilon_l^{exp}(\%)$	$\Delta\varepsilon_l^{est}(\%)$	$\Delta\varepsilon_l^{est} - \Delta\varepsilon_l^{exp}(\%)$
nominal	-8.67±0.23	-10.12±0.13	-1.45±0.26
larger m_{top}	-8.65±0.23	-10.07±0.13	-1.42±0.26
smaller m_{top}	-8.64±0.23	-10.39±0.13	-1.75±0.26
	$\Delta\varepsilon_b^{exp}(\%)$	$\Delta\varepsilon_b^{est}(\%)$	$\Delta\varepsilon_b^{est} - \Delta\varepsilon_b^{exp}(\%)$
nominal	-2.08±0.36	-1.32±0.20	0.76±0.41
larger m_{top}	-2.08±0.35	0.58±0.20	2.66±0.40
smaller m_{top}	-1.92±0.37	-2.84±0.20	-0.92±0.42

Table 5.7: The expected and estimated residual jet energy corrections and the bias on the estimations for different top quark masses as constraints in the fit package.

What is important to note from Table 5.7, is that the estimated results on the light jet energy correction do not change when using a top quark mass as the constraint which is different from what has been used to generate the physics events. This observation is due to that the light quark jet energy calibration factors are determined by the mass of the W boson which is used in the kinematic fit package and are independent from the value of the top quark mass which is used to constrain the event topology. The estimated residual b jet energy correction varies from -2.84% to 0.58% when increasing the top quark mass value from 170.2 GeV to 174.8 GeV in the fitting procedure. Therefore, compared to the nominal estimated b jet energy correction, a shift of 1.52% or 1.9% is observed when using the smaller or the larger mass of the top quark, respectively. Hence, the maximum difference between the larger and the nominal or the smller and the nominal, is taken as the systematic uncertainty due to the uncertainty on the top quark mass. It should be noted that the uncertainties obtained from Table 5.7 are overestimated since a larger deviation of the top quark mass from its mean value has been considered, as mentioned above.

Pile-Up

Together with the hard scattering event in the proton-proton collision, there are many soft processes coming from the next bunch crossing that might affect the topology of the event, specially in the high luminosity phase where the time between the two successive bunch crossings is so small that the electronic signals from the current event may not be fully collected yet. Hence the pile-up events may result to have more tracks in the track collection and would yield to reconstruct additional objects in the final state. This can vary the final estimation results. In Table 5.8, the expected and

estimated jet energy corrections for the events with and without including pile-up are given.

	$\Delta\varepsilon_l^{exp}(\%)$	$\Delta\varepsilon_l^{est}(\%)$	$\Delta\varepsilon_l^{est} - \Delta\varepsilon_l^{exp}(\%)$
nominal	-9.38±0.25	-8.75±0.14	0.63±0.29
pile-up	-9.52±0.24	-8.68±0.13	0.84±0.27
	$\Delta\varepsilon_b^{exp}(\%)$	$\Delta\varepsilon_b^{est}(\%)$	$\Delta\varepsilon_b^{est} - \Delta\varepsilon_b^{exp}(\%)$
nominal	-2.66±0.36	-2.81±0.20	-0.15±0.41
pile-up	-2.82±0.36	-3.30±0.20	-0.48±0.41

Table 5.8: The expected and estimated residual jet energy corrections and the bias on the estimations in the presence of pile-up events. The nominal refers to the situation when no pile-up is present.

As it is seen from Table 5.8, the method of estimating the residual jet energy corrections is stable with respect to adding the pile-up collisions. Obviously the expected as well as the estimated corrections are affected by the pile-up collisions. If the pile-up is well simulated, the bias should be stable. The estimated residual light and b jet energy corrections do not change much in the presence of the pile-up events. The estimated light jet energy correction changes from -8.75%, corresponding to the nominal value, to -8.68%, which is obtained when including pile-up events. Thus, the shift of 0.07% is considered as the systematic uncertainty on the estimated residual light jet energy correction due to the pile-up effects. The deviation of the estimated b jet energy correction from the nominal value corresponds to 0.49% which is taken as the systematic uncertainty on the estimated residual b jet energy correction due to the pile-up events.

ISR/FSR

When simulating physics events, various parameters are set to control the production and decay products of the hard scattering interactions. Changing any of these parameters individually, would yield to obtain a different topology of the event. Therefore, the sensitivity of the estimated results with respect to the changes of these parameters, has to be studied and the possible deviations are taken as the systematic uncertainties on the final estimation. For example, the amount of ISR/FSR that are allowed to be generated in the simulation procedure can affect the results. Two different samples with a smaller and larger amount of radiation compared to the nominal situation are processed. The estimated residual jet energy corrections are obtained for each of these samples which are given in Table 5.9. The estimated and expected results for light and b jet energy corrections are listed individually for the various samples.

	$\Delta\varepsilon_l^{exp}(\%)$	$\Delta\varepsilon_l^{est}(\%)$	$\Delta\varepsilon_l^{est} - \Delta\varepsilon_l^{exp}(\%)$
nominal	-9.38±0.25	-8.75±0.14	0.63±0.29
larger ISR/FSR	-8.85±0.26	-8.12±0.14	0.73±0.29
smaller ISR/FSR	-9.62±0.25	-9.13±0.14	0.49±0.29
	$\Delta\varepsilon_b^{exp}(\%)$	$\Delta\varepsilon_b^{est}(\%)$	$\Delta\varepsilon_b^{est} - \Delta\varepsilon_b^{exp}(\%)$
nominal	-2.66±0.36	-2.81±0.20	-0.15±0.41
larger ISR/FSR	-2.54±0.37	-2.68±0.20	-0.14±0.42
smaller ISR/FSR	-1.85±0.39	-2.49±0.20	0.64±0.44

Table 5.9: The expected and estimated residual jet energy corrections and the bias on the estimations for the samples with larger and smaller amount of radiation compared to the nominal.

Since the bias on the estimated light and b jet energy corrections is small and, in some cases, is compatible with zero within the statistical uncertainties, it can be deduced that the method is stable with respect to changes in ISR/FSR. The estimated light jet energy correction changes from -8.12%, corresponding to the sample with larger ISR/FSR, to -9.13%, which is obtained when considering smaller amount of ISR/FSR in the simulation procedure. Compared to the nominal value which equals -8.75%, the maximum deviation of the estimated light jet energy correction is taken as the systematic uncertainty due to the ISR/FSR, being 0.63%. In case of b jets, the maximum deviation would reach to 0.32% which is quoted as the systematic uncertainty on the estimated b jet energy correction due to the ISR/FSR.

Scale Factor Q^2

As already mentioned, the initial and final state radiations are governed by the use of the DGLAP equation which describes the evolution of the partonic distribution function as a function of the factorization scale Q^2. According to the DGLAP equation, the final state radiation in an event can be started from a maximum energy state Q^2_{max}, which is chosen to be the squared mass of the parton shower initiator, down to smaller values of the Q^2 scale. Since the chosen value used for the Q^2_{max} parameter might not be the optimal choice for the description of observed proton-proton collision data, the effects of altering the Q^2 value up and downwards are taken into account. Therefore, additional samples are simulated for which the initial input variable of the Q^2_{max} parameter is changed within an interval. Then, in order to check the effects of the variation of the Q^2 scale, the method of estimating the residual jet energy corrections is applied on both simulated samples with the various settings of the Q^2 parameters and the nominal sample which is simulated with the pre-defined setting of the Q^2 parameter. The results of the estimated and expected residual jet energy corrections on different settings of the Q^2 scale are summarized and a possible bias of the method is calculated

and shown in Table 5.10.

	$\Delta\varepsilon_l^{exp}(\%)$	$\Delta\varepsilon_l^{est}(\%)$	$\Delta\varepsilon_l^{est} - \Delta\varepsilon_l^{exp}(\%)$
nominal	-9.38±0.25	-8.75±0.14	0.63±0.29
Q^2 up	-9.30±0.32	-8.76±0.18	0.54±0.37
Q^2 down	-10.14±0.32	-9.20±0.17	0.94±0.36
	$\Delta\varepsilon_b^{exp}(\%)$	$\Delta\varepsilon_b^{est}(\%)$	$\Delta\varepsilon_b^{est} - \Delta\varepsilon_b^{exp}(\%)$
nominal	-2.66±0.36	-2.81±0.20	-0.15±0.41
Q^2 up	-2.52±0.49	-2.57±0.26	-0.05±0.55
Q^2 down	-3.41±0.47	-3.00±0.25	0.41±0.53

Table 5.10: The expected and estimated residual jet energy corrections and the bias on the estimations for the samples with higher and lower values of the Q^2 parameter.

As seen from the results shown in Table 5.10, the effects of changing the input Q^2 parameter on the final estimated results are minor and the estimation follows the expected value since the bias on the estimated results is compatible with zero. Also the effect of Q^2 on the bias is compatible with zero. The systematical uncertainties that can be assigned due to the changes in the Q^2 scale can be evaluated as follows. The estimated light jet energy correction changes from -8.76%, corresponding to the sample with larger Q^2 scale, to -9.20%, which is obtained when considering the simulated sample with smaller value of the Q^2 scale. Compared to the nominal value which is -8.75%, the maximum deviation of the estimated light jet energy correction is taken as the systematic uncertainty due to the changes in the Q^2 scale, being 0.45%. In case of b jets, the maximum deviation would reach 0.24% which is quoted as the systematic uncertainty on the estimated b jet energy correction due to the variation of the Q^2 scale. It should be noted that, when a more precise estimation of the jet energy corrections is aimed for, these systematic uncertainties should be estimated with larger simulated samples.

Matching Matrix Element to Parton Shower

In the procedure of simulating physics collisions, the events are fed to a parton shower program to apply the showering process on the partons which are generated in the collision, as already explained in Section 4.1.2. Subsequently a matching scheme is applied in order to avoid double counting of the radiation. Therefore, the jets produced by the parton shower method are compared with the partons produced with the matrix element method, given that the jets exceeding a pre-defined threshold on the transverse energy. The choice of this threshod might have some effects on the final results of the physics analysis. Hence alternative samples are simulated with a higher and lower threshold on the transverse energy of the jets used in the matching scheme. The effects

of the variation of the matching threshold is investigated by running the method of estimating the residual jet energy corrections on both simulated samples with the various settings of the matching thresholds and the nominal sample which is simulated with the pre-defined setting of the matching threshold. The results of the estimated and the expected residual jet energy corrections with different settings of the matching threshold are summarized in Table 5.11.

	$\Delta\varepsilon_l^{exp}(\%)$	$\Delta\varepsilon_l^{est}(\%)$	$\Delta\varepsilon_l^{est} - \Delta\varepsilon_l^{exp}(\%)$
nominal	-9.38±0.25	-8.75±0.14	0.63±0.29
matching up	-9.19±0.28	-9.31±0.16	-0.12±0.32
matching down	-9.74±0.30	-9.37±0.17	0.37±0.34
	$\Delta\varepsilon_b^{exp}(\%)$	$\Delta\varepsilon_b^{est}(\%)$	$\Delta\varepsilon_b^{est} - \Delta\varepsilon_b^{exp}(\%)$
nominal	-2.66±0.36	-2.81±0.20	-0.15±0.41
matching up	-3.07±0.41	-2.77±0.24	0.30±0.47
matching down	-3.17±0.43	-2.57±0.25	0.60±0.50

Table 5.11: The expected and estimated residual jet energy corrections and the bias on the estimations for the samples with higher and lower threshold in the matching procedure.

It is seen from the results shown in Table 5.11 that the effects of changing the matching threshold on the final estimated results are minor and the estimation follows the expected value since the bias on the estimated results is compatible with zero. The systematical uncertainties that can be assigned due to the changes in the matching threshold can be evaluated as follows. The estimated light jet energy correction changes from -9.31%, corresponding to the sample with larger matching threshold, to -9.37%, which is obtained when considering the simulated sample with smaller value of the matching threshold. Compared to the nominal value which is -8.75%, the maximum deviation of the estimated light jet energy correction is taken as the systematic uncertainty due to the changes in the matching threshold, being 0.62%. In case of b jets, the maximum deviation would reach to 0.24% which is quoted as the systematic uncertainty on the estimated b jet energy correction due to the variation of the matching threshold.

Combined Systematics

The various systematic uncertainties on the estimated results of light and b jet energy corrections are summarized in Table 5.12. The total systematical uncertainty is calculated as the square root of the sum of the squared uncertainties from the diverse sources which are considered in the table.

	m_{top}	pile-up	ISR/FSR	Q^2	matching	Total Sys. Unc.
$\delta^{sys.}(\Delta\varepsilon_l^{est})(\%)$	±0.1	±0.1	±0.6	±0.4	±0.6	±0.9
$\delta^{sys.}(\Delta\varepsilon_b^{est})(\%)$	±0.8	±0.5	±0.3	±0.2	±0.2	±1.0

Table 5.12: The systematical uncertainties on the estimated light and b jet energy correction originating from various sources. The total uncertainty is obtained by the square root of the sum of the squred of the individual errors.

As already noted, the estimated uncertainty on the light and b jet energy corrections due to the variation of the mass of the top quark are overestimated. Therefore, the uncertainties of 0.3% on the light jets and of 1.9% on the b jets are probably conservative estimates. They are divided by a factor of 2.5 to represent the uncertainties on the estimated jet energy corrections resulting from one sigma deviation of the top quark mass value from its world average value. Another point is that, some of the estimated systematical uncertainties quoted in Table 5.12, recieve large statistical uncertainties because they are limited with small simulated samples. For example, the estimated systematical uncertainty on the light jet energy correction due to pile-up source, is dominated by the statistical uncertainty, since from Table 5.8 one can have

$$\delta^{sys.}(\Delta\varepsilon_l^{est})\Big|_{pile-up}(\%) = 0.1 \pm 0.2(stat.).$$

The estimated systematical uncertainty on the b jet energy correction due to either the Q^2 or the matching source, is dominated by the statistical uncertainty, too.

5.5 Statistical Properties of the Estimator

5.5.1 Linearity of the Estimator

The linearity of the method is checked by looking at the calibration curve, which is defined as the variation of the estimated versus the expected jet energy corrections while applying an inclusive shift on the energy scales of the reconstructed jets. Applying any relative inclusive shift on the energy of the reconstructed jets followed by performing the kinematic fit method to constrain the event topology to fulfill the mass requirements, would yield to obtain the estimated results which are linearly dependent to the expected jet energy corrections. This can be understood by looking at the distributions of the $\Delta E_{l,b}^{est}$ versus $\Delta E_{l,b}^{exp}$ which are shown in Figure 5.21. The plots are made using only the e+jets $t\bar{t}$ events.

It can be seen that, when applying the L2 and L3 corrections on the input energies of the reconstructed jets, the estimator is able to predict the expected jet energy corrections. When moving away from the point of nominal L2 and L3 corrections, a deviation of the estimated results from the expected corrections is observed resulting

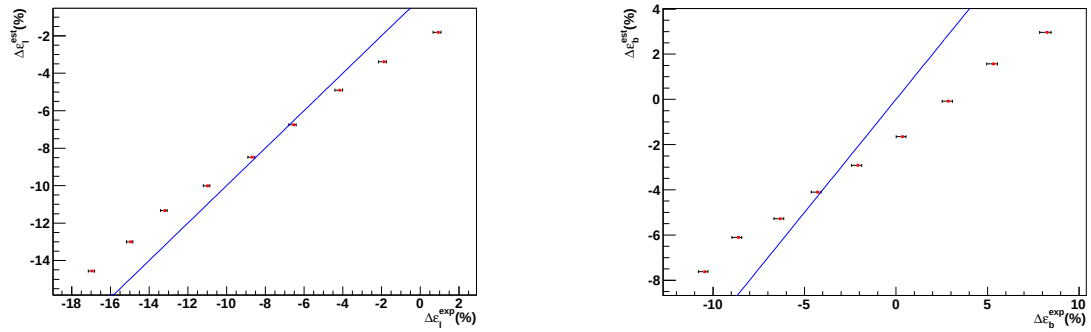

Figure 5.21: The calibration curves representing the estimated residual jet energy corrections as a function of the expected jet energy corrections when applying several inclusive shifts on the input energy of the reconstructed jets. A line with a unity slope is overimposed for comparison.

in a calibration curve with a non-unity slope. This can be fixed as explained below. The non-unity slope of the calibration curves can be solved by the use of successive iterations, which means the estimated jet energy corrections found in a first iteration are applied on the initial jet energy scales. In a second iteration, the residual jet energy corrections can again be estimated, starting with the corrected jets of the first iteration. Ideally, in a second iteration no residual jet energy correction should be needed, given that the estimated jet energy correction is equal to the expected jet energy correction in the first iteration. Othewise, the iteration procedure should be repeated until a residual jet energy correction equals to zero is estimated. This means after applying several iterations, the estimation converges to the line $\Delta\varepsilon_{l,b}^{est} = \Delta\varepsilon_{l,b}^{exp}$, means the slop of the calibration curve becomes close to the unity.

5.5.2 Pull Distribution

In order to evaluate if the statistical uncertainties on the estimators have been quoted correctly, one can start looking at the pull distributions which are defined as

$$\frac{(\Delta\varepsilon_{l,b}^{i} - < \Delta\varepsilon_{l,b} >)}{\delta\Delta\varepsilon_{l,b}^{i}},$$

where $\Delta\varepsilon_{l,b}$ and $\delta\Delta\varepsilon_{l,b}$, as already used, are the estimated residual jet energy corrections and their statistical errors, respectively. The index i is explained as follows. The initial simulated e+jets $t\bar{t}$ sample contains about 180000 events. In an analysis corresponding to 100 pb^{-1}, one would expect to obtain 2333 signal events, as it can be deduced from the cross section of the e+jets $t\bar{t}$ process. Hence instead of giving a weight to the initial simulated events, one can repeat the whole analysis on the smaller bulks of simulated events, so-called pseudo-experiments, where each contain 2333 signal events. Therefore, starting with the initial number of simulated signal events, one would end with about 77 uncorrelated pseudo-experiments. Here no correlation means that a single signal event is not processed in no more than one pseudo-experiment.

Only e+jets $t\bar{t}$ events are taken into account in order to make the pull distributions. No background is considered because not enough simulated events are available.
Running the method to estimate the residual jet energy callibration factors on each of the pseudo-experiments, one can obtain 77 different values as the estimation of the light and b jet energy corrections, corresponding to each of the pseudo-experiments. The obtained results are used to make the pull distribution.
Now it can be explained that the index i in the definition of the pull variable, refers to the i-th pseudo-experiment. Also the <> sign represents the average of the estimated results over all pseudo-experiments.
By definition, the pull distribution is peaked at zero and can be fitted with a Gaussian function. If the estimation of the statistical uncertainties is correct, then the fitted Gaussian would have a variance equal to one. Otherwise there would have been an underestimation or overestimation of the estimated statistical uncertainties, if the width of the fitted Gaussian would have a larger or smaller width, respectively.
The distribution of the pull variable for the estimated light and b jet energy corrections is shown in Figure 5.22.

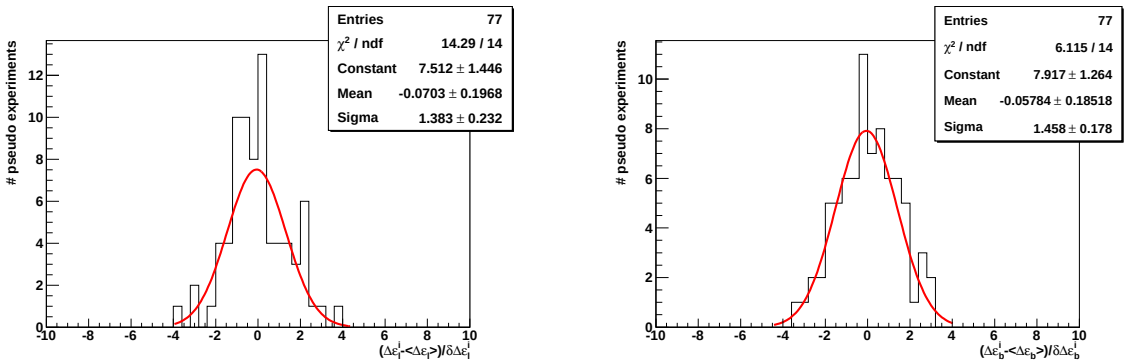

Figure 5.22: The pull distribution of the estimated (left) light and (right) b jet energy corrections. The distributions are fitted with a Gaussian function.

As it is shown in the statistical boxes of the pulls in Figure 5.22, the width of the fitted Gaussian functions is about 1.4 for both estimators. These non-unity widths can be used to correct the statistical uncertainties of the light and b jet energy corrections, which have been found for 100 pb^{-1} of integrated luminosity in Section 5.4, resulting in

$$\Delta\varepsilon_l{}^{est} = -10.1 \pm 1.5\%,$$

and

$$\Delta\varepsilon_b{}^{est} = -1.3 \pm 2.4\%.$$

Therefore, at 7 TeV center of mass energy with the use of a sample of simulated events corresponding to an integrated luminosity of 100 pb^{-1}, the residual jet energy corrections can be estimated with a statistical precision of about 1.5% and 2.4%, for light quark and b quark jets respectively. While using the full sample of simulated events which corresponds to an integrated luminosity of about 9400 pb^{-1}, the residual jet energy corrections can be estimated with a statistical precision of about 0.18% and

0.28%, for light quark and b quark jets respectively. These numbers are corrected for the non-unity width of the pull distributions, too. Therefore, the results of bias on both the light and the b jet energy corrections, which were quoted in Table 5.6, are re-estimated using the corrected statistical errors and summarized in Table 5.13.

$\Delta\varepsilon_l^{exp}(\%)$	$\Delta\varepsilon_l^{est}(\%)$	$\Delta\varepsilon_l^{est} - \Delta\varepsilon_l^{exp}(\%)$
-8.67±0.23(stat.)±1.2(sys.)	-10.12±0.18(stat.)	-1.45±0.29(stat.)±1.2(sys.)
$\Delta\varepsilon_b^{exp}(\%)$	$\Delta\varepsilon_b^{est}(\%)$	$\Delta\varepsilon_b^{est} - \Delta\varepsilon_b^{exp}(\%)$
-2.08±0.36(stat.)±2.5(sys.)	-1.31±0.28(stat.)	0.77±0.46(stat.)±2.5(sys.)

Table 5.13: The final estimated residual jet energy corrections obtained for 100 pb^{-1}. The corresponding expected residual corrections and the possible bias are also shown. The quoted estimated results are corrected for the non-unity width of the pull distributions. The systematical uncertainties on the expected results are mentioned since one cannot fit the distributions of $\Delta\varepsilon_{l,b}^{exp}$ better than the quoted values which should be taken into account when estimating the bias.

Also in Table 5.13, the systematical uncertainty on the expected residual jet energy corrections are mentioned. Since one cannot fit the distributions of $\Delta\varepsilon_{l,b}^{exp}$ better than the quoted uncertainties which have to be considered in the final estimation of the bias.

5.6 A First Application on the 2010 Collision Data

So far, the method to estimate the residual jet energy scale calibration factors has been introduced and the performance of the method on simulated data has been described. In this section, the method is applied for the first time on the collision data which is accumulated in 2010 and corresponding to 36 pb^{-1} of proton-proton collisions at a center of mass energy of 7 TeV. The event selection cuts are slightly different compared to those introduced in Section 4.4, and are listed in the next section. Also in order to illustrate how well the simulated data predict the collision data, some control plots are shown in Section 5.6.2 to compare the simulated versus collision data. The simulated samples which are compared to the collision data, are different from what have been used in previous sections because they need to match with the collision and detector parameters in the low data taking period. They are listed in Table 5.14. Also the names of 2010 collision data sets are given in the same table.

The method to estimate the residual jet energy correction factors has been already applied on the simulated data and the results are obtained for 100 pb^{-1} of integrated luminosity. Although the performance of the method was checked using a huge amount of statistics compared to the accumulated collision data, it is still interesting to inves-

Name of Data Sample
Collision Data
/Electron/Run2010B-Nov4ReReco_v1/AOD
/EG/Run2010A-Nov4ReReco_v1/AOD
Simulated Data
/TTJets_TuneD6T_7TeV-madgraph-tauola/Fall10-START38_V12-v2/AODSIM
/WJetsToLNu_TuneD6T_7TeV-madgraph-tauola/Fall10-START38_V12-v1/AODSIM
/DYJetsToLL_TuneD6T_M-50_7TeV-madgraph-tauola/Fall10-START38_V12-v2/AODSIM
/TToBLNu_TuneZ2_s-channel_7TeV-madgraph/Fall10-START38_V12-v1/AODSIM
/TToBLNu_TuneZ2_t-channel_7TeV-madgraph/Fall10-START38_V12-v2/AODSIM
/TToBLNu_TuneZ2_tW-channel_7TeV-madgraph/Fall10-START38_V12-v2/AODSIM
/QCD_Pt-20to30_BCtoE_TuneZ2_7TeV-pythia6/Fall10-START38_V12-v1/AODSIM
/QCD_Pt-30to80_BCtoE_TuneZ2_7TeV-pythia6/Fall10-START38_V12-v1/AODSIM
/QCD_Pt-80to170_BCtoE_TuneZ2_7TeV-pythia6/Fall10-START38_V12-v1/AODSIM
/QCD_Pt-20to30_EMEnriched_TuneZ2_7TeV-pythia6/Fall10-START38_V12-v1/AODSIM
/QCD_Pt-30to80_EMEnriched_TuneZ2_7TeV-pythia6/Fall10-START38_V12-v1/AODSIM
/QCD_Pt-80to170_EMEnriched_TuneZ2_7TeV-pythia6/Fall10-START38_V12-v1/AODSIM

Table 5.14: The list of collision data samples and simulated data sets which are used to evaluate the performance of the simulation.

tigate if the method works on the smaller amount of collected events. This is checked in the following.

5.6.1 Event Selection Cuts

The baseline event selection cuts developed in the Top Quark Analysis Group within the CMS collaboration are used for this part of the study. They are collected in [104] and are summarized below.

- The collision data has been collected using different trigger requirements depending on the various runs performed during the data taking period. Collision events are asked to have at least one reconstructed electron with a minimal cut on the transverse momentum exceeding 15 GeV. The cut on the electron transverse momentum becomes tighter for some of the runs. Besides applying a cut on the transverse momentum of the reconstructed electron, some identification criteria are also checked for the reconstructed electron varying run by run. The event is kept if there is a reconstructed electron passing the p_T cut and the identification requirements. No trigger criteria are applied on the simulated data samples.

- In order to reduce the huge amount of statistics, a set of pre-selection cuts is applied. All events are required to contain at least one reconstructed electron with $p_T > 15\ GeV$ and four reconstructed jets, each of which has a transverse momentum exceeding $20\ GeV$. The reconstructed electron and jets are asked to be within the tracker acceptance of $|\eta| < 2.4$.

- The next cut in the series of the event selection cuts, is to apply quality cuts on the reconstructed vertex of the event. The primary vertex of the event is asked to be reconstructed close to the nominal interaction point. At least five tracks should contribute to reconstruct the vertex, $ndof > 4$, and also it is required to have $|z| < 24\ cm$ and $\rho < 2\ cm$ relative to nominal centre of the detector. The definitions of these parameters were given in Section 4.4.1.

- The event is asked to have exactly one isolated electron passing the criteria listed below.

  - In order to reject the electrons which are reconstructed within jets, the candidate electron is required to be reconstructed very close to the primary vertex. The absolute value of the difference between the z-component of the reconstructed electron and the z-component of the primary vertex of the event is required to be less than one centimeter.
  - The threshold cut on the transverse energy of the candidate electron is increased compared to the cut applied in pre-selection step. The reconstructed electron is needed to have $E_T > 30\ GeV$.
  - The threshold cut on the pseudorapidity of the candidate electron is not changed with respect to the cut applied in pre-selection step. As an extra requirement, the electron is asked to be reconstructed outside of the transition region between the ECAL barrel and the ECAL endcap.

  - Subsequently the candidate electron is required to have an absolute impact parameter d_0 less than 0.02 cm.
  - Also the electron is asked to fulfill the identification criteria which were described in Section 4.4.1. The cut values are the same as before so that an identification efficiency of about 70% can be obtained.
  - The final criteria that the candidate electron should fulfill is applied on the relIso variable. The definition of the relIso variable was already given in Section 4.4.1. The electron is asked to have a relIso value less than 0.1.

- The event is rejected if it contains any isolated muon, which is defined as a global muon with a p_T exceeding 10 GeV and located within $|\eta| < 2.5$. The isolation criteria is applied on the relative isolation variable and the candidate muon is defined as isolated if its relIso variable is less than 0.2. In case such muon exists within the event, that event is rejected.

- In this step, the events which happen to be in the peak of the mass distribution of the Z boson are removed. The invariant mass of the selected isolated electron and an other loose electron is calculated. If the resulting invariant mass is within the range of 76 GeV to 106 GeV, the event is rejected. The loose electron is defined using the same variables as given in Section 4.4.1.

- There is another additional requirement which is applied on the selected electron. Electrons can be initiated from various sources in the event. For example, it is possible that the electron originates from a converted photon. Photons can produce an electron-positron pair while traversing the tracker material. Therefore there is a probability that the selected electron initiates from a photon conversion. In order to reject the electrons which come out of a conversion process, some algorithms are developped within the CMS collaboration. The algorithm starts by looking for a partner track reconstructed very close to the selected electron. Since the electron and positron initiated from a converted photon bend in opposite direction due to the presence of the magnetic field, it is required that the partner track has an opposite charge with respect to the electron track. If the distance between the two tracks is less than a pre-defined value, the selected electron is flagged as conversion. The detailed information of the technical implementation can be found in [105].

- The event is required to have at least 4 reconstructed calorimeter jets. The jets are reconstructed using the anti-k_T algorithm where the cone radius is set to be 0.5 as before. It should be noted that the jets are L2 and L3 corrected. The selected jets should fulfill the following requirements.

  - A tighter cut on the transverse momentum of the reconstructed jets is applied compared to the cut which has been applied in the pre-selection step. The jets are asked to have $p_T > 30\ GeV$. The threshold cut on the pseudorapidity variable of the jets is the same as the one applied in the pre-selection step, $|\eta| < 2.4$.

  - There is a lower limit on the f_{EM} variable, defined in Section 4.4.1, which is set to be 0.01. No upper limit is put on this property of the reconstructed jets.
  - The selected jets are required to have at least two reconstructed hits which contain 90% of the jet energy, n90Hits > 1. The definition of n90Hits was given in Section 4.4.1.
  - The final cut in the series of jet identification cuts is applied on f_{HPD} variable which is defined in Section 4.4.1. The selected jets are asked to have $f_{HPD} < 0.98$.
  - Also the jet collection is cleaned from electrons. Any jet which is reconstructed close to the selected electron with $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.3$, is not counted as a reconstructed jet.

- The event is required to have at least one b-tagged jet among the four selected jets, which is defined as a jet with a b-tagging value returned by "Track Counting High Efficiency" algorithm exceeds 4 as was used in Section 4.4.1.

The summary of the selection cuts together with the number of events passing the different steps is shown in Table 5.15.

It is obvious from Table 5.15 that after applying all event selection cuts, one does obtain 176 data events versus 201 events expected from the simulated events. The obtained results are in reasonable agreement within the statistical uncertainties. Also it can be understood that the dominating background in the selected sample of the simulated events comes from the other $t\bar{t}$ events. The significance parameter, which is defined as $\frac{S}{\sqrt{B}}$, can be a measure of how well the signal events are selected against to the backgrounds. According to Table 5.15, the significance is calculated to be aroud 20. This shows that the final selected sample is well populated by the e+jets $t\bar{t}$ events. The distributions of the various properties of the data events versus simulation are compared and shown in the next section.

5.6.2 Data-Simulation Comparison

Now that the baseline event selection cuts are applied, it is interesting to look at the different properties of the selected events and compare the resulting distributions betweeen simulation and collision data. In Figure 5.23, the distributions of the selected electron transverse momentum and pseudorapidity are shown. The events remaining after applying all event selection cuts are taken into account.

It is understood that the collision data and the simulated events are in good agreement within the statistical errors.
Another interesting property of the selected electron is the relative isolation variable. The distribution of the selected electron relIso is shown in Figure 5.24.

Again a good agreement between the collision data and the simulation can be observed.

	$t\bar{t}$ $\rightarrow e + jets$	$t\bar{t}$ $\rightarrow$ other	W+jets $\rightarrow l + jets$	Z+jets $\rightarrow l^+l^- + jets$	t $\rightarrow l + jets$	QCD multijets	Data
Initial	951.7	5473.2	1582609.6	154045.9	1188.8	1.9762447E+8	8.5413535E+7
Pre-Selection	777.2	2076.4	7607.7	2953.6	221.0	2119019.0	1065204
PV	777.1	2076.0	7606.2	2952.8	221.0	2118614	1065116
One Electron	425.0	149.1	3234.6	1255.4	64.6	3384.8	8485
Muon Veto	423.8	111.6	3232.6	1252.7	63.4	3384.2	8441
Z Veto	420.9	106.6	3228.2	693.8	62.9	3381.2	8011
Conversion Rejection	397.0	100.1	3023.2	648.4	59.3	1230.5	5118
Four Jets	194.0	30.5	159.2	38.6	10.6	65.8	402
One B Jet	146.8	23.4	15.1	3.5	7.5	4.6	176

Table 5.15: Table of event selection cuts. The number of events are equivalent to an integrated luminosity of 36.1 pb^{-1}.

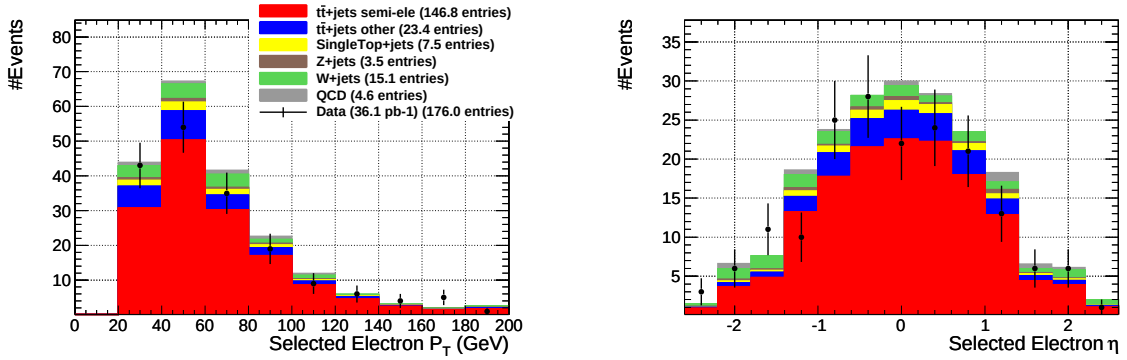

Figure 5.23: The distributions of the selected electron transverse momentum and pseudorapidity. The collision data events are overimposed to the simulation for comparison purpose.

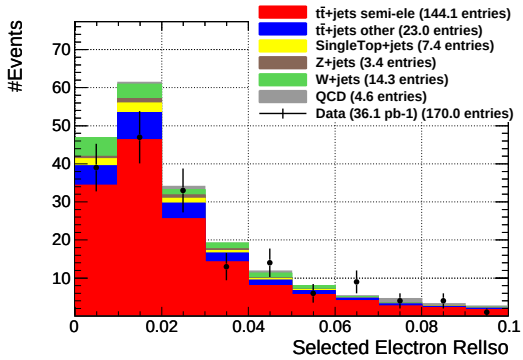

Figure 5.24: The distribution of the relative isolation of the selected electron. The collision data events are overimposed to the simulation for comparison purpose.

Also, one can look at the kinematics of the selected jets. In Figure 5.25, the distributions of the transverse momentum and pseudorapidity of the selected jets are shown. The events remaining after applying all event selection cuts are taken into account.

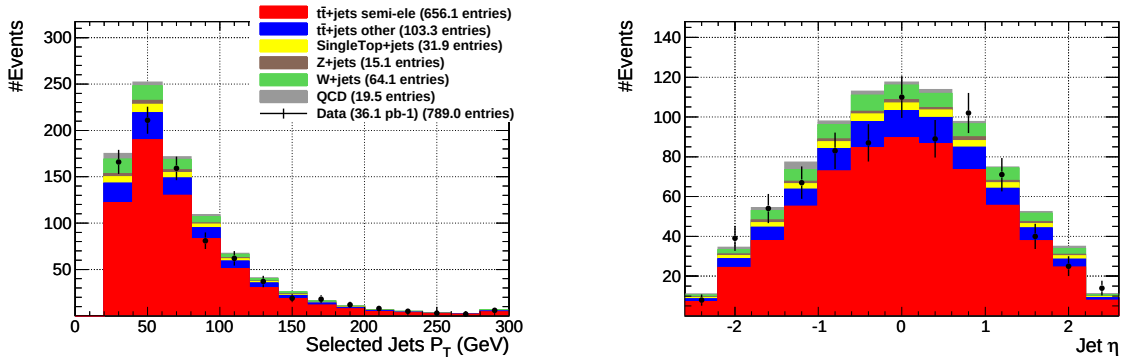

Figure 5.25: The distributions of the transverse momentum and pseudorapidity of the selected jets. The collision data events are overimposed to the simulation for comparison purpose.

As it is seen from Figure 5.25, the data events follow the same behaviour as predicted by the simulation. In other words, the available theoretical models are able to describe the collision data for top quark topologies with this small integrated luminosity.
As another property of the selected events, one can look at the distribution of the ΔR distance between the selected electron and the closest jet among the four selected jets. Since the jet-electron cleaning has been already applied during the event selection cuts, then the selected jets are expected to be reconstructed in a distance greater than 0.3 in (η, ϕ) space with respect to the selected electron. This can be clearly seen from Figure 5.26. Again there is a good agreement between collision data and simulated events which can be observed from the distribution of the minimal distance between the selected electron and a jet among the selected jets.

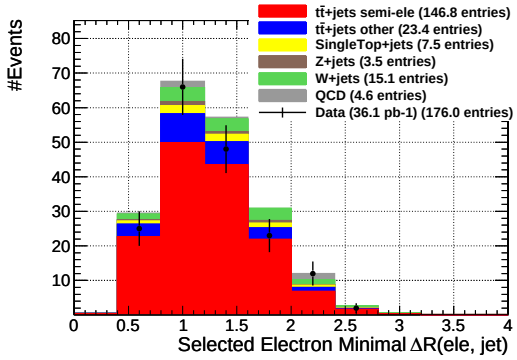

Figure 5.26: The distribution of the minimal distance between the selected electron and a jet among the four selected jets. The collision data events are overimposed to the simulation for comparison purpose.

The next observable of the event which is interesting to check, is the distribution of the missing transverse energy E_T^{miss}. Since the source of E_T^{miss} in the e+jets $t\bar{t}$ events, is the neutrino which appears in the decay product of the W boson, then it is expected to obtain a clear peak around 40 GeV, as it is clearly seen from Figure 5.27.

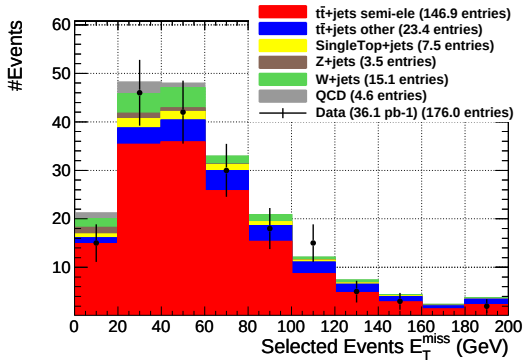

Figure 5.27: The distribution of the missing transverse energy E_T^{miss}. The collision data events are overimposed to the simulation for comparison purpose.

It can be concluded from Figure 5.27 that the distribution of the missing transverse energy in the collision data is comparable with the one in the simulated data.
Although the four-vector of the W boson can not be fully reconstructed due to the unknown z component of the neutrino momentum, it is still possible to reconstruct another variable that contains the information of the mass of the W boson. The transverse mass of the W boson, M_T, is defined as

$$M_T = \sqrt{2p_T^e E_T^{miss}(1 - \cos\phi_{e\nu})},$$

where e denotes the selected electron and $\phi_{e\nu}$ is the azimuthal angle between the selected electron and the missing transverse energy which is measured in the transverse plane. The distribution of the M_T is shown in Figure 5.28.

By definition, if the longitudinal component of the momentum of the W boson is equal to zero, which means the W boson is produced in the transverse plane, then the transverse mass of the W boson, as defined above, becomes the invariant mass of the W boson. In other words, the mass of the W boson is the upper limit for the transverse mass variable, $M_T < 80.4$. As it can be seen from Figure 5.28, there is a clear peak around the mass of the W boson. Also again a good agreement between the collision data and the simulation can be observed.
There is another important observable, called M3, that can be used to compare the behaviour of the data versus simulation. The M3 variable can also be used as an estimator for the mass of the reconstructed top quark. Before applying the MVA method, which makes it possible to label the selected jets and associate them to the hadronic quarks coming from the hadronic decay of the top quark $t \to Wb \to q\bar{q}b$, it is still possible to construct an estimator that can be a measure of the invariant mass

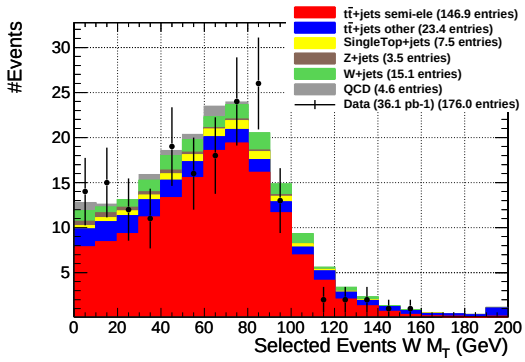

Figure 5.28: The distribution of the transverse mass of the W boson M_T. The collision data events are overimposed to the simulation for comparison purpose.

of the top quark. The M3 variable is defined as the invariant mass of the three jets among all selected jets for which the transverse momentum of the resulting vectorial sum is the maximum. The distribution of M3 variable for all selected events can be found in Figure 5.29.

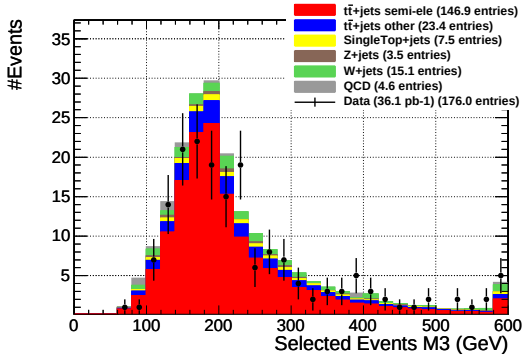

Figure 5.29: The distribution of the M3, defined as the invariant mass of the three jets which yield the maximum transverse momentum of the vectorial sum. The collision data events are overimposed to the simulation for comparison purpose.

The peak of the M3 distribution happens around 180 GeV which is close to the mass of the top quark that was used when simulating the physics events. Then it can be concluded that indeed the M3 variable represents a good estimator of the invariant mass of the top quark. Again a good agreement between data and simulated events can be observed.
Now that a good compatibility between the collision data and the simulation is observed considering various properties of the selected events, one can run the MVA package in order to reconstruct the topology of the e+jet $t\bar{t}$ event. The same variables as introduced in Section 5.1.4, are used as input in order to train the Likelihood Ratio method. As already mentioned, the Likelihood Ratio method assigns a value to each jet-parton combination and returns the chosen jet-parton combination corresponding to the maximum Likelihood Ratio value among the 12 possible jet-parton combinations.

The distributions of some variables, which were used as input to train the MVA method, are shown for the collision data and the simulated events in Figure 5.30. They are calculated using the chosen jet-parton combination which is returned by the MVA method.

Looking at the high-level variables which are shown in Figure 5.30, the good agreement between data and simulation can be observed.
With the jets labeled to the partons in the $t \rightarrow Wb \rightarrow q\bar{q}b$ decay using the MVA method, it is time to reconstruct the W boson and the top quark four-vectors. The distributions of the mass of the W boson and the top quark are shown in Figure 5.31.

Although the chosen jet combination by the MVA method does not correspond to the true combination, what is observed from Figure 5.31 is that the data events follow the same behaviour as the simulation and a good agreement can again be obtained.
In the following the selected events are fed to the kinematic fit package. As before, the kinematic fit is performed 41×41 times per event while scanning a region corresponding to a window of ±40% around the non-corrected energies of the light as well as the b jets. The fit is performed in steps of 2% in both the light and b jet energy corrections to make a two-dimensional plot of the fit probabilities. The output of the kinematic fit method for a typical collision data event is shown in Figure 5.32. The information of the event, including the run number and the event number, are quoted on the plot.

It can be seen from Figure 5.32, that the fit has been converged in each point in the two-dimensional space which is spanned in the direction of the residual light and b jet energy corrections. Also it is clearly observed that the maximum of the fit happens somewhere in the scanned region which confirms that the mass constraints are maximally fulfilled at that point. Therefore, the observed collision data event can represent an e+jets $t\bar{t}$ event.
The distribution of the kinematic fit probability in the selected events when no correction is applied $P_{KinFit}(\Delta E_l = 1, \Delta E_b = 1)$, is shown in Figure 5.33.

The peak at zero justifies the need for the residual jet energy corrections. It is also interesting to look at the distribution of the maximum probability returned by the kinematic fit $P^{max}_{KinFit}(\Delta E_l, \Delta E_b)$ per event, which is shown in Figure 5.34.

Again what can be seen from Figure 5.34 is the compatibility of the collision data and the simulated events within the statistical uncertainties.
In order to apply the method of estimating the residual jet energy corrections on the data, the same extra event selection cuts, as described in Section 5.3, are applied on the collision data. These additional event selection cuts are summarized in Table 5.16.

According to the results that were obtained using 100 pb^{-1} of the integrated luminosity shown in Table 5.3, the number of e+jets $t\bar{t}$ events survived, for example, the cut on the maximum probability of the kinematic fit $P^{max}_{KinFit} > 0.98$, is around 75. Therefore, in an analysis which corresponds to an integrated luminosity of 36.1 pb^{-1},

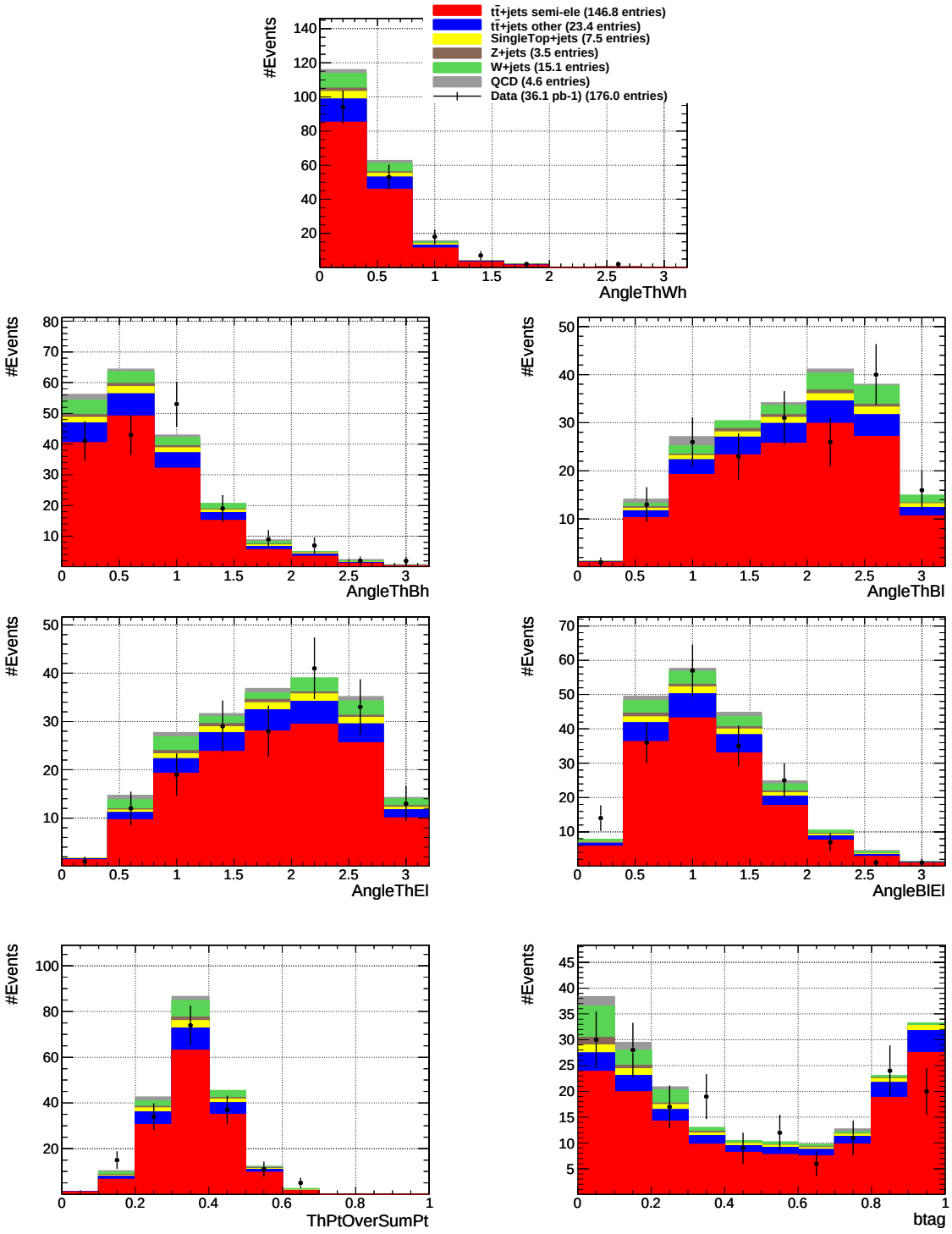

Figure 5.30: The distribution of the observables which are calculated using the four-vectors chosen by the Likelihood Ratio method. These variables were used for training the Likelihood Ratio method. The collision data events are overimposed to the simulation for comparison purpose.

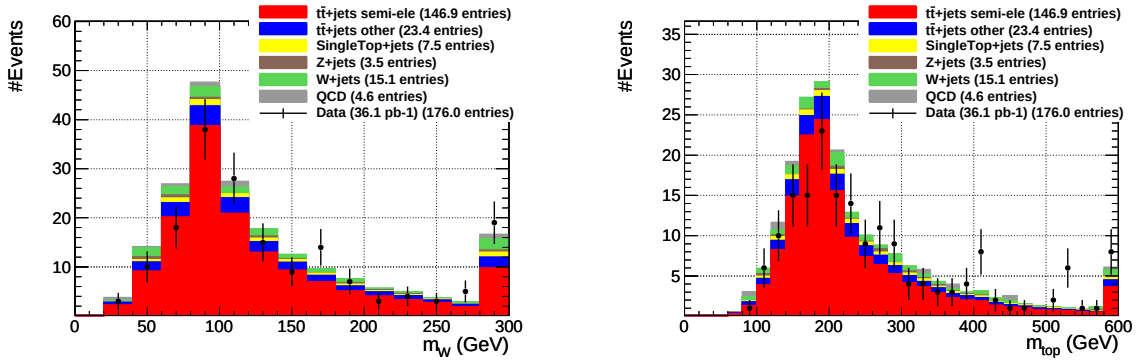

Figure 5.31: The mass distributions of the W boson (left) and the top quark (right) using the four-vectors returned by the MVA method. The collision data events are overimposed to the simulation for comparison purpose.

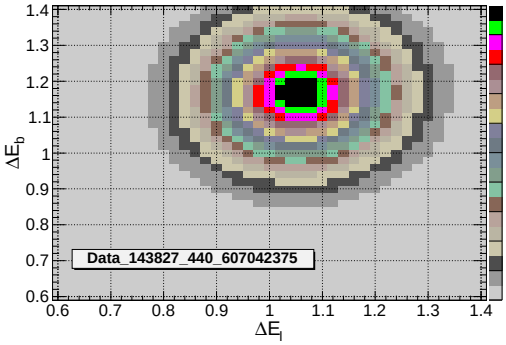

Figure 5.32: The kinematic fit output for a typical collision data event. The third dimension contains the probability of the kinematic fit $P_{KinFit}(\Delta E_l, \Delta E_b)$.

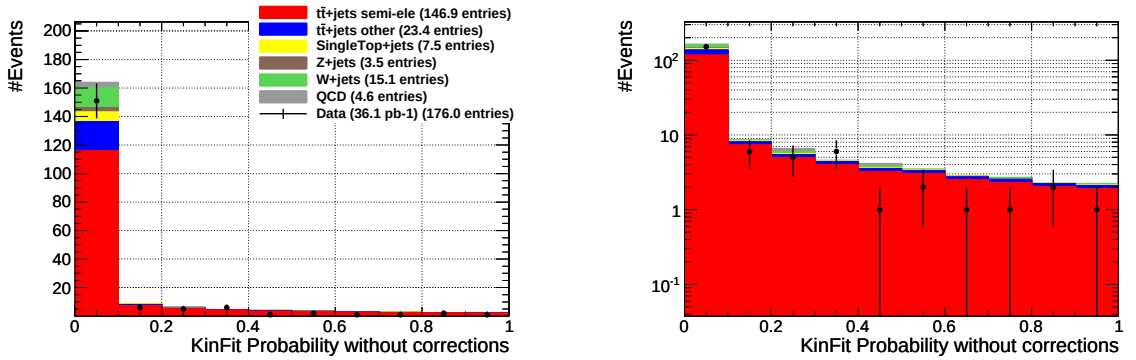

Figure 5.33: The distribution of the kinematic fit probability when no correction is applied shown in normal scale (left) and logarithmic scale (right). The collision data events are overimposed to the simulation for comparison purpose.

	$t\bar{t}$ $\rightarrow e + jets$	$t\bar{t}$ $\rightarrow$ other	W+jets $\rightarrow l + jets$	Z+jets $\rightarrow l^+l^- + jets$	t $\rightarrow l + jets$	QCD multijets	Data
Selection Table 5.15	146.8	23.4	15.1	3.5	7.5	4.6	176
Radiation Veto	96.6	16.1	12.2	2.5	5.9	3.7	115
P^{max}_{KinFit}	41.9	4.8	2.2	0.5	1.3	0.9	40
Gap Veto	36.7	4.0	1.9	0.4	1.1	0.9	34

Table 5.16: Table of extra event selection cuts. The numbers are equivalent to an integrated luminosity of 36.1 pb^{-1}.

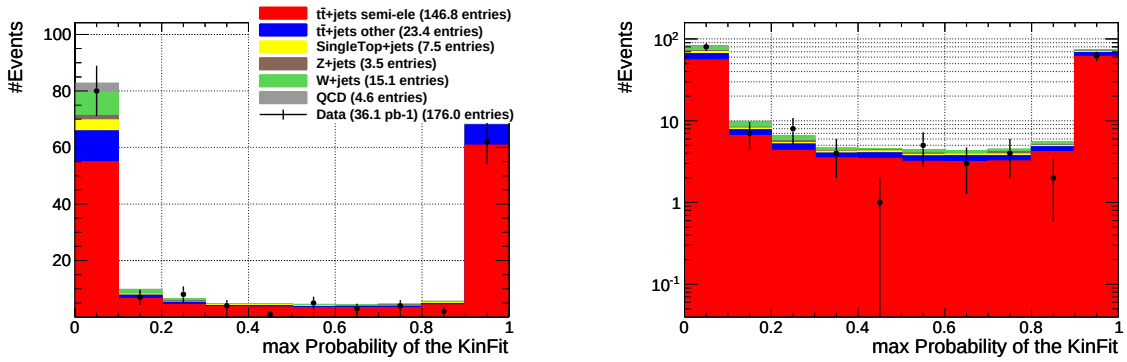

Figure 5.34: The distribution of the maximum probability returned by the kinematic fit shown in normal scale (left) and logarithmic scale (right). The collision data events are overimposed to the simulation for comparison purpose.

one may expect to obtain a number of e+jets $t\bar{t}$ events which survive the P^{max}_{KinFit} to be about $\frac{36.1}{100} \times 75 \sim 27$, given that the same event selection cuts are applied. As it can be seen from Table 5.16, there are about 42 e+jets $t\bar{t}$ events remained at the same level of the selection procedure, hence more events than what is expected. This can be explained since a looser set of cuts has been applied on the reconstructed jets when running on the data.
With the 34 events in the collision data passing the full event selection, one can start estimating the residual light and b jet energy correction factors. For each of the survived data events, the twoi-dimensional grid of points which is filled with the kinematic fit probabilities, $P_{KinFit}(\Delta E_l, \Delta E_b)$, is transformed to a χ^2 denoted as $\chi^2(\Delta E_l, \Delta E_b)$. The χ^2 values in each point of the grid are summed for all of the final selected data events. The results can be found in Figure 5.35.

The final estimated light and b jet energy corrections are determined at the minimum point of the $\chi^2(\Delta E_l, \Delta E_b)$ distribution which is shown in Figure 5.35. The projected χ^2 distribution in the direction of the ΔE_l and ΔE_b, shown in Figure 5.36, are fitted with a second-degree polynomial and the points with the minimum values are quoted as the light and b jet energy calibration factors, respectively. The final estimated residual jet energy corrections, which are obtained using the 36.1 pb^{-1} of the accumulated collision data, are listed in Table 5.17.

In order to compare the results derived from the collision data events with the results obtained from the simulation, the same procedure to determine the residual jet energy corrections can be applied using only the final survived simulated events which are listed in Table 5.16. The results of the residual jet energy corrections estimated based on the simulation are also shown in Table 5.17.
It should be noted that the statistical uncertainties on the estimated jet energy corrections listed in Table 5.17, are corrected for the non-unity width of the pull distributions. The systematical uncertainties on the estimated residual light and b jet energy corrections derived from the simulation are also quoted. Clearly the uncertainties on the estimated results are dominated by the statistics. The statistical uncertainties can be

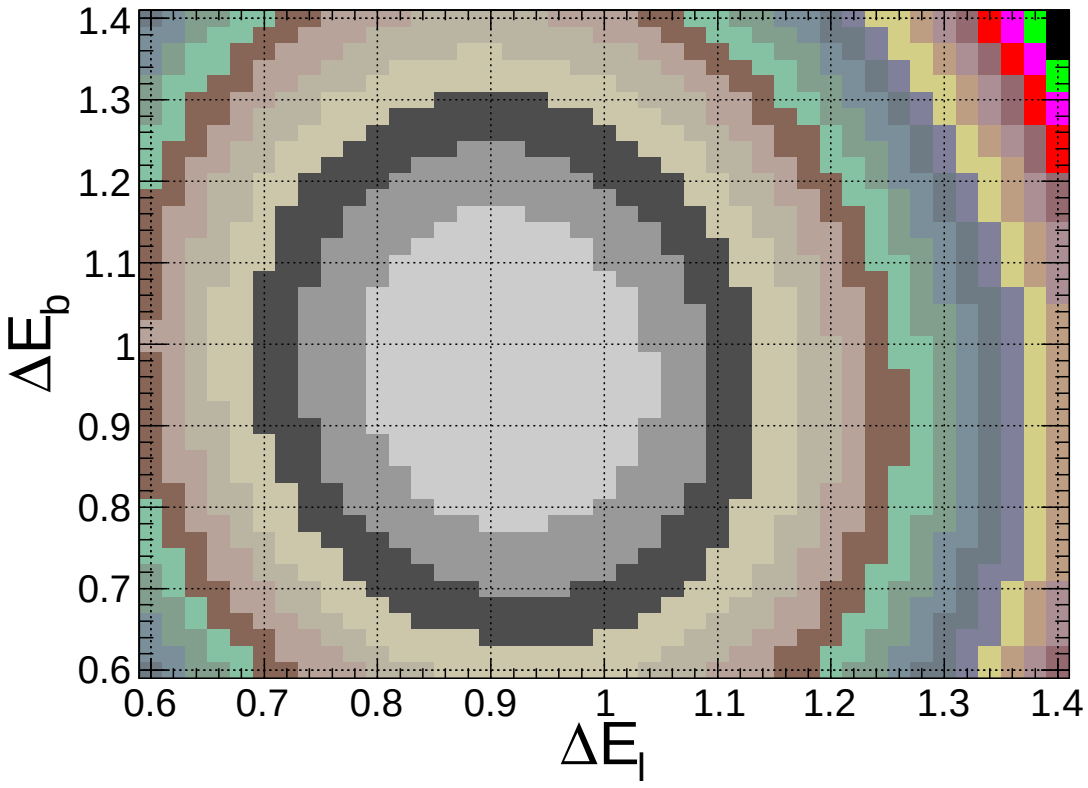

Figure 5.35: The distribution of the $\chi^2(\Delta E_l, \Delta E_b)$ which is obtained by the sum of the χ^2 distributions for all selected collision data events.

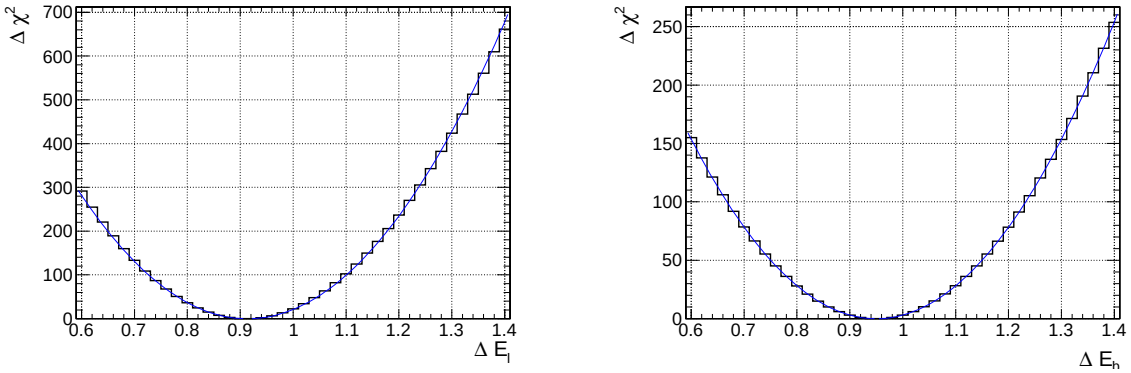

Figure 5.36: The projected distribution of the $\chi^2(\Delta E_l, \Delta E_b)$ into the (left) ΔE_l and (right) ΔE_b axis which are fitted with a second-degree polynomial. The results are obtained for 36.1 pb^{-1} of real proton collision data.

	simulation	data
$\Delta\varepsilon_l^{est}(\%)$	-8.5±2.4(stat.)±0.9(sys.)	-8.6±2.7(stat.)
$\Delta\varepsilon_b^{est}(\%)$	-3.2±3.5(stat.)±1.0(sys.)	-5.0±3.9(stat.)

Table 5.17: The final estimated residual jet energy corrections obtained in the 36.1 pb^{-1} data set. The results found on the simulation and the collision data events, are listed separately. The statistical uncertainties are corrected for the non-unity width of the pull distributions.

reduced by using a larger sample of data events, as will be discussed in more detail in the next chapter. Obviously the residual corrections are found to be equal in data and simulation. They are in reasonable agreement within the quoted uncertainties.

Summary of the Analysis

The method to estimate residual jet energy corrections has been applied on 36.1 pb^{-1} of integrated real proton collisions collected by the CMS experiment in 2010. The residual corrections are estimated based on the e+jets $t\bar{t}$ candidates. Event selection cuts have been applied to enhance the signal over background ratio followed by imposing the mass constraints of the W boson and the top quark with the use of a kinematic fit technique. The kinematic fit procedure provides a χ^2 value used to fill a two-dimensional distribution in the directions of ΔE_l and ΔE_b, on which the final estimation is based. As it is seen from Table 5.17, the estimated residual jet energy corrections which are obtained using the collision data events are compatible with the results derived based on the simulation. Then it can be claimed that the method works reasonably on the 36.1 pb^{-1} of accumulated 2010 collision data taken by the CMS experiment at the LHC.

Chapter 6

Conclusion

The Standard Model of particle physics is able to describe successfully the elementary particles and their interactions. The parameters of the model have been widely measured in diverse experiments which made the Standard Model an excellent candidate to be a reference theory for the description of the nature at the energy scale of Electroweak interactions. Despite all the successes of the model, there is a part for which no experimental evidence has been observed yet, being the Higgs mechanism which is responsible for giving mass to the elementary particles. This was discussed in detail in Chapter 1.

A possible solution would be provided by the Large Hadron Collider (LHC) as it mainly aimes to search for the missing particle predicted by the mechanism, namely the Higgs particle. Since March 2010, the LHC has been colliding beams of protons at a center of mass energy of 7 TeV and the collision data has been recorded by detectors installed at four points around the LHC main ring, such as the CMS experiment (Compact Muon Solenoid). The full description of the LHC and the CMS experiment were given in Chapter 2.

When protons collide, various objects can be produced at the final state, such as isolated leptons and quarks. According to the theory, quarks carry color-charges and therefore can not be observed in free forms. They hadronize and make a cascade of hadrons, which is referred to as jets. One of the main challenges at hadron colliders is how to define a jet and how to measure its energy. A simple definition of a jet is to collect all the particles produced in the hadronization process within a cone around the initial quark. This works as a first guess but one should note that due to the existence of the magnetic field, the charged particles produced in the cascade are bent and move away from the direction of the initial quark. This yields to a measurement of the energy of the jets lower than its true energy, as the energy deposits of the escaping particles are not taken into account in the jet area considered in the jet reconstruction algorithm. Therefore, jets need to be calibrated so that their energies are corrected back to their true values. The various jet definitions and calibration strategies which are introduced within the CMS collaboration, were explained in Chapter 3.

Since jets are produced in final states of many channels of interest, it is desired to determine their energy as precise as possible. Starting with the baseline calibration recipe, the energy of the jets can be further calibrated with the use of constraints

which are imposed on the system. In this thesis, the e+jets decay channel of $t\bar{t}$ events is chosen as the signal $t\bar{t} \to W\bar{W}b\bar{b} \to e\nu_e q\bar{q}b\bar{b}$, which contains an isolated electron that can be used to suppress the QCD multijets background and at least four reconstructed jets where two of them originate from heavy quarks. Several cuts are applied to reject the huge backgrounds and to select the signal events which were discussed in detail in Chapter 4. Having applied the event selection cuts listed in Table 4.6 on a sample of simulated events corresponding to 100 pb^{-1} of integrated luminosity, there are about 308 e+jets $t\bar{t}$ events remained resulting in a signal over background ratio of about 2.5.

As there are at least four reconstructed jets in the final state of e+jets $t\bar{t}$ events, which are indistinguishable if no tagging tools are used, a first task is to label the reconstructed jets and associate them to the hard scatter partons. This can be done with the use of Multi-Variate Analysis techniques (MVA). A Likelihood Ratio method is used to return a chosen jet combination based on the information which is obtained during the training of the MVA method. The chosen jet combination by the MVA method is not all the time the true combination which represents the combination where jets are matched to the partons. The performance of the MVA method strongly depends on the input variables which are chosen for the training. A full discription of the MVA method and a list of variables used for the training of the Likelihood Ratio method were given in Chapter 5. It was found that in around 17% of the selected e+jets $t\bar{t}$ events, the Likelihood Ratio method is able to return the correct combination where the three reconstructed jets match to the quarks from the hadronic top quark decay $t \to Wb \to q\bar{q}b$. This is a great improvement compared to the random choice of the correct jet-parton combination which is equal to 8%.

Considering the hadronic branch of the signal $t \to Wb \to q\bar{q}b$, the masses of the W boson and the top quark can be used to constrain the system by requiring the four-vectors of the reconstructed jets of the chosen jet combination to fulfill the mass constraints. The procedure of applying constraints is performed with the use of a kinematic fit, which was explained in detail in Chapter 5. The $t\bar{t}$ events are excellent candidates for the jet calibration purpose because, in addition to the high rate of top pair production, they contain both the light and the b jets in their final state which yields to obtain calibration factors for different flavours of the quarks. Starting with the three jets in the hadronic branch of the selected event $t \to Wb \to q\bar{q}b$, which are now assigned to the hard scatter partons by means of the MVA method, the energy of the jets are altered within a window of $\pm 40\%$ around the reconstructed energies. Changes are made in steps of 2%. As a result, a two-dimensional space is spanned in the directions of the light jet energy scale calibration factor ΔE_l and the b jet energy scale calibration factor ΔE_b. For each selected event and in each point of the two-dimensional grid, the kinematic fit returns a probability $P_{KinFit}(\Delta E_l, \Delta E_b)$, which reflects how likely the hypothesis of the three jets to originate from the top quark in the chosen jet combination, is true. The probability of the kinematic fit can be transformed into a $\chi^2(\Delta E_l, \Delta E_b)$ space. The information of all the selected events is combined by the sum of the χ^2 distribution of the individual events, which makes a two-dimensional parabola. The resulted parabola can be projected in each of the two dimensions and the minimum of each of the two projected parabolas, refers to an estimation of the residual jet energy scale calibration factors for both light and b quark jets.

The method to estimate the residual jet energy scale calibration factors has been tested on simulated proton collisions and the performance of the method has been discussed in Chapter 5. The same method has also been applied on the real proton collisions which have been collected by the CMS experiment in 2010, corresponding to an integrated luminosity of 36 pb^{-1} of proton-proton collisions at a center of mass energy of 7 TeV which was described in detail in Chapter 5. The estimated residual jet energy corrections resulting from the application of the method on the simulation as well as on the real proton collisions are summarized in Table 6.1. The statistical uncertainties quoted on the estimated results are corrected for the non-unity width of the pull distributions. The complete description of the statistical properties of the estimator including the pull variable can be found in Chapter 5.

	simulation	data
$\Delta\varepsilon_l^{est}(\%)$	-8.1±2.4(stat.)±0.9(sys.)	-8.6±2.7(stat.)
$\Delta\varepsilon_b^{est}(\%)$	-3.2±3.5(stat.)±1.0(sys.)	-5.0±3.9(stat.)

Table 6.1: The final estimated residual jet energy corrections obtained in the 36.1 pb^{-1} data set. The statistical uncertainties are corrected for the non-unity width of the pull distributions.

Obviously the uncertainties on the estimated results are dominated by the limited statistics which can be solved when larger samples of data events become available. The LHC runs also in 2011 at a center of mass energy of 7 TeV. So far, more than $2fb^{-1}$ of proton collisions have been collected by the CMS detector. One can then apply the method on this larger amount of real proton collisions to improve the results, where the statistical uncertainties would be reduced by a factor of $\sqrt{\frac{36}{2000}} \sim 0.1$, as the statistical uncertainty $\delta(stat.)$, decreases while the number of events N, increases $\delta(stat.) \sim \frac{1}{\sqrt{N}}$. This can be understood by looking at the plots in Figure 6.1, where the statistical as well as the systematical uncertainties on the light and b jet energy corrections are shown as a function of an increasing integrated luminosity.

In Figure 6.1, the straight lines represent the systematical uncertainties on the light and b jet energy corrections. It is clear that, at some point which in case of for example the light jet energy correction corresponds to $\int \mathcal{L}dt \sim 300\ pb^{-1}$, the systematical uncertainty starts to be dominated. Therefore, the total uncertainty would reach a certain amount which is determined by the systematical limits and the final estimation of the jet energy corrections cannot be determined better than this precesion. This is the limit of the applied calibration technique where the precision cannot be improved even if larger samples of collision data are recorded and used in the analysis. Also the best estimate of m_{top} from other measurements provides an uncertainty which translates into a limit on the precision of ΔE_b. The bias of the estimators has an intrinsic uncertainty due to the fits which are applied to determine the expected jet

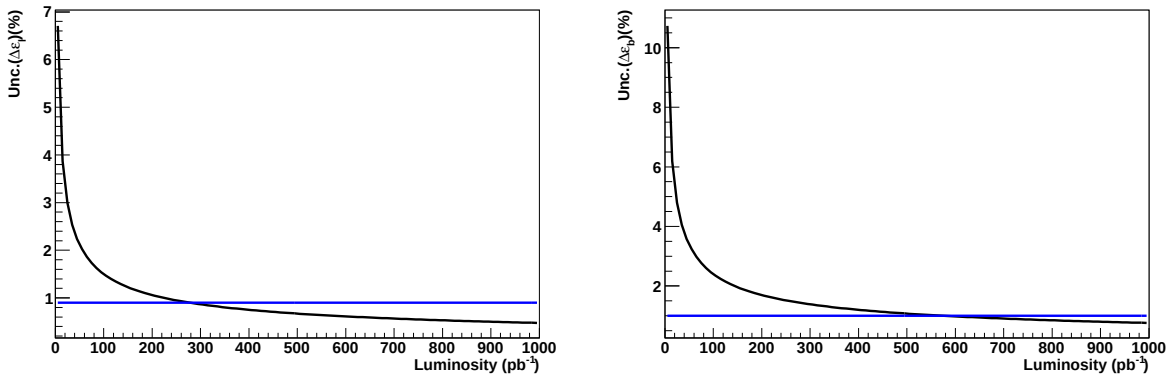

Figure 6.1: The absolute statistical (curved line) as well as the systematical (straight line) uncertainties on the (left) light and (right) b jet energy corrections as a function of the luminosity. From some point onward the systematical uncertainties dominate the statistical uncertainties.

energy correction values on simulation, also this is a limit of the precision of ΔE.

Although the method is able to estimate the jet energy corrections inclusively, it has also the capability to provide a differential estimation of the jet energy corrections. When larger amounts of statistics are recorded, the jets can be categorized in terms of their kinematics, such as pseudorapidity η or transverse momentum p_T. Then, the same method can be applied on the events which populate a particular bin of these kinematic variables and consequently the estimated results would be obtained for that region of the phase space. This yields to obtain a differential estimation of the jet energy corrections. With 36 pb^{-1} where a low number of events survived the selection cuts, a differential estimation is not feasible. Using a higher amount of statistics of for example $2fb^{-1}$, there are sufficient events remaining in the various regions of the phase space that makes the differential method applicable and would result in meaningful estimations. More detailed information on this aspect can be found in [106].

The calibration method explained in this thesis will help the search for new physics phenomena at the LHC, because when jets are better calibrated, the kinematic topology of the new physics events can be better reconstructed and differentiated from the backgrounds. Moreover, since in almost all the analyses, the jet energy scale calibration is the main source of systematical uncertainty, a precise determination of the energy of the jets would also reduce the uncertainties originating from the jet energy scale calibration factors and makes it possible to perform precision tests even at hadron colliders like the LHC.

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] CDF Collaboration, F. Abe *et al.*, Phys. Rev. Lett. **73** (1994) 225–231.

[4] CDF Collaboration, F. Abe *et al.*, Phys. Rev. Lett. **74** (1995) 2626–2631.

[5] D0 Collaboration, S. Abachi *et al.*, Phys. Rev. Lett. **74** (1995) 2632–2637.

[6] CDF Collaboration, F. Abe *et al.*, Phys. Rev. **D51** (1995) 4623–4637.

[7] CDF Collaboration, F. Abe *et al.*, Phys. Rev. Lett. **75** (1995) 3997.

[8] *The Tevatron Collider*, `http://www-bdnew.fnal.gov/tevatron/`.

[9] F. Englert and R. Brout, Phys. Rev. Lett. **13** (1964) 321–322.

[10] P. Higgs, Phys. Rev. Lett. **13** (1964) 508–509.

[11] G. Guralnik, C. Hagen, and T. Kibble, Phys. Rev. Lett. **13** (1964) 585–587.

[12] S. L. Glashow, J. Iliopoulos, and L. Maiani, Phys. Rev. D **2** (Oct, 1970) 1285–1292.

[13] ALEPH Collaboration, Phys. Rept. **427** (2006) 257–454.

[14] D. Schaile and P. M. Zerwas, Phys. Rev. D **45** (May, 1992) 3262–3265.

[15] LEP Collaboration, *A Combination of preliminary electroweak measurements and constraints on the standard model*, `hep-ex/0412015`.

[16] ALEPH Collaboration, J. Alcaraz *et al.*, *A Combination of preliminary electroweak measurements and constraints on the standard model*, `hep-ex/0612034`.

[17] T. E. W. Group, CDF, and D. Collaborations, *Combination of CDF and DO results on the mass of the top quark using up to 5.8 fb-1 of data*, `1107.5255`.

[18] The LEP Electroweak Working Group
http://lepewwg.web.cern.ch/LEPEWWG/.

[19] LEP Working Group for Higgs boson searches Collaboration, R. Barate *et al.*, Phys. Lett. **B565** (2003) 61–75.

[20] S. F. Novaes, *Standard model: An Introduction*, hep-ph/0001283.

[21] T. CDF, D. Collaborations, t. T. N. Phenomena, and H. W. Group, *Combined CDF and D0 Upper Limits on Standard Model Higgs Boson Production with up to 8.6 fb-1 of Data*, 1107.5518.

[22] *The CDF experiment*, http://www-cdf.fnal.gov/physics/physics.html.

[23] *The D0 experiment*, http://www-d0.fnal.gov/.

[24] M. Beneke *et al.*, *Top quark physics*, hep-ph/0003033.

[25] U. Baur, M. Buice, and L. H. Orr, Phys. Rev. **D64** (2001) 094019.

[26] N. Cabibbo, Phys. Rev. Lett. **10** (1963), no. CERN-TH-342, 531–533.

[27] M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **49** (1973) 652–657.

[28] *Particle Data Group*, http://pdg.lbl.gov/.

[29] O. S. Bruning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, *LHC Design Report*. CERN, Geneva, 2004.

[30] S. Myers and E. Picasso, Contemp. Phys. **31** (1990) 387–403.

[31] CMS Collaboration, D. G. d'Enterria, (Ed.) *et al.*, J. Phys. **G34** (2007) 2307–2455.

[32] L. Evans and P. Bryant, JINST **3 S08001** (2008).

[33] M. Benedikt, P. Collier, V. Mertens, J. Poole, and K. Schindl, *LHC Design Report*. CERN, Geneva, 2004.

[34] Particle Data Group Collaboration, K. Hagiwara *et al.*, Phys. Rev. **D66** (2002) 010001.

[35] *CMS Luminosity - Public Results*,
https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults2010.

[36] CERN, *Chamonix 2010 Workshop on LHC Performance*. CERN, Geneva, 2010.

[37] CERN, *Evian 2010 Workshop on LHC Commissioning*. CERN, Geneva, 2010.

[38] The CMS Collaboration, JINST **3 S08004** (2008).

[39] The ATLAS Collaboration, JINST **3 S08003** (2008).

[40] The ALICE Collaboration, JINST **3 S08002** (2008).

[41] The LHCb Collaboration, JINST **3 S08005** (2008).

[42] The LHCf Collaboration, JINST **3 S08006** (2008).

[43] The TOTEM Collaboration, JINST **3 S08007** (2008).

[44] M. Della Negra, L. Foà, A. Hervé, and A. Petrilli, *CMS Physics Technical Design Report Volume I: Detector Performance and Software*. Technical Design Report CMS. CERN, Geneva, 2006.

[45] V. Karimaki, *The CMS tracker system project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.

[46] *The CMS tracker: addendum to the Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 2000.

[47] *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.

[48] P. Bloch, R. Brown, P. Lecoq, and H. Rykaczewski, *Changes to CMS ECAL electronics: addendum to the Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 2002.

[49] F. Ferri and P. Govoni, *The CMS Electromagnetic Calorimeter Pre-calibration with Cosmic Rays and Test Beam Electrons*, Tech. Rep. CMS-CR-2007-012. CERN-CMS-CR-2007-012, CERN, Geneva, 2007.

[50] P. Adzic *et al.*, JINST **2** (2007) P04004.

[51] *The CMS hadron calorimeter project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.

[52] CMS HCAL Collaboration, S. Abdullin *et al.*, Eur. Phys. J. **C55** (2008) 159–171.

[53] *The CMS muon project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.

[54] G. L. Bayatyan *et al.*, *CMS TriDAS project: Technical Design Report; 1, the trigger systems*. Technical Design Report CMS.

[55] S. Cittolin, A. Rcz, and P. Sphicas, *CMS trigger and data-acquisition project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 2002.

[56] G. L. Bayatyan, M. Della Negra, Fo, A. Herv, and A. Petrilli, *CMS computing: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 2005. Submitted on 31 May 2005.

[57] C. Eck *et al.*, *LHC computing Grid: Technical Design Report. Version 1.06 (20 Jun 2005)*. Technical Design Report LCG. CERN, Geneva, 2005.

[58] I. Antcheva *et al.*, Comput. Phys. Commun. **180** (2009), no. 12, 2499–2512.

[59] *Physics Analysis Toolkit (PAT)*, `https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookPAT`.

[60] *CMS Belgian T2 Wiki*, `http://mon.iihe.ac.be/trac/t2b/wiki`.

[61] T. S. Virdee, *Calorimetry. oai:cds.cern.ch:529421*, Tech. Rep. CMS-CR-1998-026, CERN, Geneva, 1999.

[62] CMS Collaboration, The CMS Collaboration, CMS AN-2008/097 (2008).

[63] S. Baffioni, C. Charlot, F. Ferri, D. Futyan, P. Meridiani, I. Puljak, C. Rovelli, R. Salerno, and Y. Sirois, *Electron reconstruction in CMS*, Tech. Rep. CMS-NOTE-2006-040. CERN-CMS-NOTE-2006-040, CERN, Geneva, Feb, 2006.

[64] E. Meschi, T. Monteiro, C. Seez, and P. Vikas, *Electron Reconstruction in the CMS Electromagnetic Calorimeter*, Tech. Rep. CMS-NOTE-2001-034, CERN, Geneva, Jun, 2001.

[65] M. Pioppi *et al.*, *Electron Pre-identification in the Particle Flow framework*, CMS AN-2008/032.

[66] W. Adam *et al.*, *Electron Reconstruction in CMS*, CMS AN-2009/164.

[67] The CMS Collaboration, *Jet Performance in pp Collisions at 7 TeV*, CMS PAS JME-10-003.

[68] V. Chetluru *et al.*, *Jet Reconstruction Performance at CMS*, CMS AN-2009/067.

[69] The CMS Collaboration, *Performance of Jet Algorithms in CMS*, CMS PAS JME-07-003.

[70] A. Heister *et al.*, *Measurement of Jets with the CMS Detector at the LHC*, CMS AN-2005/053.

[71] P. Schieferdecker *et al.*, *Performance of Jet Algorithms in CMS*, CMS AN-2008/001.

[72] M. Cacciari, G. P. Salam, and G. Soyez, JHEP **04** (2008) 063.

[73] The CMS Collaboration, *Plans for Jet Energy Corrections at CMS*, CMS PAS JME-07-002.

[74] S. Esen *et al.*, *Plans for Jet Energy Corrections at CMS*, CMS AN-2007/055.

[75] R. Harris *et al.*, *MC Truth L2 & L3 Factorized Jet Corrections at CMS*, CMS AN-2008/003.

[76] The CMS Collaboration, *Measurement of the Jet Energy Resolutions and Jet Reconstruction Efficiency at CMS*, CMS PAS JME-09-007.

[77] S. Beauceron *et al.*, *Jets-Electron Cleaning via CaloTowers Selection*, CMS AN-2009/149.

[78] S. Esen *et al.*, E_T^{miss} *Performance in CMS*, CMS AN-2007/041.

[79] M. Dobbs and J. B. Hansen, Computer Physics Communications **134** (2001), no. 1, 41 – 46.

[80] R. K. Ellis, W. J. Stirling, and B. R. Webber, *QCD and Collider Physics*, Cambridge University Press (2003) 435 p.

[81] *The Coordinated Theoretical-Experimental Project on QCD*, `http://www.phys.psu.edu/~cteq/`.

[82] *Online Plotting and Calculation of Parton Distributions*, `http://hepdata.cedar.ac.uk/pdf/pdf3.html`.

[83] T. Stelzer and W. F. Long, Comput. Phys. Commun. **81** (1994) 357.

[84] M. Mangano *et al.*, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, `arXiv:hep-ph/0206293`.

[85] Y. L. Dokshitzer, Sov. J. Phys. JETP **46** (1977) 641.

[86] V. N. Gribov and L. N. Lipatov, Sov. J. Nucl. Phys. **15** (1972) 438.

[87] G. Altarelli and P. G, Nucl. Phys. **B126** (1977) 298.

[88] T. Sjöstrand, S. Mrenna, and P. Skands, JHEP **05** (2006) 026.

[89] G. Corcella *et al.*, JHEP **01** (2001) 010.

[90] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber, JHEP **0111** (2001) 063.

[91] F. Krauss, JHEP **0208** (2002) 015.

[92] 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`.

[93] 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`.

[94] P. Perez, *W W cross section and W branching ratios at* $\sqrt{s} = 183$ *GeV. Measurement of WW cross section and W branching ratios at LEP2 with the ALEPH detector*, Tech. Rep. CERN-OPEN-99-310. CERN-ALEPH-98-019, CERN, Geneva, Jul, 1998.

[95] A. Ealet, *WW cross sections and W branching ratios*,.

[96] *Standard Model Cross Sections for CMS at 7 TeV*, https://twiki.cern.ch/twiki/bin/viewauth/CMS/StandardModelCrossSections.

[97] *Reference Page for Top-Quark Analyses: Samples and Cross-Sections*, https://twiki.cern.ch/twiki/bin/view/CMS/CrossSections_3XSeries.

[98] *Electron Identification Based on Simple Cuts*, https://twiki.cern.ch/twiki/bin/view/CMS/SimpleCutBasedEleID.

[99] A. Rizzi, F. Palla, and G. Segneri, *Track impact parameter based b-tagging with CMS*, Tech. Rep. CMS-NOTE-2006-019. CERN-CMS-NOTE-2006-019, CERN, Geneva, Jan, 2006.

[100] *TMVA Toolkit for Multivariate Data Analysis with ROOT*, http://tmva.sourceforge.net/.

[101] *Top Quark Analysis Framework page: Jet-Parton Association Algorithms*, https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideTQAFLayer2.

[102] J. M. Bauer, *Kinematic Fit for the Radiative Bhabha Calibration of BaBar's Electromagnetic Calorimeter*, Tech. Rep. hep-ex/0011019. SLAC-PUB-8650. UMS-HEP-2000-09, SLAC, Stanford, CA, Nov, 2000.

[103] J. D'Hondt, S. Lowette, *et al.*, *Fitting of Event Topologies with External Kinematic Constraints in CMS*, CMS Note-2006/023.

[104] *Reference Selection for the $t\bar{t} \to e + jets$ Channel*, https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopLeptonPlusJetsRefSel_el.

[105] F. Bostock *et al.*, *Plans for measurement of $t\bar{t}$ cross section in the semileptonic electron channel with 20* pb^{-1} *of data at* $\sqrt{s}$ *of 10 TeV*, CMS AN-2009/075.

[106] P. Van Mulders, *Calibration of the jet energy scale using top quark events at the LHC*, CMS TS-2010/028.

Summary

Jets appear in the final state of many processes of interest at the Large Hadron Collider and a precise measurement of their energies is of great importance when searching for new physics phenomena. Since the reconstruction of the jet objects at hadron colliders is a challenging task, calibration strategies are developed to compensate for the non-perfection of the jet energy measurements. Historically, jets at hadron colliders are calibrated using QCD dijet events. The resolution on the jet energy measurement can be further improved considering the physics constraints which can be imposed on an observed top quark pair system. Due to the precise measurement of the top quark mass at the Tevatron collider, top quarks can be used as a calibration tool to improve the performance of the diverse analyses. Imposing the top quark and W boson mass constraints on the system containing a top quark decaying to jets, and altering the jet energies within their uncertainties, one can obtain the residual jet energy calibration factors.

In this thesis the method to estimate the residual jet energy corrections is applied on the events where top quark pairs are produced in proton-proton collisions and subsequently decay to an isolated electron and at least four reconstructed jets and missing transverse energy, hence $pp \to t\bar{t} \to W^+W^-b\bar{b} \to e\nu_e q\bar{q}b\bar{b}$ processes. The leptonic side of the event topology is mainly used to suppress the QCD multijet background events and the hadronic side is used for the calibration purpose. Event selection cuts are applied to enhance the ratio of signal to background events to about $S/B \simeq 2.5$, which is obtained based on a sample of simulated proton collisions corresponding to $100pb^{-1}$ of integrated luminosity at a center of mass energy of 7 TeV. In order to reconstruct the topology of the hadronic decay of the top quark in e+jets $t\bar{t}$ events, it is required to assign the three jets arising from the hadronic top to the three partons. A Multi-Variate Analysis method is used to perform the association of the jets to the quarks. In about 17% of e+jets $t\bar{t}$ events, the MVA method is able to return the correct combination, where a matching between jets and partons is found. Applying an event-by-event kinematic fitting technique with Lagrange multipliers, the masses of the W boson and of the top quark are used to constrain the e+jets $t\bar{t}$ events in the hadronic branch. The kinematic fit is applied in a two-dimensional range of residual jet energy corrections $(\Delta E_l, \Delta E_b)$. The points $(\Delta{\varepsilon_l}^{est}, \Delta{\varepsilon_b}^{est})$ minimizing the χ^2 of the fit are the estimated residual jet energy corrections. The precision of the method to estimate the residual jet energy corrections is limited to a systematical uncertainty of about 0.9% for light quark jets and of about 1.0% for b quark jets. The main source of the systematical uncertainty on the estimation of the b quark jet energy correction originates from the uncertainty on the measured top quark mass provided by the Tevatron collider.

Also the method of estimating the residual jet energy calibration factors has been applied on the real proton collisions recorded by the CMS experiment in 2010. The estimated results on the simulation and real proton collisions are summarized below. There is a good agreement between simulation and data and the residual corrections are found to be equal.

	simulation	data
$\Delta\varepsilon_l^{est}(\%)$	-8.1±2.4(stat.)±0.9(sys.)	-8.6±2.7(stat.)
$\Delta\varepsilon_b^{est}(\%)$	-3.2±3.5(stat.)±1.0(sys.)	-5.0±3.9(stat.)

Table 6.2: The final estimated residual jet energy corrections obtained in the $36.1 pb^{-1}$ data set. The statistical uncertainties are corrected for the non-unity width of the pull distributions.

Compared to the residual calibration factors provided by other data driven methods such as QCD dijet events, the gain that can be obtained by applying the calibration method of this thesis is twofold. A first one is that the calibration method described in this analysis is able to provide the residual jet energy corrections which is differentiated with respect to the flavour of quarks, namely light quark jets and b quark jets. A second one is that, while the methods based on QCD dijet events determine the residual jet energy corrections and correct the energy of the jets back to the generator level jets, the method which is described in this thesis is able to calibrate the energy of the jets back to the energy of the initiating quarks, instead of the generator level jets.

Samenvatting

Hadronische jets komen voor in de meest relevante proton-proton botsingen nabij de Large Hadron Collider te CERN, en een precieze meting van hun energie is cruciaal in de zoektocht naar nieuwe fysica fenomenen. De reconstructie van jet objecten nabij hadron versnellers is een uitdaging. Calibratietechnieken worden ontwikkeld om de meting van de detector te corrigeren. Traditioneel worden de jets gecalibreerd door middel van QCD botsingen met twee jets in de eindtoestand. De resolutie van de meting van de jet energie correcties kan significant verbetert worden door gebruik van de fysica verbanden aanwezig in top quark paar gebeurtenissen. Met behulp van de precieze meting van de top quark massa door de Tevatron experimenten, kunnen top quark gebeurtenissen gebruikt worden voor de calibratie van diverse analysemethoden. Rekening houdend met de onzekerheden op de gemeten kinematische eigenschappen van de jets, kan men jet energie correcties schatten die ervoor zorgen dat de fysica verbanden in top quark paar gebeurtenissen uitkomen.

In deze thesis is een methode toegepast die de correcties op de gemeten jet energie in te schatten, en dit met behulp van top quark paar processen waar een gesoleerd elektron in voorkomt, alsook ten minste vier jets en ontbrekende energie, namelijk het proces $pp \to t\bar{t} \to W^+W^-b\bar{b} \to e\nu_e q\bar{q}b\bar{b}$. De leptonische kant van de geobserveerde gebeurtenis is voornamelijk gebruikt om in de selectie procedure van deze topologie een onderscheid te maken met de enorme mutli-jet QCD achtergrond. De hadronische kant op zijn buurt is gebruikt voor het schatten van de energie correcties op de jets. Een selectie procedure is opgesteld en toegepast voor een dataset met een gentegreerde luminositeit van $100pb^{-1}$ bij een massamiddelpuntsenergie van 7 TeV, en resulterende in een verhouding van signaal ten opzichte van achtergrond van $S/B \simeq 2.5$. Om de topologie van de hadronisch vervallende top quark te reconstrueren moeten we drie van de geobserveerde jets in de eindtoestand identificeren met de topologie $t \to Wb \to qqb$. Hiervoor werd een Multi-Variate Analyse methode gebruikt die in ongeveer 17% van de geselecteerde signaal gebeurtenissen de correcte keuze maakt indien een juiste keuze mogelijk was. Doormiddel van een kinematische fit techniek toe te passen op elke gebeurtenis werden met Lagrange variabelen de massa verbanden van het W boson en de top quark opgelegd aan de geobserveerde gebeurtenis. Deze kinematische fit werd toegepast in een twee-dimensionale parameter ruimte $(\Delta E_l, \Delta E_b)$, namelijk die van de twee jet energie correcties voor zowel jets van lichte quarks en jets van b-quarks. De punten in deze twee-dimensionale ruimte die de gecombineerde χ^2 van de fit over alle gebeurtenissen minimaliseren, komen overeen met de geschatte jet correcties. De precisie op deze jet correcties heeft een ondergrens door systematische onzekerheden die 0.9% en 1.0% bedragen voor respectievelijk lichte en b-quarks. De belangrijkste

bron van systematische onzekerheid in de schatting van de jet correcties voor de b-quark jets is de onzekerheid op de top quark massa zoals gemeten door de Tevatron experimenten.

De methode is toegepast op de eerste proton botsingen gecumuleerd en gereconstrueerd door het CMS experiment in 2010. De bekomen resultaten worden weergegeven in onderstaande tabel. Men kan een goede overeenkomst waarnemen tussen de verwachte correcties uit simulatie en de geschatte correcties met de botsingsdata.

	simulation	data
$\Delta\varepsilon_l^{est}(\%)$	-8.1±2.4(stat.)±0.9(sys.)	-8.6±2.7(stat.)
$\Delta\varepsilon_b^{est}(\%)$	-3.2±3.5(stat.)±1.0(sys.)	-5.0±3.9(stat.)

Table 6.3: De resulterende jet energie correcties bekomen met een dataset van $36.1pb^{-1}$. De onzekerheden zijn gecorrigeerd om een correcte pull verdeling te bekomen.

In een vergelijking met jet correcties bekomen door middel van andere technieken, zoals QCD dubbel-jet gebeurtenissen, is de methode toegepast in deze thesis beter op twee manieren. Ten eerste bekomen we met deze techniek jet correcties voor zowel jets komende van lichte en b-quark jets afzonderlijk. Ten tweede hebben we met deze methode jet correcties die de geobserveerde jet energie calibreert naar de energie van de originele quark en niet enkele naar het zogenaamde deeltjes niveau na hadronisatie.

Acknowledgements

There are many different people that I would like to acknowledge. First of all, I would like to thank the IPM (Institute for Research in Fundamental Sciences, Tehran, Iran) which gives me the opportunity to perform my PhD thesis in the experimental particle physics while being based at CERN. Herewith, I would like to thank Professor Hessamaddin Arfaei, the head of Particle and Accelerator School at IPM, who provides me this great chance for which I would be grateful forever.

While I was mainly belonged to the IUT (Isfahan University of Technology, Isfahan, Iran), a joint-PhD program between my home-university and VUB (Vrije Universiteit Brussel, Brussels, Belgium), was established. This contract was agreed upon by the professors from both universities whose efforts on this aspect are really appreciated. Concerning this, I would like to thank Dr. Mohammad Esmaeil Hamedani Golshan (the head of the university), Dr. Ahmad Shirani (the head of the physics department), Dr. Mansour Haghighat (my promotor from IUT) and Dr. Farhad Shahbazi (the person in charge of graduate students) for providing this joint-PhD program. I would also like to give my special thanks to Professor Jorgen D'Hondt (my promotor from VUB). It is indeed a great honor for me to get a PhD degree under his supervision. I would like to thank Roger Wolf for his great helps when I was new to this exciting field. I never forget Stephanie Beauceron for her patience and comments on my job. I also appreciate 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 IPM that helped me in various ways. I would like to thank Dr. Mojtaba Mohammadi Najafabadi, Dr. Mohsen Khakzad, Dr. Majid Hashemi, Dr. Saied Paktinat and Dr. Batool Safarzadeh for the scientific discussions we have always had in our national weekly meetings. I am grateful to my Iranian colleagues at CERN, namely Nadjieh Jafari, Hamed Bakhshian, Ali Fahim and Rohollah Mohammadi, for their help in computational aspects and our nice physics discussions.

I also would like to thank the people in Top Brussels group, specially Petra Van Mulders, Stijn Blyweert and Michael Maes, for their patience and great helps in fixing my computational issues and in providing the codes on which my main analysis is based.

And finally I would also like to thank my family for their love and patience. I am really happy that my PhD period is coming to end with lots of good memories. It is indeed a pleasure that I had the opportunities to meet great people and make nice friends during this period.