Promotor: Prof. Dr. Jorgen D'Hondt

Departement Natuurkunde

Faculteit Wetenschappen

Vrije Universiteit Brussel

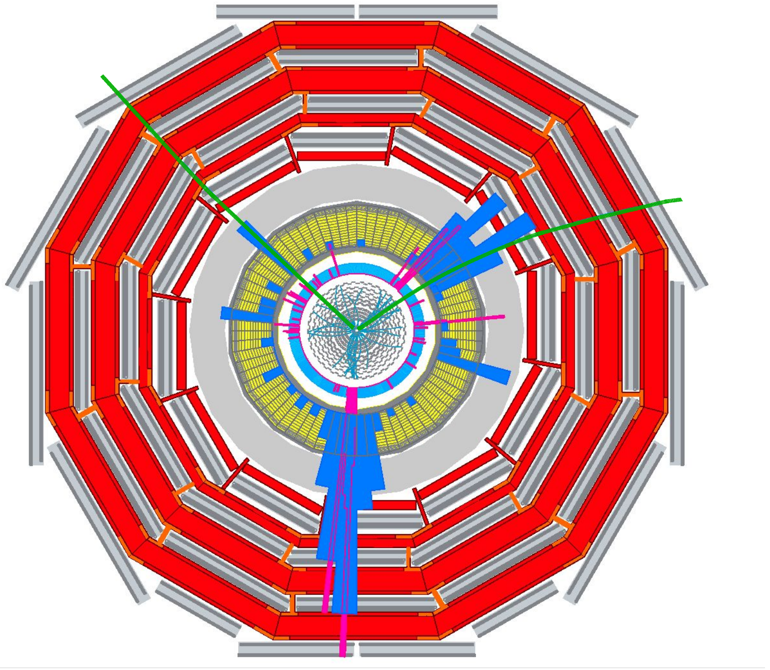
11 april 2008

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

Jan Heyninck – The CMS Top Quark Physics Potential with 1 fb^{-1} of Data – 2008

Vrije Universiteit Brussel

The CMS Top Quark Physics Potential with 1 fb^{-1} of Data



Jan Heyninck

Cover illustration: Visualisation of the reconstructed particles in the transverse plane of the CMS detector, resulting from a semi-muonic decaying $\mathrm{t\bar{t}}$ event. Four jets with high transverse energy, an isolated high p_T muon and a muon produced in the decay of a b-quark jet can be distinguished.

Vrije Universiteit Brussel

Faculteit Wetenschappen
Departement Natuurkunde

The CMS Top Quark Physics Potential with $1\,\mathrm{fb}^{-1}$ of Data

Jan Heyninck

Promotor: Prof. Dr. Jorgen D'Hondt

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

April 2008

Doctoral examination commission

Chair: Prof. Dr. Stefaan Tavernier (VUB)
Supervisor: Prof. Dr. Jorgen D'Hondt (VUB)
Prof. Dr. Roberto Tenchini (Universita di Pisa)
Dr. Roberto Chierici (IPNL)
Prof. Dr. Ben Craps (VUB)
Prof. Dr. Robert Roosen (VUB)
Prof. Dr. Alex Hermanne (VUB)

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

Het mooiste wat we kunnen meemaken,
is het raadselachtige.
Daar komt alle ware kunst en wetenschap uit voort.

Albert Einstein

Contents

Acknowledgements

The thesis you are holding in your hands is the result of an exciting journey of four years through the unique field of elementary particle physics. Firstly, I would like to thank the Inter-University Institute for High Energies (I.I.H.E.) and IWT-Vlaanderen for offering me this opportunity. I also would like to thank the members of the jury for their valuable comments that certainly improved the quality of the manuscript. In addition, there are however several people who deserve a special word of gratitude for their support and friendship I enjoyed during my PhD period.

As soon as I arrived at the I.I.H.E., two highly motivated colleagues were eager to guide me around: my promotor Jorgen D'Hondt and Steven Lowette. I think all three of us can look back with great satisfaction not only on our physics output, but especially on the naturalness, fellowship and gastronomy that came along with it. Jorgen, it was a pleasure working with you; I am sure your energy and enthusiasm will still inspire a lot of other young people! The last two years of my PhD I had the pleasure to witness the extension of our 'top quark team' by five great people. Many thanks Gregory, Ilaria, Joris, Petra and Volker for the nice time together! It would be great if you can keep me informed about the progress of the CMS experiment. A big thanks is also in order to Rosine, Daisy and Marleen for always offering that helping hand, and to Prof. H. Eisendrath, for giving me the chance to contribute to the exhibition 'Einstein, anders bekeken'.

Because of the international character of the research, I had the pleasure to meet many other interesting people. Participating in conferences, workshops or a stay at CERN I always conceived as a perfect example of an open-minded international collaboration in its purest form. Many thanks to all of you for the great insights and conversations, whether it be about physics or about other mind broadening topics.

In addition, there are of course several other friends that deserve some special attention. Daan, Sofie, Kevin, Salua, Severine, Pieter, Amina, Bram, Evi, ... many thanks for your support, and the many hours of relaxing rock climbing. Also congratulations to my office mate Cedric, who survived four years of almost living together with me.

Last but not least, it is difficult to overemphasise my gratitude towards the people closest to my heart. Eileen, your love and patience were an invaluable aid in the realisation of this PhD. I will do my ultimate best to create a similar cocoon of support around you as the one surrounding me during the redaction of my thesis. Finally, I would like to mention my parents for their belief in me and for always being there for me. Thanks, and enjoy this moment, as this is also your merit!

Introduction

The behaviour of all known subatomic particles can be described within a single theoretical framework called the Standard Model (SM). This quantum field theory incorporates the quarks and leptons as well as their interactions through all fundamental forces except gravity (the strong and the weak nuclear force and electromagnetism). The constituents of matter are classified in the Standard Model in three generations of fermions (quarks and leptons), which only differ from one another by an increased mass. The forces and their corresponding force-carrying particles are introduced via the requirement for the theory to be invariant under local gauge transformations. In the Standard Model, particles are proposed to acquire their mass via the interaction with the Higgs boson field, that is created as a result of a non-zero vacuum expectation value and the principle of spontaneous symmetry-breaking. Today, however, the Higgs boson has not yet been observed.

The detection of the Higgs boson's signatures is one of the main goals of the upcoming experiments at the Large Hadron Collider (LHC). The LHC is a proton-proton collider currently being completed in the former LEP tunnel at the CERN Laboratory near Geneva, Switzerland. The LHC will provide two proton beams, circulating in opposite directions, at an energy of 7 TeV each. The resulting centre-of-mass energy of $\sqrt{s} = 14\,\text{TeV}$ and the increased collision rate allow for the first time to study directly and in detail the TeV scale region. Apart from elucidating the electroweak symmetry breaking mechanism, the LHC might provide evidence of physics beyond the Standard Model. The study of the heaviest of all quarks, the top quark, is generally considered to be a highly sensitive window for this search of new physics. Also, a better knowledge of the top quark properties such as its mass and the $t\bar{t}$-production cross-section will result in an ultimate confinement test of the Standard Model. Scenario's in which a special role of the top quark is predicted or models resulting in anomalies on the top quark production rate or branching ratios will be accessible experimentally. At the LHC, every second several $t\bar{t}$-pairs are expected to be produced. For the first time, this huge statistics of top quark events will be exploited to commission and calibrate the detectors.

All physics analyses presented in this thesis envisage the estimation of the physics potential with $1\,\text{fb}^{-1}$ of data accumulated by the multi-purpose Compact Muon Solenoid (CMS) detector. The extreme dimensions of the 4 T solenoid allow to incorporate the electromagnetic and hadronic calorimeter, what leads to an improved energy resolution. Additionally, CMS is characterised by its excellent muon spectrometer and trigger system, and the high hermiticity of the calorimeters, what allows a relatively accurate determination of the transverse energy of non-interacting particles.

This thesis starts with an introduction of the Standard Model and the top quark sector as part of it. The importance of this particle is emphasised, and the today's knowledge of its main properties such as its mass and production cross-section is summarised. Because the LHC starts to collide protons in the summer of 2008, all physics analyses are based on the study of Monte-Carlo generated event samples. A second chapter elaborates on the different steps that can be distinguished in the simulation of a proton-proton collision. The choice for the multi-parton leading order matrix element generators such as the `AlpGen` program is motivated. Next, the CMS experiment and the LHC is described in Chapter 3, while Chapter 4 concentrates on the reconstruction aspects and selection of physics objects such as tracks, muons, jets, ... The issue of jet reconstruction and the calibration of the reconstructed jet energy is emphasised, and a method to compare the jet reconstruction performance of several jet clustering algorithms is introduced. The object selection requirements defined in Chapter 4 will be applied in all following physics analyses. Also, only semi-muonic decaying $t\bar{t}$-events are considered as signal events. A first analysis, presented in Chapter 5, illustrates the use of top quark events in the determination of the jet energy scale. Next, in respectively Chapter 6 and Chapter 7 the estimators of the $t\bar{t}$ production cross-section and the top quark mass are introduced. In both analyses extra event selection criteria are defined, and an extended study of the systematic uncertainties is given. Finally, the main conclusions and lessons to take away from this thesis are summarised in a last chapter.

Chapter 1

Top quark Physics as a part of the Standard Model

Elementary particle physics research is the quest for understanding the smallest constituents of matter and their interactions. Over the last 60 years, great breakthroughs are accomplished in this field as a result of a fruitful interplay between improved experimental data-taking and ingenious theoretical insights which allowed an elegant description of the observed phenomena. In the early seventies a model known as the 'Standard Model (SM)' was derived able to describe three of the four known fundamental interactions between the elementary particles that make up all matter: the electromagnetic, the weak and the strong force [1, 2]. The Standard Model has yet to be disproved by any experimental test. This model will be further elaborated upon in the first section of this chapter.

One of the particles predicted by the Standard Model and experimentally observed only 13 years ago is the top quark. As this work fully concentrates on the measurement of top quark properties from data collected by the CMS experiment at the Large Hadron Collider (Chapter 3), a second section will focus on the current experimental knowledge and the theoretical importance of this particle.

1.1 The Standard Model of Elementary Particles

The dynamics of both matter and energy in nature are presently best understood in terms of the kinematics and interactions of elementary particles. To date, science has managed to reduce the numerous amount of laws introduced in all of its disciplines to a small core of fundamental laws and theories. One of the main goals of physics is to find the 'common ground' that would unite all of these into one integrated model of everything, in which all the other laws we know of would be special cases, and from which the behaviour of all matter and energy can be derived. As we will illustrate, this idea of unification is somehow indicated by the observations and has already proven to be a very powerful guidance in the construction of the Standard Model itself.

In a first subsection, an overview of all elementary particles comprised in the Standard Model is given. A differentiation is made between matter particles and particles responsible for the mediation of the fundamental forces. The somewhat

	1^{st} generation	2^{nd} generation	3^{rd} generation
quarks	up u down d	charm c strange s	top t bottom b
leptons	electron neutrino ν_e electron e	muon neutrino ν_μ muon μ	tau neutrino ν_τ tau τ

Table 1.1: Overview of the fermions building up the Standard Model

more mathematical description of the way interactions among the elementary particles are described in the SM is the subject of paragraph 1.1.2. In this subsection the $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge group building up the Standard Model and the importance of local gauge invariance are introduced. The Standard Model is a grouping of two major theories, the electroweak theory (EW) and quantum chromodynamics (QCD). Masses are generated in the SM according to the Higgs mechanism, summarised in paragraph 1.1.3. Despite the success of the Standard Model, this theory cannot completely describe elementary particles. An overview of its main shortcomings, which indicate a more comprising theory should exist, is given in paragraph 1.1.4. Finally, some possible extensions of the SM such as supersymmetry (SUSY) are briefly discussed in a last paragraph of this section.

1.1.1 Fundamental Particles and Forces

According to the Standard Model, all matter, whether it is observed in our daily life, in galaxies at the edge of our universe or created as the result of high energy collisions, is built from only 12 elementary matter particles or fermions. They all have spin-1/2, which means they obey the Pauli exclusion principle[1]. Each of these fermions has its anti-particle, which has the same mass, but opposite electrical charge, and is denoted with a bar over the particle's symbol. The 12 fermions comprise three generations of leptons and quarks. Each extra generation is an exact copy of the first generation, apart from the increasing particle mass. The particles filling in the first generation are well known, as they are the building blocks of all matter surrounding us. The up and down quark build up the protons and neutrons, and consequently the nuclei of all chemical elements. Electrons neutralise these nuclei to form the atoms and the chemical bounds in molecules. Finally, the electron neutrino might be observed in the decay of radioactive isotopes. An overview of all fermions is given in Table 1.1. The up-type quarks have electrical charge $2/3\,e$, while down-type quarks have charge $-1/3\,e$ (with $-e$ the electrical charge of the electron). Neutrino's are neutral.

Apart from the elementary building blocks of matter, the SM introduces force carrying spin-1 particles, called bosons, that mediate the interactions between fermions. Where the photon is responsible for the electro-magnetic (EM) force, three massive particles (the charged $W^{\pm}$ and the neutral Z^0-boson) mediate the weak nuclear force. Eight gluons finally carry the strong force. All bosons introduced in the Standard Model are listed with their respective mass in Table 1.2.

---

[1] This principle states that one and only one fermion can be found in a given quantum state.

force	force carrier	mass (GeV/c^2) [3]
electro-magnetism	photon γ	0
weak	$W^{\pm}$; Z^0	80.40; 91.19
strong	eight gluons g	0

Table 1.2: Overview of the bosons, and their measured masses, responsible for the mediation of the three fundamental forces comprised in the Standard Model.

Because photons are massless particles, the EM force has an infinite range and interacts with all particles carrying electric charge. The coupling strength of the EM interaction is equal to the fine structure constant $\alpha_{EM} \simeq 1/137$ at low energies, but increases as the energy increases[2]. The theory of EM interactions was independently formulated first as Quantum Electrodynamics (QED) and showed a remarkable agreement with the experimental observations. Since QED itself is a unified theory of electricity and magnetism, the road was set to include another type of force.

This was accomplished independently by Weinberg [4] and Salam [5] via the construction of the electroweak theory. In this model, the photon and the three weak bosons are the physical manifestation of the four gauge fields generated by the $SU(2)_L \times U(1)_Y$ local gauge group. It has been observed that the weak force only interacts with the left-handed fermions [3]. This indicates that the universe has a preference for a given chirality, and that the conservation of parity is broken.

Together with the electroweak theory, also the theory of the strong interaction, Quantum Chromodynamics (QCD), is successfully incorporated and described in the Standard Model, although it is not unified with the EW theory. Of all fermions only the quarks interact via this force. Characteristic to the strong interaction is the particular property that the coupling constant decreases with energy, a phenomenology which is often referred to as ‘asymptotic freedom’. As a consequence, the quarks are only observed as bounded or confined states called hadrons. Adding extra energy to separate the quarks in a hadron would only result in the production of extra quark/anti-quark pairs created from the increasing energy density in the intermediate gluon field. In QCD, these phenomena are explained in terms of ‘colour’-charge. Each quark has one of the three allowed colours, which are traditionally chosen blue, red and green, while an anti-quark receives an anti-colour. Only colour neutral states or colour singlets are observed as free particles. These are formed either by the combination of a colour and an anti-colour, or via the addition of one of each (anti-)colours in a three quark state. Both types of hadrons are respectively known as mesons and baryons. The QCD mediators or gluons are also colour charged, and will hence interact with one another.

Finally, it should be noted that although the Standard Model combines all three forces together in one framework it is not fully a true unified theory in that it is not based on a single representation governed by one coupling constant. Such a unification might however be possible at the ‘Grand Unified Theories’ or GUT scale of 10^{16} GeV.

---

[2] This effect is known as ‘running coupling constants’.

[3] Although chirality is only identical to helicity for massless particles, the same nomenclature is traditionally used. Right-handed fermions have a spin lined up to the direction of motion, while for left-handed particles the scalar product of both vectors is negative.

1.1.2 The massless Standard Model Lagrangian

Behind these particles and forces lies the elegant mathematics of the Standard Model. Since these sub-atomic elementary particles are very small and tend to travel with velocities close to the speed of light, the Standard Model is described in terms of a quantum field theory, which incorporates both quantum mechanics and special relativity. In such a theory each physical process corresponds to an interaction of particle fields.

Quantum field theories are based on Hamilton's principle of least action S,

$$\delta S = \delta \left(\int \mathcal{L}(\psi(x), \partial^\mu \psi(x))\, d^\mu x \right) = 0\,, \tag{1.1}$$

with $\mathcal{L}$ the Lagrangian density depending on the particle's field wave functions $\psi(x)$ and their first derivatives $\partial^\mu \psi(x)$.

Dirac spinors

In the Standard Model, the wave function $\psi(x)$ of all fermions is represented by Dirac spinors, for which the Lagrangian function becomes:

$$\mathcal{L}^{Dirac} = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi \ . \tag{1.2}$$

The requirement of least action results in the so-called Euler-Lagrange equations which translates for the above Lagrangian in the well known Dirac equation[4]. Four solutions exist, of which two correspond to the two spin states of the fermion [7]:

$$\psi(p) = \sqrt{E+m} \begin{bmatrix} \phi \\ \frac{\vec{\sigma}.\vec{p}}{E+m}\,\phi \end{bmatrix}, \tag{1.3}$$

with $\phi^T = [1\ 0]$ for spin up, $\phi^T = [0\ 1]$ for spin down and $\vec{\sigma}$ the Pauli matrices[5]. The other two solutions define the two spin states of the anti-fermion.

Gauge invariant theories

The Lagrangian written in Eq. 1.2 only describes the free moving fermions. To also include interactions between the matter particles, the Lagrangian is asked to be invariant to local phase transformations. Where global phase transformations yield to the extraction of conservation laws [6], the requirement for local gauge invariance will entail

---

[4] The Dirac equation is a relativistic quantum mechanical wave equation formulated by the British physicist Paul Dirac in 1928 [6], that provides a description of an elementary spin-1/2 particle, such as an electron. The equation also predicted the existence of anti-matter, which was confirmed with the discovery of the positron.

[5] The Pauli matrices are a set of 2×2 complex Hermitian and unitary matrices given by:

$$\sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \qquad \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{1.4}$$

[6] This is a direct consequence from Noether's theorem, which states that each symmetry of a physical system implies that some physical property of that system is conserved, and conversely that each conserved quantity has a corresponding symmetry [8].

the introduction of interaction terms that represent the coupling of the particle to the gauge bosons. In this way, the couplings between gauge bosons and fermions just falls out of the theory.

Suppose a fermion's wave function changes under a local phase transformation with rotation parameters $\vec{\epsilon}(x)$ in an internal space, characterised by the generators $\vec{\tau}$ of a given Lie-group, as

$$\psi' = U\psi = e^{i\vec{\epsilon}(x)\cdot\frac{\vec{\tau}}{2}}\psi, \tag{1.5}$$

such that all quantum-mechanical observables that depend on $|\psi|^2$ remain invariant. In general however, the Dirac Lagrangian is not invariant under such a transformation. To restore the symmetry in the theory, a covariant derivative is introduced:

$$\mathcal{D}_\mu = \partial_\mu - ig\frac{\vec{\tau}}{2}\vec{A}_\mu \tag{1.6}$$

with ∂_μ the usual space-time derivative and $\vec{A}_\mu$ a new interacting vector field. The factor g represents the interaction strength associated to the field. Substituting the covariant derivative into the Lagrangian of Eq. 1.2 gives

$$\begin{aligned}\mathcal{L}^{Dirac} &= i\bar{\psi}\gamma^\mu\mathcal{D}_\mu\psi - m\bar{\psi}\psi \\ &= i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - ig\bar{\psi}\gamma^\mu\frac{\vec{\tau}}{2}\vec{A}_\mu\psi,\end{aligned} \tag{1.7}$$

where the last term expresses the coupling between the fermion field and the new vector field. By requiring that

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

such that the Dirac Lagrangian is invariant under Eq. 1.5, the transformation relations for the components of the field $\vec{A}_\mu$ are derived to be

$$\frac{\tau^i}{2}A^{i\prime}_\mu = -\frac{i}{g}(\partial_\mu U)U^{-1} + U\frac{\tau^i}{2}A^i_\mu U^{-1}\,. \tag{1.9}$$

If the theory is made invariant under a local phase-space transformation of an Abelian group[7], only interactions between the fermion and the gauge field are obtained. In the case of a non-Abelian group, also interactions among the gauge bosons are expected. Hence, the couplings among the gauge bosons themselves are predicted by requiring that the non-Abelian gauge groups also satisfy the local gauge invariance.

The electroweak theory

Constantly guided by new experimental observations, the gauge symmetry group able to give an appropriate description of the observed electroweak phenomena was determined to be the $SU(2)_L \times U(1)_Y$[8] group. Requiring the Lagrangian to be gauge

---

[7] An Abelian group is defined as a group represented by commutating generators $\vec{\tau}$: $[\tau_i, \tau_j] = 0$.

[8] The subscript Y stands for the hypercharge which is defined as

$$Y = 2(Q - I_3) \tag{1.10}$$

where Q is the electromagnetic charge and I_3 is the third component of the weak isospin.

invariant towards local phase-space transformations of this group allowed to unify the weak nuclear force with the electromagnetic force, up to then described by Quantum Electrodynamics (QED). The EM force is characterised by unitary $e^{i\epsilon(x)}$ phase transformations in one dimension according to the $U(1)$ group symmetry. The weak force on the other hand is described by $SU(2)$. Consequently, it is convenient to group the fermions into doublets interacting under the weak force:

$$\Psi_{EW} = \begin{pmatrix} u \\ d \end{pmatrix}, \begin{pmatrix} c \\ s \end{pmatrix}, \begin{pmatrix} t \\ b \end{pmatrix}, \begin{pmatrix} \nu_e \\ e \end{pmatrix}, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix} \tag{1.11}$$

Each of these doublets corresponds to a field comprising two Dirac spinors. Any $SU(2) \times U(1)$ local phase-space transformation can be written as:

$$\Psi'_{EW} = e^{i\vec{\epsilon}(x).\vec{\sigma}}\ e^{i\theta(x)}\ \Psi_{EW}. \tag{1.12}$$

The covariant derivate which makes the Lagrangian invariant under these $SU(2)_L \times U(1)_Y$ gauge transformation takes the form

$$D_\mu = \partial_\mu + ig\frac{\vec{\sigma}}{2}\vec{W}_\mu + ig'\frac{Y}{2}B_\mu\ , \tag{1.13}$$

with $\vec{W}_\mu$ and B_μ the gauge fields associated to the $SU(2)$ and $U(1)$ group respectively. Hence, for the electroweak theory to be gauge invariant, one scalar gauge boson B^0 and three vector gauge bosons W^α (α=1,2,3) are required. The latter bosons can only couple to left-handed fermion doublets; right-handed fermion fields remain unchanged under the $SU(2)_L$ gauge transformation. This way the parity-violating nature of the weak interactions is incorporated in the theory.

The electroweak symmetry is not only expected to spoil parity, is must also be broken as a rather high mass is measured for its vector bosons. However, adding explicit mass terms to the Lagrangian would break the gauge invariance, and is hence no option. In the next paragraph an elegant solution known as spontaneous symmetry breaking will be introduced. It will be shown that with this electroweak symmetry breaking procedure, the mass terms for the more familiar physical states [9]

$$W^\pm_\mu = \sqrt{\frac{1}{2}}\left(W^1_\mu \mp iW^2_\mu\right) \tag{1.14}$$

$$\begin{aligned} Z^0_\mu &= W^3_\mu \cos\theta_w - B_\mu \sin\theta_w \\ A_\mu &= W^3_\mu \sin\theta_w + B_\mu \cos\theta_w\ , \end{aligned} \tag{1.15}$$

arise naturally from the Higgs mechanism. In the above equations, θ_w denotes the Weinberg mixing angle, defined as

$$\tan\theta_w = \frac{g'}{g} \tag{1.16}$$

---

[9] From the definition of the physical vector boson states in Eq. 1.15 it is clear that the $W^\pm$ bosons still couple only to the left-handed fermions. The Z^0-boson on the contrary can couple to right-handed fermions too because of its mixing with the B_μ-field.

Quantum chromodynamics

QCD is formulated in an analogue way to QED and the electroweak theory, in the sense that also the strong force mediating gluon fields are introduced in the theory by requiring gauge invariance. For this interaction the relevant gauge group is found to be the $SU(3)_C$ symmetry group, where the C subscript corresponds to the quark colour triplets. Restoring the gauge invariance of the theory with respect to local $SU(3)$ phase-space transformations invokes the introduction of eight gauge fields, corresponding to the eight $SU(3)$ group generators[10]. The non-Abelian character of the $SU(3)$ group will ensure that gluons self-interact, which is expected as gluons are colour charged themselves.

The observation of CP violation and processes violating the conservation of strangeness, are allowed in the Standard Model by the assumption that the strong force eigenstates of the quarks slightly differ from their weak force eigenstates. This mismatch of quantum states is given by the Cabibbo-Kobayashi-Maskawa (CKM) matrix:

$$\begin{pmatrix} d^{weak} \\ s^{weak} \\ b^{weak} \end{pmatrix}_L = \begin{pmatrix} V_{ub} & V_{us} & V_{ub} \\ V_{cb} & V_{cs} & V_{cb} \\ V_{tb} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L \tag{1.18}$$

This CKM matrix describes the probability of a transition from one quark q to another q' due to a flavour changing weak interaction, and is proportional to $|V_{qq'}|^2$. Because minimum three generations of fermions are required to have a CP violating complex phase in a unitary matrix[11] and CP violation was observed, this result was the start of a quest to the at that time undiscovered bottom and top quark.

Finally, it should be noted that in addition to the Lagrangian terms describing the free fermion propagation and the terms introducing the interactions of these fermions to the various gauge bosons, for each gauge field a gauge invariant kinetic term has to be introduced. These terms will allow the propagation of free gauge bosons.

Renormalisation of the theory

Despite the elegance and simplicity of the SM Lagrangian, the derivation of predictions from the theory is a highly non-trivial task. This is simplified by the introduction of Feynman diagrams and rules for their calculation, enabling a diagrammatic approach to calculations of probabilities associated to specific processes [9]. Quantum-mechanical corrections need to be accounted for when performing such calculations, which introduce extra loops and vertices in the Feynman diagrams. By ordering all the diagrams as a function of the number of vertices, a series is formed with increasing powers of the coupling constant. Such an expansion can be used for perturbative calculations,

---

[10] The eight generators of the $SU(3)$ group are the 3×3 Gell-Mann matrices T_a, which satisfy the commutation rule

$$[T_a, T_b] = i\, f_{abc}\, T_c \ , \tag{1.17}$$

where f_{abc} denote the structure constants of the group.

[11] In general, the number of complex phases for a unitary $N \times N$ matrix is given by $(N-1)(N-2)/2$.

provided the coupling constant is smaller than unity. In such calculations, however, one is confronted with divergences, even in the easiest case of electromagnetism. In the Standard Model, however, these infinities can always be absorbed into the unobservable bare parameters (such as the electric charge) of the Lagrangian by a technique called renormalisation [10]. Finite measured quantities would in general imply divergent bare quantities, and should be defined at a certain renormalisation scale. To minimise the contribution of loop diagrams to a given calculation (and therefore make it easier to extract results), this scale it typically chosen close to the energies and momenta actually exchanged in the interaction. However, in principle each physics quantity should be invariant under this choice. Changes in the renormalisation scale will only affect how much of the result comes from Feynman diagrams without loops (‘tree level diagrams’) and how much comes from the leftover finite parts of loop diagrams.

1.1.3 Introducing mass: the Higgs Mechanism

The coupling of the SM fermions to electroweak gauge bosons has required the introduction of the gauge group $SU(2)_L \times U(1)_Y$. To guarantee massive W and Z bosons and consequently weak interactions with a short range compared to the EM force, the electroweak symmetry has to be broken in a way that conserves the gauge invariance and renormaliseability of the theory. This is realised by means of the Higgs mechanism. It is based on the idea of spontaneous symmetry breaking, according to which the vacuum state of a system does not possess the same symmetry as the Lagrangian density [12]. This fundamental observation was first made in the relativistic context by R. Brout, F. Englert and P.W. Higgs [11, 12].

The simplest way to break the $SU(2)_L \times U(1)_Y$ gauge symmetry is to introduce an extra scalar field in the doublet representation of $SU(2)$ to the SM Lagrangian:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \tag{1.19}$$

with ϕ^+ and ϕ^0 complex fields. The Lagrangian density for such a field can be written with a specific potential in the form:

$$\mathcal{L}_\phi = \left(\mathcal{D}^\mu \phi\right)^\dagger \mathcal{D}_\mu \phi - V(\phi) = \left(\mathcal{D}^\mu \phi\right)^\dagger \mathcal{D}_\mu \phi - \mu^2 - \lambda \left(\phi^\dagger \phi\right)^2, \tag{1.20}$$

where $\mathcal{D}^\mu$ is the covariant derivative in the $SU(2)_L \times U(1)_Y$ gauge given in Eq. 1.13, μ^2 a mass parameter and $\lambda > 0$ the strength of the Higgs boson field’s self interaction. By requiring $\mu^2 < 0$, the symmetry is spontaneously broken because the minimum of the potential is no longer unique, but takes a value on a continuous ring in the complex plane as shown in Figure 1.1. In this ring the vacuum expectation value v is equal to:

---

[12] The Higgs mechanism is often compared to the global symmetry in a ferro-magnets. At rather high temperatures, all spins will be randomly directed as the configuration maximising the entropy will be most likely. This system hence contains a symmetry for the SO(3) rotation group. When the magnet however is cooled down, the spin interaction energy will be minimal if all spins are aligned. Hence, the system is ‘spontaneously’ broken to a system with a preferred spin direction, what spoils the SO(3) symmetry.

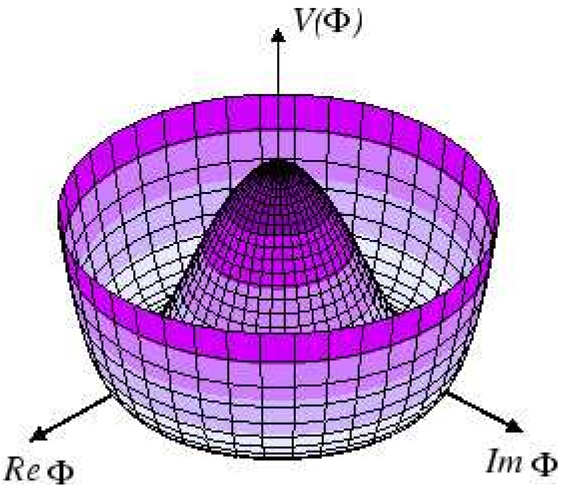


Figure 1.1: The 'sombrero' shaped Higgs potential.

$$< \phi^\dagger \phi >= v^2 = \frac{|\mu^2|}{\lambda} > 0 \tag{1.21}$$

An arbitrary choice of the vacuum ('unitary gauge') allows to write Eq.1.19 as

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} , \tag{1.22}$$

in which $h(x)$ represents the actual Higgs boson field, expressed as a quantum fluctuation about its vacuum expectation value. After this choice, only one out of four degrees of freedom from the original Higgs doublet remains. The other three are absorbed by the three vector bosons while acquiring mass. This can be seen when we actually calculate the Lagrangian term $(\mathcal{D}^\mu \phi)^\dagger (\mathcal{D}_\mu \phi)$ using Eq. 1.13 and Eq. 1.22. This gives:

$$\begin{aligned} (\mathcal{D}^\mu \phi)^\dagger (\mathcal{D}_\mu \phi) &= \frac{1}{2} \partial_\mu h \; \partial^\mu h \\ &+ \frac{1}{8} (v+h)^2 \; g^2 \; (W^1_\mu + iW^2_\mu)\,(W^{1\mu} - iW^{2\mu}) \\ &+ \frac{1}{8} (v+h)^2 \; (g' B_\mu - g W^3_\mu)\,(g' B^\mu - g W^{3\mu}) \end{aligned} \tag{1.23}$$

By writing the electroweak gauge boson fields as given in Eq. 1.15, the following masses for the physical gauge bosons can be derived from the mass terms in the above equation:

$$m_W = \frac{1}{2}\, g\, v \qquad\qquad m_Z = \frac{1}{2}\, v \sqrt{g^2 + g'^2}. \tag{1.24}$$

With Eq. 1.16 these relations can be merged to $m_W = cos\theta_W \, m_Z$[13].

This symmetry breaking mechanism predicts the existence of an additional neutral massive Higgs boson, with mass $M_H = \sqrt{2\lambda v^2}$.

---

[13] This relation however only holds at tree level.

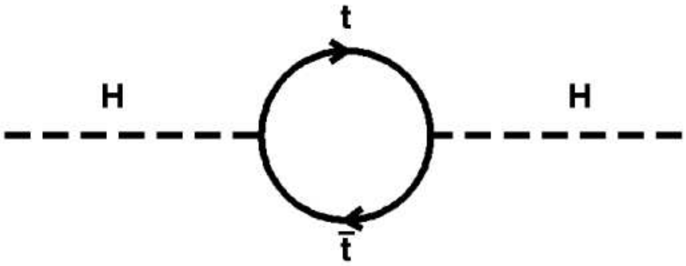


Figure 1.2: A fermionic top quark loop diagram responsible for extremely high radiative corrections to the Higgs boson mass.

In contrast to the mass of the bosons, the mass of the fermions is not generated by a gauge principle, but by the addition of extra gauge invariant Yukawa couplings in the Lagrangian density, resulting also in an increase of the number of free parameters in the SM. Nevertheless, a gauge invariant way of introducing fermion masses in the theory exists. Important for the motivation of top quark physics is that the coupling strength of the fermion to the Higgs boson field is proportional to the mass of the fermion. Hence, due to the extremely high top quark mass, the theoretical prediction of the most probable Higgs boson mass is extremely sensitive to the value of the top quark mass (cfr. Section 1.2.4).

1.1.4 Shortcomings of the Standard Model

Despite the enormous success of the Standard Model to describe the interactions of fermions in a gauge invariant quantum field theory, strong indications exist that this model is totally inadequate as a final theory.

A first important conceptual problem addresses the fact that the SM cannot naturally explain the relative low mass scale at which the electroweak symmetry is broken ($\sim 10^2\,\mathrm{GeV/c^2}$) compared to the Planck mass ($\sim 10^{18}\,\mathrm{GeV/c^2}$). Also, for the theory to be perturbative ($\lambda << 1$), the Higgs boson mass is supposed to be below 1 TeV. However, the calculation of this Higgs boson mass is extremely sensitive to higher-order quadratic and logarithmic divergences which occur as a result of radiative corrections. While the logarithmic divergences can be treated by using renormalisation, the quadratic divergences cannot. To illustrate, a fermionic top quark loop correction as shown in Figure 1.2 results in a Higgs boson mass correction term

$$\Delta m_H^2 = \frac{|\lambda_f|^2}{16\pi^2}\left[-2\Lambda_{UV}^2 + 6\,m_f^2\,\ln(\Lambda_{UV}/m_f) + \ldots\right]. \qquad (1.25)$$

The Ultra-Violet momentum cut-off Λ_{UV} corresponds to the mass scale at which manifestations of new physics are expected and the SM is to be replaced by a higher energy theory. If the SM is the only playing theory, this scale equals the Planck scale, and the correction term in Eq. 1.25 becomes typically 15 orders of magnitude higher than the electroweak scale. This is known as the naturalness or 'hierarchy' problem as the only way to manage it within the limits of the Standard model is to fine-tune all the constants in the theory to an extremely precise degree. This is a consequence of the fact that in the SM no symmetry protects the mass of the Higgs boson, which is rather embarrassing.

In the next paragraph a solution known as supersymmetry will be sketched, that also enjoys much interest from experimental point of view.

Besides the hard conceptual problems of the SM illustrated above and the fact that gravity is not incorporated in the theory, many other issues strengthen the believe that the SM does not offer the final answer up to the Planck scale. The main criticism is that the SM perfectly describes phenomena, but often is not capable to explain our observations. For example:

- Why are there three generations of fermions ?
- Why are there still 19 free parameters in the SM. Should it not be possible to find a theory predicting their values ?
- Why is the mass hierarchy observed in the fermion sector so extreme ?
- Why the disequilibrium of matter and anti-matter in the universe ?

1.1.5 Extensions of the Standard Model

Beyond the Standard Model, SUperSYmmetry (SUSY)[13] is considered to be a good candidate for resolving the problem of the potentially diverging mass of the Higgs boson, the so-called ‘hierarchy problem’ as introduced in the previous paragraph. The idea of supersymmetry is that, for each boson in Nature, there is a corresponding fermion, with the same number of degrees of freedom (and vice-versa). Also, supersymmetry relates their couplings. In the Minimal SuperSymmetric Model (MSSM), the quantum correction to the Higgs boson mass due to a fermionic loop is now compensated by a bosonic loop correction term [14]:

$$\Delta M_H^2 \propto (M_{SM}^2 - M_{SUSY}^2)\log \Lambda_{UV}^2 \tag{1.26}$$

The crucial point is that the extra loop correction due to the SUSY partner comes in with a minus sign, what cancels the quadratic divergences. Because no supersymmetric particle has been observed yet, SUSY should be broken and the residual logarithmic divergences remain. Moreover, Eq. 1.26 indicates that the difference between M_{SM} and M_{SUSY} should be in the order of the mass of the Higgs boson. This is a strong motivation for supersymmetry and also the reason why one expects to discover SUSY particles at TeV colliders. Moreover, requiring the conservation of R-parity[14] in these models delivers a perfect candidate, the so-called Lightest Supersymmetric Particle (LSP), for the dark matter problem in the universe.

---

[14] The R-parity quantum number of a particle with spin S, lepton number L and baryon number B is defined as

$$P_R = (-1)^{3(B-L)+2S}. \tag{1.27}$$

Conservation of R-parity prevents the decay of SUSY to SM particles, and consequently results in the prediction of a lightest supersymmetric particle in the model.

While SUSY is shown to be capable of stabilising the Higgs boson field, it leads not only to a whole spectrum of new particle types to be discovered (squarks, gluinos and neutralinos to name a few), but also to at least five different types of Higgs bosons in the MSSM. Furthermore it requires the measurement and tuning of 105 constants in addition to those already included in the Standard Model.

Alternative to SUSY, the introduction of extra space dimensions in the theory comprises a completely different strategy to solve the hierarchy problem [15, 16]. Instead of actually trying to cancel the quadratic divergences, the Planck scale is brought down to the TeV scale. As a consequence, at short distances (∼2 mm in the case of two extra space dimensions) the Newton's $1/r^2$ gravity law no longer holds. Since not much is known about gravity at short distances yet, these scenario's are not excluded by observations.

An overview of many other theories beyond the Standard Model can be found in [17–19].

1.2 Top Quark Physics

The discovery of the top quark [20] at Fermilab's $p\bar{p}$ Tevatron collider [21] in 1995 by the CDF [22] and DØ [23] collaborations opened up the new field of top quark physics. The relative late direct measurement is due to the extremely high top quark mass in comparison to the other fermion masses (∼35 times the mass of the next heaviest quark) and the need for a collider with sufficient centre-of-mass energy to produce top quarks. To date, only the 1.96 TeV Tevatron Collider is able to concentrate enough energy to generate these massive particles. As a consequence, all experimental top quark results referred to in this section are obtained by both the CDF and DØ collaborations.

In this section the rich field of top quark physics is introduced. Starting with a paragraph on the top quark production and decay aspects, paragraph 1.2.3 concentrates on the latest results on the measurement of the top quark mass. In a last paragraph the importance of top quark physics is motivated. Precision measurements of top quark properties will not only allow an even more stringent test of the Standard Model, the top quark sector it also known to be a highly sensitive window to new physics. Additionally, at the upcoming 14 TeV proton-proton collider (cfr. Chapter 3), the abundantly produced top quark pairs will serve as an important dataset to calibrate and commission the detector, illustrated by the study presented in Chapter 5. Finally, $\mathrm{t\bar{t}}$ production will represent an important channel background contributions in many other studies. Therefore, an appropriate study of for example its production cross-section is of particular interest here.

1.2.1 Top quark production

At hadron colliders two distinct Standard Model production mechanisms are expected to produce top quark event: the dominant $\mathrm{t\bar{t}}$ pair production via the strong interaction, and single-top production via the electroweak (EW) interaction. In the SM, $t\bar{t}$ pairs are produced either via quark-antiquark ($q\bar{q}$) annihilation or gluon fusion. Figure 1.3

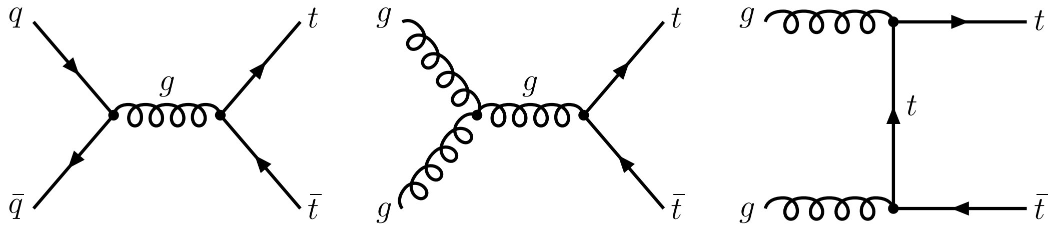


Figure 1.3: Leading order Feynman diagrams for $t\bar{t}$ production via the strong interaction.

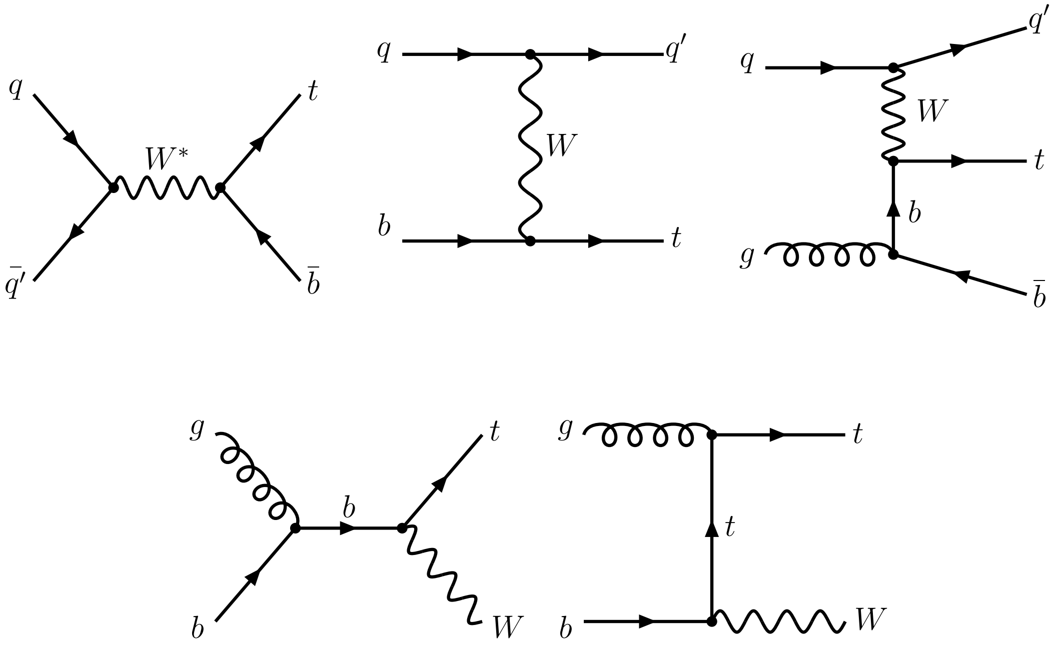


Figure 1.4: Leading order Feynman diagrams for single-top production via the weak interaction.

and Figure 1.4 show the leading order (LO) Feynman diagrams for respectively the $\mathrm{t\bar{t}}$-pair and single top production.

Because this thesis focuses on the study of the final state of $\mathrm{t\bar{t}}$-pairs, we will briefly describe the current theoretical and experimental knowledge about the top quark pair production cross-section.

In general, the total tree level $\mathrm{t\bar{t}}$ cross-section is calculated as a convolution of the parton distribution functions (PDF's) for the incoming hadrons and the cross-section of the partonic processes $q\bar{q}, gg \to t\bar{t}$:

$$\sigma(s, M_{top}^2) = \sum_{i,j} \int_0^1 dx_i \int_0^1 dx_j \, f_i(x_i, \mu_F^2) \, f_j(x_j, \mu_F^2) \, \hat{\sigma}_{ij}(\hat{s}, M_{top}, \alpha_s(\mu_R^2)), \qquad (1.28)$$

where i, j are the possible combinations of incoming gluon or quark/anti-quark pairs and $f(x, \mu_F^2)$ are the PDF's, evaluated at some factorisation scale μ_F, and a value

$(\mu_R = \mu_F = M_{top})$	LO (pb)	NLO (pb)	NLO+NLL (pb)
Tevatron ($\sqrt{\hat{s}}$=1.96 TeV)	4.37	$6.70^{+0.40}_{-0.74}$	$6.97^{+0.15}_{-0.47}$
LHC ($\sqrt{\hat{s}}$=14 TeV)	471	803^{+90}_{-89}	833^{+52}_{-39}

Table 1.3: The total $\mathrm{t\bar{t}}$ production cross-section expected at the Tevatron and LHC collider, calculated to LO [25], NLO and NLO+NLL [24] precision with M_{top}=175 GeV/c^2.

x which is the fraction of the incoming (anti-)proton energy that the parton carries. For more information on the definition, measurement and evaluation of these PDF functions, we refer to Section 2.2 of the next chapter. The partonic subprocess cross-sections $\hat{\sigma}_{ij}$, integrated over phase space, depend on the centre-of-mass energy in the collision, $\sqrt{\hat{s}}$, the value of the top quark mass M_{top}, and the QCD strong coupling constant α_s evaluated at a renormalisation scale μ_R. In principle, physical observables in a renormalisable field theory do not depend on any scale. But this is true only if all orders in perturbation theory are accounted for, which is impossible to calculate. When only fixed order calculations are available, the scale dependence might become significant. To minimise the effect, both the factorisation and renormalisation scale should be taken a value in the order of the scale of the process. For the calculation of the $\mathrm{t\bar{t}}$ production cross-section $\mu_R = \mu_F = M_{top}$ is a relevant choice. Varying the scale at a given order gives one an idea of the residual calculation's uncertainty.

The uncertainty in $\sigma^{LO}_{t\bar{t}}$ at hadron colliders is very large ($\sim 50\%$), mainly due to direct or indirect (e.g. α_s running) scale effects. Additional terms in the perturbative expansion for the cross-section are hence required for a meaningful comparison to the measurements. Today, the theoretical calculation with the highest accuracy is described in [24] and corresponds to a complete NLO+NLL resummation calculation. The SM production cross-sections expected both at the Tevatron and the upcoming Large Hadron Collider (cfr. Chapter 3) and their uncertainties are summarised in Table 1.3 for the LO, NLO and NLO+NLL calculation. The listed uncertainties for both NLO and NLO+NLL correspond to a variation of the scales between $M_{top}/2$ and $2\,M_{top}$. Where at the Tevatron a total $\mathrm{t\bar{t}}$ production cross-section is expected of ~7 pb, at the LHC this value increases by about two orders of magnitude (~833 pb).

The predicted SM value for the $\mathrm{t\bar{t}}$ production is confirmed by the Tevatron experiments. The CDF collaboration found a value $\sigma_{\mathrm{t\bar{t}}}$=7.3±0.5(stat.)±0.6(syst.)±0.4(lum.) pb for 760 pb^{-1} of accumulated data [26], where for DØ a slightly smaller value is found [27].

1.2.2 Decay of the top quark

Because the lifetime of a top quark ($\sim 10^{-25}$ s) is even an order of magnitude smaller than the characteristic hadronisation time of QCD, the top quark is the only quark that can be studied as a free quark. As a result, the decay of top quarks offers a

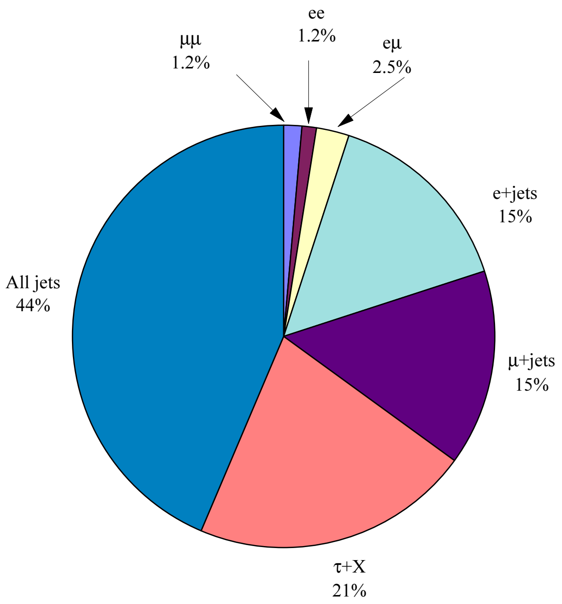


Figure 1.5: Branching fractions of $\mathrm{t\bar{t}}$ due to the various W-boson decays. All final states two b-quark jets arising directly from the $\mathrm{t\bar{t}}$-decay.

unique window on the properties of a bare quark free from the long-range effects of QCD, such as confinement. Actually measuring the top quark lifetime, or its inverse the decay width $\Gamma_{top} = \hbar/\tau_{top}$ ~1.4 GeV/c^2, is very challenging. Although the width exhibits itself in the Breit-Wigner invariant mass distribution of the decay products, this distribution is unfortunately much more narrow than the experimental resolution at a hadron collider. Nevertheless, in a CDF study the top quark decay width is measured to be less than 12.7 GeV/c^2 within a 95% confidence level using 1 fb^{-1} of data [28].

Within the SM, the top quark decays almost exclusively (99.8%) into a $W\,b$ pair[15]. A produced $\mathrm{t\bar{t}}$-pair will hence decay to two b-quark jets and two W-bosons, which will on their turn decay almost immediately either hadronically ($\mathcal{B}(W \rightarrow q\bar{q}) = 2/3$) or leptonically ($\mathcal{B}(W \rightarrow l\bar{\nu}_l) = 1/3$). As a consequence, three main $\mathrm{t\bar{t}}$ decay topologies can be distinguished: the fully-hadronic, the semi-leptonic and the fully-leptonic decay. Often, however, $\mathrm{t\bar{t}}$ decays involving the production of τ-leptons are regarded as an extra category. A graphical representation of the various SM branching fractions of top quark pairs is shown in Figure 1.5. All analyses presented in this thesis will start from 'semi-muonic' $\mathrm{t\bar{t}}$ decays, in which one W-boson decays to a muon, and the other one hadronically, which occurs in ~14.8% of the cases.

Other decays allowed in the SM are not only rare, but also mostly too difficult to identify. After $t \rightarrow bW$, the next most likely modes are the off-diagonal CKM decays $t \rightarrow Ws, Wd$ (cfr. Eq. 1.18). Flavor-changing neutral currents (FCNC) decays,

---

[15] In the DØ study presented in [29], this fraction is measured to be $\mathcal{B}(t \rightarrow Wb) = 1.03^{+0.19}_{-0.17}$, hence in good agreement with the Standard Model. This measurement can be translated in a lower bound on $|V_{tb}|$ of 0.78 at 95% C.L. (230 pb^{-1}).

$t \rightarrow X^0 q$, where $X^0 = g, \gamma, Z, H$ and $q = c, u$, are loop induced and highly suppressed by the GIM mechanism [30]. In the presence of new physics, many decay channels might compete with top quark SM decays.

In the SM, $t \rightarrow Wb$ is described purely by the universal $V - A$ charged-current interaction. Being on-shell, however, the W-boson's helicity in top decays is very different from that in the decays of any other quark, where the W-boson is highly virtual. The amplitude for positive W-boson helicity state is suppressed by a chiral factor $\frac{m_b^2}{M_W^2}$, so the W-boson helicity is a superposition of just the zero and negative helicity states. At tree level in the SM, the fraction $\mathcal{F}_0$ of the zero helicity W bosons in the top rest frame is [31]:

$$\mathcal{F}_0 = \frac{M_{top}^2/M_W^2}{1 + M_{top}^2/M_W^2} = 0.701 \pm 0.016 \tag{1.29}$$

for $M_{top} \gg M_W$. The large top mass exposes the longitudinal mode of the W, so a precise measurement of $\mathcal{F}_0$ serves as a stringent test of the SM. At CDF both the fraction of zero helicity and positive W-boson helicity states were measured using $318\,\mathrm{pb}^{-1}$ of accumulated data [32]. This resulting in the values $F_0 = 0.85^{+0.15}_{-0.22}(\text{stat.}) \pm 0.06(\text{syst.})$ and $F_+ = 0.05^{+0.11}_{-0.05}(\text{stat.}) \pm 0.03(\text{syst.})$, which are perfectly in agreement with the SM prediction.

1.2.3 Experimental knowledge about the top quark mass

While the top quark is the least well-studied quark in terms of quantum properties, its mass, M_{top}, is more accurately known than the mass of any other quark. This is also extremely important, as the top quark mass plays a crucial role in SM precision fits, as will be illustrated in the next paragraph.

The top quark mass can be extracted from each of the main $\mathrm{t\bar{t}}$ decay channels introduced in the previous section: the fully-leptonic, the semi-leptonic and the fully-hadronic channel[16]. Across these channels, three general approaches can be distinguished to measure M_{top} experimentally. In order of increasingly sophisticated usage of the information measured for each event, they are:

- **Employing the M_{top}-correlation of kinematic or decay observables**
  The value of M_{top} is reflected in several individual observables of the top quark event. Due to the immediate decay of the top quark for instance, the b-quark jet spectrum as well as the decay length distribution is highly correlated to the top quark mass. By fitting the spectrum of one or more of these observables one can discriminate between top quarks of different masses.

- **Kinematic reconstruction of the $\mathrm{t\bar{t}}$ decay**
  Because of the specific decay chain of the $\mathrm{t\bar{t}}$-pair, there are connections between the kinematics of the reconstructed four-momenta of the final state particles. For example, two light-quark jets reconstructed in the semi-leptonic or fully-hadronic

---

[16] In the Tevatron experiments, more common names for these decay channels are the dileptons, the ℓ+jets and the all-jets $\mathrm{t\bar{t}}$ decay channels.

$t\bar{t}$ final state are known to originate from the decay of the W-boson, and should consequently have an invariant mass equal to the W-boson mass. Optimal use of this information allows to build event solutions that provide a mass estimator for each candidate event. By fitting these mass estimators for a sample of events, the mass of the top quark can be extracted.

- **Matrix element fitting**
  This special case of kinematic reconstruction involves the full use of the information about the top quark production and decay. By using the leading-order matrix elements, in conjugation with a full knowledge of the experimental resolutions of the final state object momenta, a fit can be performed on the data. This is used to provide a probability that an observed event configuration is consistent with a top quark of a given mass.

In general, the more statistical information is used in a given measurement, the lower the statistical uncertainty on the resulting top quark mass estimate. However, the methods based on kinematic or decay observables often have the great advantage of being much less sensitive to systematic uncertainties, because for example no jet energy scale knowledge is required. Consequently, several of these techniques allow to extract M_{top} with comparable precision as the methods based on the full kinematic reconstruction, once enough data is available to reduce the statistical uncertainty. This scenario is expected for the top quark mass measurement based on the mean decay length of b-hadrons in $t\bar{t}$-events [33] and the determination of M_{top} via the $\ell\, J/\Psi$ invariant mass spectrum, where ℓ originates from the leptonic W-boson decay and $J/\Psi \rightarrow \ell\ell$ from the fragmentation of the b-quark [34]. The first method predicts a total M_{top} uncertainty of $5\,(2.5)\,\mathrm{GeV/c^2}$ for $8.5\,(10)\,\mathrm{fb^{-1}}$ of integrated luminosity at the Tevatron (LHC) collider [17], while 1-2 $\mathrm{GeV/c^2}$ is assumed to be achievable for the first $20\,\mathrm{fb^{-1}}$ at the LHC using the second technique. Another advantage of both methods is that one can consider them fully uncorrelated with the kinematic reconstruction methods.

Our current knowledge of the top quark mass is summarised in Figure 1.6, in which an overview is shown of the best top quark mass analysis results obtained from data collected at the Tevatron collider [35]. Both the yet published Run-I (1992-1996) measurements and the most recent preliminary Run-II (2001-present) measurements are combined using up to $1\,\mathrm{fb^{-1}}$ of data. Clearly, the best top quark mass estimates are inferred from the kinematic reconstruction and the matrix element fitting techniques applied on the semi-leptonic decaying $t\bar{t}$-events. In contrast to the fully-leptonic channel, this decay channel is over-constrained. Also, it contains a high momentum lepton which reduces impressively the channel background contributions compared to the rate expected for the fully-hadronic decay. Therefore it is often referred to as the 'golden channel'.

Taking into account the correlations between the uncertainties obtained for the individual measurements results in a preliminary world average mass of the top quark

---

[17] Today, the best top quark mass estimate achieved with this technique is $M_{top} = 183.9 \pm \begin{matrix}15.7\\13.9\end{matrix} \pm 5.6\,\mathrm{GeV/c^2}$ [35]. This measurement is also included in Figure 1.6 under the tag 'CDF II L_{xy}'

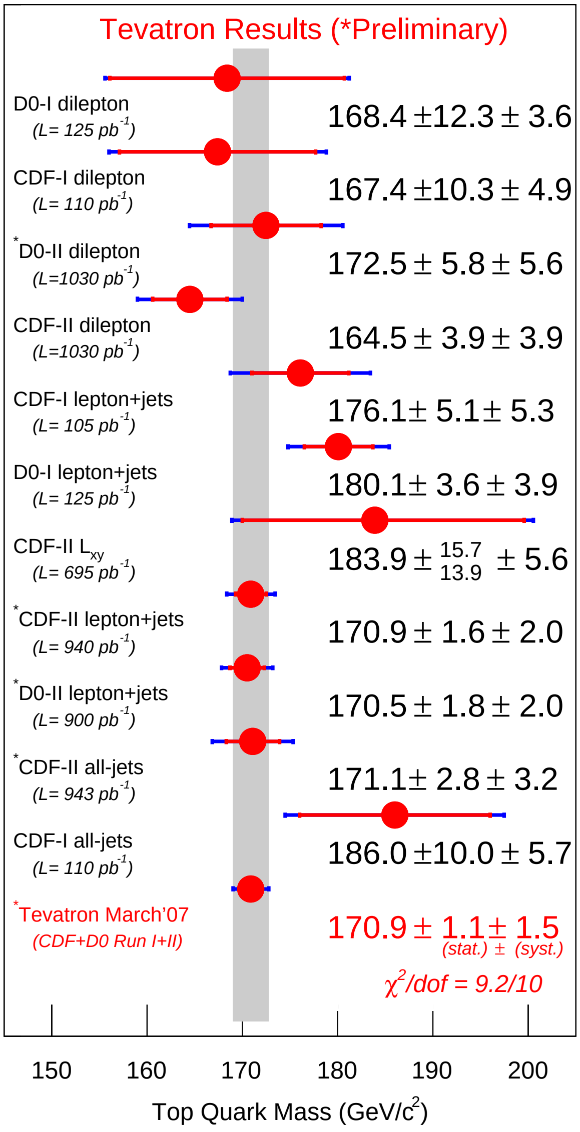


Figure 1.6: A combination of the Run-I and the most recent preliminary Run-II top quark mass measurements using up to 1 fb^{-1} of data collected at the Tevatron collider [35].

of $M_{top} = 170.9 \pm 1.1\,(\text{stat.}) \pm 1.5\,(\text{syst.})\,\text{GeV}/c^2$. Hereby, the systematic uncertainties were assumed Gaussian. Adding the statistical and systematic uncertainty quadratically yields a total uncertainty of 1.8 GeV/c^2, what corresponds to a relative precision of 1.1 % on the top quark mass. The main source of systematic uncertainty in these measurements are the uncertainties in the jet energy scale (JES). However, in Run-II both CDF and DØ take advantage of the increased $t\bar{t}$-statistics to employ new analysis techniques to reduce these uncertainties. In particular, the JES is now constrained using the in-situ calibration based on the invariant mass of the $W \to qq'$ decays in the semi-leptonic and fully-hadronic channels [36–38]. Residual η and E_T dependencies as

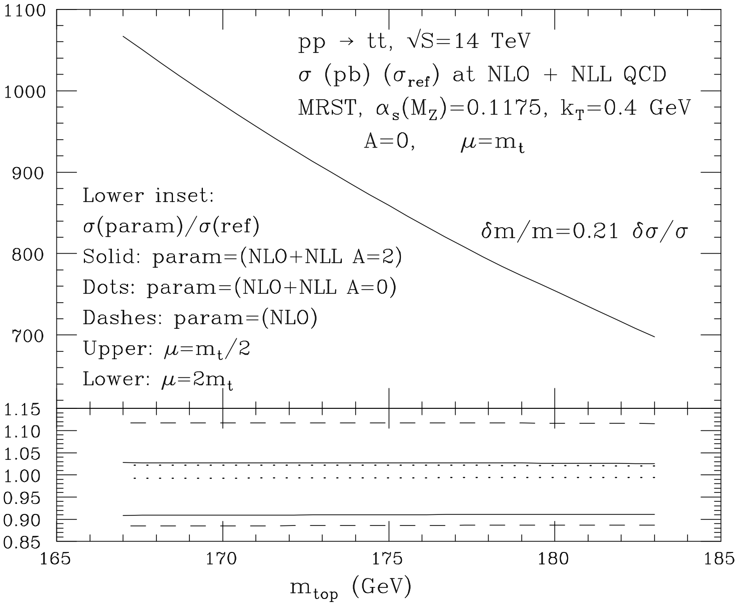


Figure 1.7: The $t\bar{t}$ production rates as a function of the top quark mass. The scale dependence of a NLO and a NLO+NLL calculation is tested in the lower inset [40].

well as uncertainties specific to the response of b-quark jets are treated as seperate error categories. On the contrary, the Run-II CDF dilepton measurement uses a JES determined from external calibration samples. Consequently, the weights and the mutual correlations between the listed measurements, which are employed in the calculation of the combined top quark mass, are strongly affected by e.g. the use of the same JES calibration correction factors [35, 39].

Once the full Run-II data set is on tape and analysed, the Tevatron aims for a final precising on the top quark mass of 1 GeV/c^2. One of the limiting factors to achieve this goal will be the modelling of the b-quark jet.

Finally, we would like to introduce the important correlation between the top quark mass and the value of the $t\bar{t}$ production cross-section [40]. This dependence of the cross-section on the mass is illustrated in Figure 1.7. A fit to the distribution shows that $\Delta\sigma/\sigma \sim 5\Delta M_{top}/M_{top}$. As a result, a 5% uncertainty on the measurement of the total cross section is equivalent to a determination of M_{top} to the level of 1%. However, this correlation assumes the correctness of the Standard Model. External production mechanisms of $t\bar{t}$ events might be detectable as a discrepancy between the indirect top quark mass measurement obtained from the cross-section estimate and the direct measurement.

1.2.4 Importance of top quark physics

Indirect measurement of the SM Higgs boson mass

The top quark mass and the Higgs boson mass both contribute to radiative terms in theoretical calculations of many observables that have been measured with good precision by LEP, SLD [41] and low-energy neutrino scattering experiments. As a consequence, a precisely known W-boson and top quark mass would allow to measure the relative impact of these radiative correction, what can be translated in a constraint on the SM Higgs boson mass. Using the latest values for the precision measurement of M_W and M_{top}, the result shown in Figure 1.8 is found [42]. The plot compares the today's accuracy of the top quark and W-boson masses with the values determined indirectly via loop corrections using the LEP-1 and SLD data. Clearly, the direct measurement of M_{top} did reduce the parameter-space impressively, resulting in a shift of the most probable SM Higgs boson mass to lower values. As shown in Figure 1.9, the current χ^2 fit to the EW data sets an upper limit for the mass of the Higgs boson mass of 144 GeV/c^2 (one-sided 95 percent confidence level including both the experimental and the theoretical uncertainty) [42]. This limit increases to 182 GeV/c^2 when including the LEP-2 direct search limit of 114.4 GeV/c^2. However, due to the logarithmic dependence of the M_H to both M_W and M_{top}, the accuracy of the indirect SM Higgs boson mass measurement is strongly determined by the precision achieved on the direct top quark and W-boson mass estimation. Therefore, a reduction of the total top quark mass uncertainty from 2 to 1 GeV/c^2 as aimed for at the Large Hadron Collider would induce a major improvement in the estimation of M_H. Finally, we would like to stress that today the impact of the uncertainty on M_W is larger than the one arising from M_{top}, hence also an improvement of the W-boson mass is crucial.

Yukawa coupling equals one

The high sensitivity of the Higgs boson mass to the top quark mass is mainly because the latter mass is found to be very close to the energy scale of Electroweak Symmetry Breaking (EWSB), resulting in a Yukawa coupling curiously close to 1:

$$Y_{top} = \sqrt{2}\, M_{top}/v \sim 1 \tag{1.30}$$

with the Higgs boson vacuum expectation value $v = 246$ GeV/c^2 [18] and $M_{top} =$ 170.9 GeV/c^2. From theory point of view this is considered often as no coincidence, what led to speculation that important new physics may be accessed via top quark studies, or that at least the top quark plays an important role in this EWSB. If there is any new physics associated with the generation of mass, it may be more apparent in the top quark sector than within any other sector of the SM. However, deriving direct information on top quark properties such as the Yukawa coupling itself turns out to be very difficult.

Sensitivity to new physics

The large mass of the top quark is particularly important in various extensions of the SM, as many of the predicted 'beyond the SM' particles have mass-dependent couplings. Top quark production at hadron colliders, be it $t\bar{t}$ or single top, is for example an ideal place to look for new physics, as its manifestations might be observed in the top quark production rate or in the obtained kinematic observable's distributions.

---

[18] The study of the decay width of muons, $\Gamma(\mu \rightarrow \nu_\mu e \bar{\nu_e})$, resulted in $v = (\sqrt{2}G_F)^{-2} = 246\,\mathrm{GeV/c^2}$ (with G_F the Fermi coupling constant).

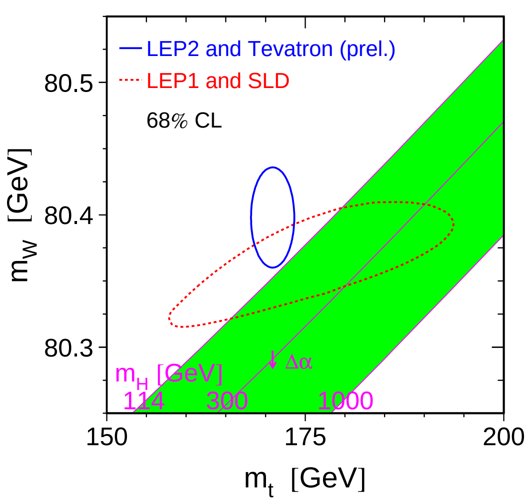


Figure 1.8: A comparison of the indirect W-boson and top quark mass measurements based on the LEP-1 and LSD data with the result obtained from direct measurements at LEP-2 and Tevatron. The logarithmic dependence of the most probable SM Higgs boson mass on the value of M_W and M_{top} is illustrated.

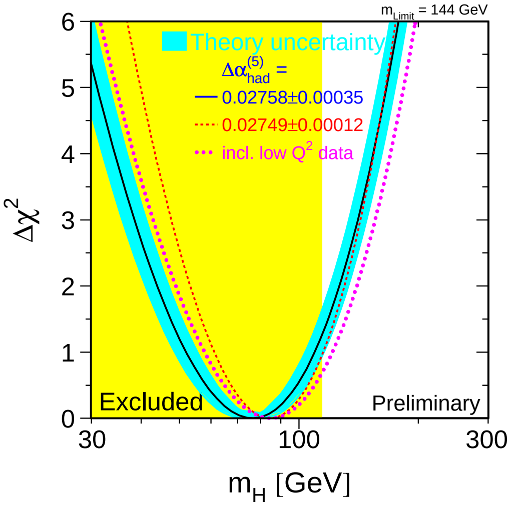


Figure 1.9: The predicted Higgs boson mass obtained from the χ^2 fit to the EW precision data (Winter 2007).

In $t\bar{t}$ production, it is especially interesting to study the invariant mass distribution of the top pairs, $d\sigma/dm_{t\bar{t}}$, since it can reveal resonant production mechanisms [43]. Many theories postulate heavy resonances decaying to $t\bar{t}$, such as technimesons in technicolor models [44, 45], other models of strong EWSB [46, 47], as well as in extra-dimension scenarios in which new scalar bosons might have couplings preferential to the third generation [48] ($Z' \to \mathrm{t\bar{t}}$). The current limit set by the DØ collaboration on the minimal mass of a Z' resonance decaying to $\mathrm{t\bar{t}}$ is 680 GeV/c^2 at 95% C.L. [49].

Top quarks as Channel Background

Another argument to study in detail the production of top quark is because they will constitute an important background contribution to many searches for new physics. Especially at TeV scale colliders, the abundantly produced $\mathrm{t\bar{t}}$ pairs are expected to make several analyses a lot more challenging, such as:

- the search of the Higgs boson in the $gg \to \mathrm{t\bar{t}}H$ channel [50] ($\sigma[gg \to \mathrm{t\bar{t}}H] = 0.664\,pb$ for $M_H = 120\,\mathrm{GeV/c^2}$)
- the study of the discovery potential for MSSM charged Higgs bosons in $gg \to tbH^{\pm}$ [51] ($\sigma[gg \to \mathrm{t\bar{t}}H^{\pm}] = 0.570\,pb$ for $M_{H^{\pm}} = 311\,\mathrm{GeV/c^2}$)
- the E_T^{miss}-measurement in the search for a LSP [52].

Detector Calibration and Commissioning

At the LHC typically about one top quark pair will be produced per second. For the first time, this huge statistics will not only allow detailed measurements of top quark parameters, also the use of top quark events for the calibration and commissioning of the detector will be feasible. This will be illustrated in Chapter 5, where semi-muonic decaying $t\bar{t}$-events are employed to calibrate the light quark jet energy scale. This is one of the first studies that opens a whole new application of top quark physics. Another example of the possible calibration techniques with top quarks, is the measurement of the b-tagging performance [53]. A perfect understanding of both the jet energy scale and the b-tagging performance are crucial for the performance and the potential of many physics analyses to be performed at the LHC.

Chapter 2

Monte-Carlo Event Generation and Simulation

To help in the design of physics experiments in general, and to define experimental strategies and potentials, theoretical predictions of the expected physics are needed. These predictions should reproduce as closely as possible the collision processes taking place in a real detector. A largely successful way of achieving this goal is through the so-called event generator and cross-section integrator codes.

Following the excellent performance and large statistics accumulated by the Tevatron collider, and the upcoming startup of the LHC physics era, the recent years have witnessed an impressive progress in the development of improved tools for the simulation of the complex final states produced in hadron-hadron collisions [54]. The most significant advances have been the inclusion of next-to-leading-order (NLO) matrix elements in the shower Monte-Carlo (MC) codes, and the consistent merging of shower MCs with the leading-order (LO) calculations for final states with many hard partons.

The aim of this chapter is to show the importance of both improvements for top quark physics at the LHC, and to describe the different compounds that can be distinguished in the production chain of Monte-Carlo simulated events.

2.1 The event simulation chain

An advantage in the Monte-Carlo generation of a detectable particle flow arising from a beam collision, is the possibility to factorise its simulation into several compounds, as is illustrated in Figure 2.1.

This chain can be summarised as follows [55]:

- Two incoming hadrons can be viewed as a bag of partons and are heading to a collision. The probability density information for each parton in the hadron is available by means of measured 'Parton Distribution Functions' (cfr. Section 2.2).

- A collision occurs between two partons and gives the hard process of interest (cfr. Section 2.3). However, in most bunch crossings rather soft, simple elastic or diffractive events are produced. These 'minimum bias' events will be piled up to the hard event (cfr. Section 3.1.2).

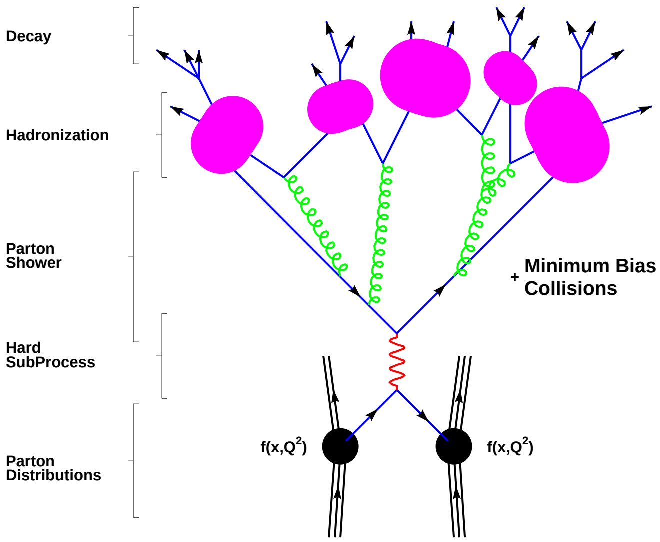


Figure 2.1: Schematic representation of the hadronic interaction generation chain. For simplicity, the parton showering is not shown for the initial state partons.

- Both on the incoming as outgoing partons, gluon radiation will occur due to the acceleration of the colour charges in the collision. A popular strategy to model these emissions is by means of a 'parton shower' (cfr. Section 2.4).

- As time goes by, and partons recede from each other, perturbative QCD becomes no longer valid as its coupling constant approximates unity. Hadronisation models need to be introduced to describe the formation of the ultimate, detectable and stable colour singlet states (Section 2.7).

- Meanwhile, the colour charged hadron remnants from the hard collision will radiate and hadronise while travelling essentially in the same direction. The modelling of this extra source of energy flux in the detector, known as the 'underlying event', is described in Section 2.8.

Many event generators are available, from general-purpose ones as `PYTHIA` [56] and `HERWIG` [57] to more specialised ones. Among this last class of generators, many deal with the matrix elements for some specific set of processes, others concentrate on topics as parton showers or particle decays. However, these specialised programs have in most cases no way to handle hadronisation. Consequently, many are used as 'plug-ins' to the general purpose ones. In general, more accurate theoretical predictions are obtained by the use of a specialised hard event generator combined with the parton showering and hadronisation implemented in general purpose programs.

2.2 Parton Distribution functions

The calculation of any production cross-section relies upon a knowledge of the distribution of the momentum fraction x and the virtuality scale Q^2 of the partons (quarks and gluons) in the incoming hadrons. These parton densities or parton distribution functions (PDF's) can not be calculated perturbatively from first principles but rather are determined by global fits to data from deep inelastic scattering (DIS), Drell-Yan (DY), and jet production at current energy ranges. Two major groups, CTEQ [58] and MRST [59], provide semi-regular updates to the parton distributions when new data and/or theoretical developments become available. The main source of information to the parton distribution $f_{q/g}(x, Q^2)$ is provided by the measurement of the DIS structure functions in lepton-hadron scattering and the lepton pair production in hadron-hadron collisions. In addition, jet final states detected at the Tevatron experiments contribute significantly to the leading order gluon distribution functions $f_g(x, Q^2)$.

Currently, global parton distribution fits are carried out to next-to-leading order with a remarkable consistency between the data and theory.

Even though the data used in the PDF fits cover a wide range in x and Q^2, an extrapolation has to be made to cover the range accessible at the LHC. Nowadays, programs in use by CTEQ and MRST should carry out this evolution to an accuracy of a few percent over the hadron collider's kinematic range, except maybe at very large and very small x-regions.

For the production of a $\mathrm{t\bar{t}}$-pair at the LHC, the relevant proton momentum fraction x and virtuality scale Q^2, are given by:

$$Q = M_{\mathrm{t\bar{t}}}; \qquad\qquad x_{1,2} = \frac{M_{\mathrm{t\bar{t}}}}{\sqrt{s}} \exp^{\pm y}, \tag{2.1}$$

with:

$\sqrt{s} = 14\,\mathrm{TeV}$: the centre-of-mass energy of the colliding protons
y: the rapidity of the $\mathrm{t\bar{t}}$-system

Hence, in order to produce an unboosted $\mathrm{t\bar{t}}$-system ($y = 0$) with an invariant mass $M_{\mathrm{t\bar{t}}} = 350\,\mathrm{GeV/c^2}$, both momentum fractions x_1 and x_2 should be equal to $x_{\mathrm{t\bar{t}}} = 0.025$. The parton distributions from the CTEQ5l PDF release[1] are plotted in Figure 2.2, where the scale was set at $Q^2 = (350\,\mathrm{GeV/c^2})^2$ [60]. From this plot, we can calculate the fraction of $\mathrm{t\bar{t}}$-pairs produced as a result of gluon-fusion:

$$\begin{aligned} \frac{\sigma(gg \to \mathrm{t\bar{t}})}{\sigma(pp \to \mathrm{t\bar{t}})} &= \frac{xF^g(x_{\mathrm{t\bar{t}}}, Q^2_{\mathrm{t\bar{t}}})^2}{xF^g(x_{\mathrm{t\bar{t}}}, Q^2_{\mathrm{t\bar{t}}})^2 + 2 \times \sum_{q_i\bar{q}_i} xF^{q_i}(x_{\mathrm{t\bar{t}}}, Q^2_{\mathrm{t\bar{t}}}) \times xF^{\bar{q}_i}(x_{\mathrm{t\bar{t}}}, Q^2_{\mathrm{t\bar{t}}})} \\ &\approx 95\% \quad \text{for } M_{\mathrm{t\bar{t}}} = 350\,\mathrm{GeV/c^2} \\ &\approx 83\% \quad \text{for } M_{\mathrm{t\bar{t}}} = 1\,\mathrm{TeV/c^2} \end{aligned} \tag{2.2}$$

Hence, the vast majority of top quark pairs expected at the LHC will be produced from a gluon-fusion, in particular $\mathrm{t\bar{t}}$-pairs with a relatively low invariant mass. The opposite is true for top quark pairs produced at the Tevatron collider. The much lower centre-of-mass energy ($\sqrt{s}$=1.96 TeV) requires initial partons with a significantly higher x-value of the order of 0.18 for $y_{\mathrm{t\bar{t}}} = 0$. As a result, $\mathrm{t\bar{t}}$-pairs will be mainly produced due

---

[1] This release version is used in the production of all Monte-Carlo simulated event samples analysed in this work.

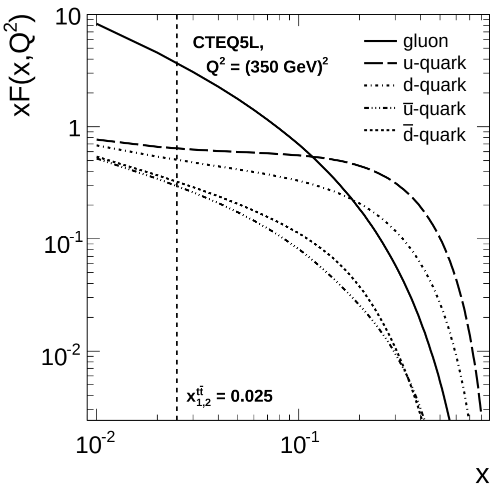


Figure 2.2: Several PDF's from the CTEQ5L set are plotted at a Q^2 value of $(350GeV)^2$, corresponding to the threshold of the tt̄-pair invariant mass spectrum.

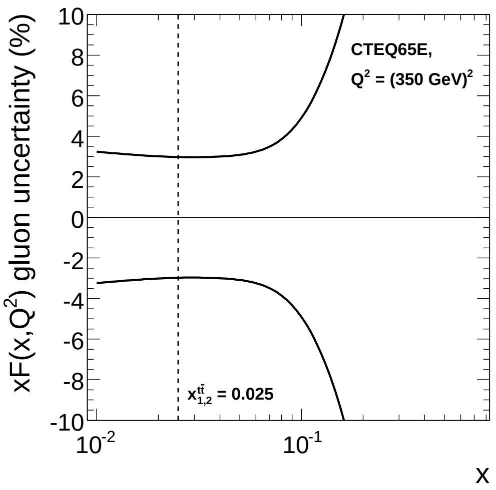


Figure 2.3: The uncertainty on the gluon PDF as a function of the momentum fraction x. At the typical tt̄-production x-values, an uncertainty of 3% is expected.

to quark/anti-quark annihilation. Consequently, at the LHC the uncertainty on the gluon PDF will be the main contribution to the overall PDF systematic uncertainty on a top quark measurement, and is expected to be of the level of 3% (Figure 2.3). This PDF uncertainty was calculated based on the Hessian technique [61, 62], wherein a large matrix with dimensions equal to the number of free parameters in the fit, is diagonalised. In the case of CTEQ, this results in 20 orthogonal eigenvector directions. To actually evaluate the PDF uncertainty, 40 PDF sets are build, where in each of the sets one of the eigenvectors is shifted in positive or negative direction. The final PDF systematic error on on a measured quantity A (e.g. M_{top}) can then be calculated by summing over the 20 couples of PDF sets in the following way:

$$\delta A = \pm 1/2 \sqrt{\sum_{eigenvectors\ i} \left[A(+,i) - A(-,i)\right]^2}, \tag{2.3}$$

where $A(+,i)$ represents the value of A given a PDF with an excursion of eigenvector i in positive direction. This procedure will be applied in this work to estimate the systematic effects on the various defined estimators due to the PDF uncertainties.

2.3 The simulation of the hard event

The QCD factorisation theorem for short-distance inclusive processes allows to write the cross-section σ for the production of a certain final state X as:

$$\sigma_X = \sum_{a,b} \iint dx_1 dx_2 f_a(x_1, Q^2) f_b(x_2, Q^2) \times \hat{\sigma}_{ab \to X} \tag{2.4}$$

In the calculation of this inclusive cross-section, all contributions (spin, colour, ...) from the different initial partons are summed, while integrating over the momentum fraction of each of the partons a and b. The factors $f_{a/b}(x_{1/2}, Q^2)$ represent the evaluation of the parton density functions for a parton with momentum fraction x and scale Q^2. Each of the subprocesses has a differential cross-section $\hat{\sigma}_{ab\rightarrow X}$.

In order to address the physics of interest, already within the Standard Model a large number of processes have to be available in the generators. Although many can be found in general-purpose programs as `PYTHIA` and `HERWIG`, often only lowest order processes are available, while experimental interest may be in higher orders, or in a precise description of extra hard jets in the final state, either as signal or as a potential background.

A new strategy to obtain the most accurate predictions in a detector simulation friendly way is to match fixed-order calculations and parton showers. This fixed-order calculations might be a fully exclusive description of events correct at NLO in normalisation and distributions. However, as these calculations are difficult and therefore often not available, a popular alternative is a LO matrix element calculation of a many parton final state.

These parton-level generators are able to compute tree-level matrix elements with a fixed number of legs (i.e. a fixed number of partons in the final state). The codes are based either on the direct computation of the relevant Feynman diagrams, or on the solutions of the underlying classical field theory. The goal is to maintain a consistent leading logarithm (LL) accuracy in the prediction of a final state accompanied by a varying number of extra jets. This is realised by the convolution of a LO matrix element for n hard partons with a full shower evolution obtained with a shower MC. However, this merging step contains two difficulties:

1. Common parton level generators sum and average over flavours and colours, and do as a consequence not provide sufficient information on the flavours and colour of the events necessary for the parton shower. In order to reliably evolve a multiparton state into a multijet configuration, it is however necessary to associate a specific colour-flow pattern to each generated parton-level event. A colour-flow or colour-configuration describes the set of colour connections among the partons and defines the set of dipoles for a given event (Fig. 2.4). These dipoles are the starting point for the emission of gluon radiation, that is implemented in the parton showering code.

2. Showering often leads to $m > n$ hard partons in the final state, resulting in a non-zero probability that the same n-jet configuration is generated starting from a $(n - m)$-parton configuration. Solutions to this risk of double-counting are known and will be described in Section 2.5.

Despite these difficulties, the combination of tree-level matrix element generators with parton showering programs is essential for analyses based on multi-jet configurations (such as a top quark mass precision measurement), where the standard showering codes are basically unable to describe the kinematics of those processes correctly.

In the CMS experiment, two LO matrix element generators are widely used in the study of top quark physics: `AlpGen` and `MadGraph`/`MadEvent`.

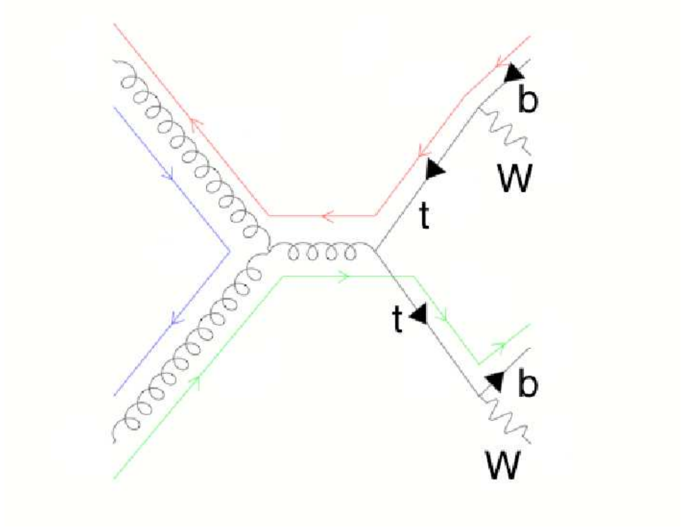


Figure 2.4: An example of a colour flow in a t$\bar{\text{t}}$-production process.

The `AlpGen` generator

`AlpGen` [63] is designed for the generation of a pre-defined list of Standard Model processes in hadronic collisions, with emphasis on final states with large jet multiplicities, where both the bottom and top quark mass are included. We are in particular interested in the following available processes:

- t$\bar{\text{t}}$ $+\ N\,jets$ (with $N \leq 4$)
- $(W \rightarrow l\,\nu_l)\ +\ N\,jets$ (with $N \leq 6$)
- $(W \rightarrow l\,\nu_l)\,b\,\bar{b}\ +\ N\,jets$ (with $N \leq 4$)

The program has two main modes of operation. In the first mode the code performs the parton-level calculation of the matrix elements relative to the selected hard processes, and generates weighted events. These events give easy access to the cross-sections in presence of some overall generation cuts. In the second mode of operation the code generates unweighted parton level events for subsequent evolution via a parton showering routine provided by general purpose packages as `HERWIG` and `PYTHIA`. In this run the matrix element calculation generates all the flavour and colour information necessary for the complete shower evolution.

The calculation of the cross-section for a given hard process is performed via several integration cycles. In a first cycle, the goal is to explore how the cross-section for the selected process is distributed in phase-space and among the possible contributing subprocesses. Event by event, the subprocess, phase-space, and spin and colour configuration of the partons is chosen randomly[2]. A weight is then obtained from the

---

[2] The sum over several subprocesses is simplified by the possibility to factorise out trivial terms as parton densities and CKM factors from the single, flavour independent, matrix element.
E.g. the overall contribution for the process $q\bar{q}' \rightarrow WQ\bar{Q}gg$ is given by:

$$[\ u_1\bar{d}_2\cos^2\theta_c\ +\ u_1\bar{s}_2\sin^2\theta_c\ +\ c_1\bar{s}_2\cos^2\theta_c\ +\ c_1\bar{d}_2\sin^2\theta_c\ ]\ \times\ |M(q\bar{q}' \rightarrow WQ\bar{Q}gg))|^2$$

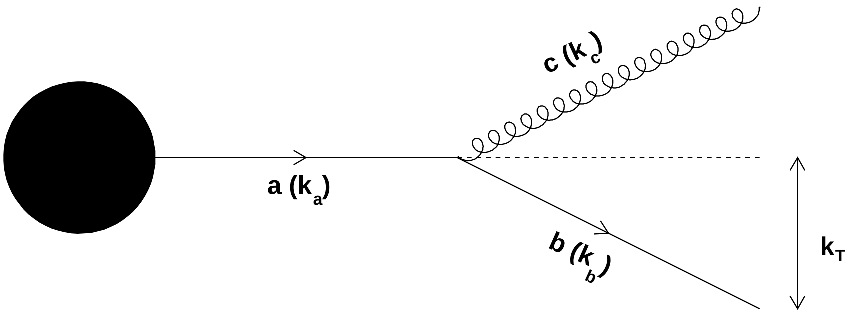


Figure 2.5: Gluon emission from a final state quark.

calculated matrix element, combined with the phase space and the initial state parton luminosity factors. At the end of the first integration iteration, a map of the cross-section distribution among the different subprocesses and phase-space is available. This grid will then be used in a second cycle to produce the weighted parton level events.

In the consequent unweighting step, the code will sequentially read in the weight and initial random seed of the previous events, and use this knowledge to reconstruct all information about the event (kinematics, flavours, spins and colours of the initial partons). The colour configuration for the event is selected according to the algorithm described in [64, 65].

The `MadGraph/MadEvent` generator

`MadEvent` [66] is a multi-purpose, tree-level event generator which is powered by the matrix element generator `MadGraph` [67]. Given a user process, `MadGraph` automatically enumerates all Feynman diagrams contributing to that process, generates their amplitudes and produces the mappings for the integration over the phase space. This process-dependent information is packaged into `MadEvent`, and a stand-alone code is produced that allows the user to calculate cross-sections, and to obtain unweighted events that can be passed directly to a shower MC program. Interfaces are available for both `HERWIG` and `PYTHIA`. Limitations of the code are related to the maximum number of final state QCD particles. Currently, the package is limited to ten thousand diagrams per subprocess. The simulation of $W + 5\,jets$-events is for example close to its practical limit.

2.4 Parton Showering

In every process that contains accelerated (colour) charged objects in the initial or final state, gluon of photon radiation could occur, and change the overall event topology. To illustrate the importance of a dedicated software tool to describe these emissions, we consider the branching of a final state quark in a quark-gluon pair, $q \rightarrow q\, g$ (Figure 2.5).

With the momenta of b and c parametrised as:

$$k_b \;=\; z\, k_a + k_T \qquad \& \qquad k_c \;=\; (1-z)\, k_a - k_T, \tag{2.5}$$

the $a \rightarrow b\,c$ branching cross-section will contain a singular term, given by:

$$d\sigma^{(a\rightarrow b\,c)}|_{singular} = \frac{\alpha_s}{2\pi} \int dk_T^2 dz\; C_F\, \frac{1+z^2}{1-z}\, \frac{1}{k_T^2}\, d\sigma^{(a)}, \tag{2.6}$$

with $C_F = (N_c^2-1)/(2N_c) = 4/3$ the colour factor, and $\sigma^{(a)}$ the cross-section as if a wouldn't have split into b and c. We notice that the $q \rightarrow q\; g$ branching cross-section diverges when the gluon either becomes collinear with the quark ($k_T \rightarrow 0$) or when the gluon energy vanishes ($z \rightarrow 1$). In QCD, compared to QED, this divergence is further enhanced by a larger coupling constant α_s, that actually increases when the emission becomes softer. Furthermore, the non-Abelian character of QCD leads to $g \rightarrow gg$ branchings with similar divergences, without any correspondence in QED. The third main branching, $g \rightarrow qq$ does not have the soft divergence and is less important.

Now, as the rate for one emission of a gluon is big, also the rate for two or more is big, resulting in the need for high orders and many loops in matrix-element-based calculations. As an alternative, parton-showering provides a sensible, but approximate, tool to simulate the structure of emissions in soft and collinear regions, based on the following two concepts:

1. an iterative structure that combines simple expressions for $q \rightarrow qg$, $g \rightarrow gg$ and $g \rightarrow qq$ branchings to build up the complex multiparton final state
2. a Sudakov factor that offers a physical way to handle the cancellation between real and virtual divergences and that ensures the conservation of total probability.

Both concepts will be further elaborated in the next paragraphs.

2.4.1 The shower approach

A shower may be viewed as a sequence of $1 \rightarrow 2$ branchings $a \rightarrow b\,c$, where a is called the mother parton and b and c the two daughters. Each daughter is free to branch in its turn, so that a tree-like structure evolves.

In general, a distinction is made between radiative processes on initial hard event particles, known as 'initial state radiation (ISR)', and emissions of outgoing particles, denoted as 'final state radiation (FSR)'. To see where this separation comes from, we consider another time the branching of a gluon from a quark, as was illustrated in Figure 2.5. We define a as moving along the z-axis, and introduce the lightcone momenta

$$p_\pm \;=\; E \,\pm\, p_z, \tag{2.7}$$

so that the relation

$$p^2 \;=\; E^2 \,-\, p_x^2 \,-\, p_y^2 \,-\, p_z^2 \;=\; m^2, \tag{2.8}$$

translates to

$$p_+p_- \;=\; m^2 \,+\, p_x^2 \,+\, p_y^2 \;=\; m^2 \,+\, p_T^2. \tag{2.9}$$

After defining the splitting of p_+ by $p_+^b = z\,p_+^a$ and $p_+^c = (1-z)\,p_+^a$, the demand for the conservation of p_- gives:

$$p_-^a = p_-^b + p_-^c \iff m_a^2 = \frac{m_b^2}{z} + \frac{m_c^2}{1-z} + \frac{p_T^2}{z\,(1-z)}, \tag{2.10}$$

where we used $p_T^b = p_T^c = p_T$.

In ISR the incoming a should be massless, and if c does not interact any further as well this particle should have mass zero. In this case, $m_b^2 = -(1-z)p_T^2 < 0$, i.e. the b-quark is off-shell. This space-like virtuality is, according to Heisenberg's uncertainty principle, only possible for very small time intervals just before the hard interaction. In the case of FSR, and in the assumption that b and c don't radiate anymore because of being produced massless, we have that $m_a^2 = p_T^2/(z\,(1-z)) > 0$. Hence, after the branching, the original a-quark looses its time-like virtuality. The virtuality scale Q^2 for each parton p taking part of a branching is defined as $|m_p^2|$, and gives an approximate sense of time-ordering. In a initial-state shower, Q^2-values are gradually increasing, while Q^2 is decreasing in a final state shower. In the next sections we will introduce the theoretical and statistical tools that are necessary in the description of both types of showering.

2.4.2 The DGLAP equations

The evolution of the parton densities is given by the DGLAP, Altarelli-Parisi or evolution equations [68]. In terms of the already defined z-variable and the scale dependent variable t,

$$t = ln(Q^2/\Lambda_{QCD}^2) \implies dt = \frac{dQ^2}{Q^2} \quad \text{(with Λ_{QCD} the QCD scale of α_s)}, \tag{2.11}$$

the differential splitting probability $d\mathcal{P}$ for a parton a to branch is

$$d\mathcal{P} = \sum_{b,c} \frac{\alpha_{abc}}{2\pi} P_{a\to b\,c}(z)\,dt\,dz\ . \tag{2.12}$$

All allowed branchings are summed, and α_{abc} is taken α_{em} or α_s depending on the branching. $P_{a\to b\,c}$ are the splitting functions, given by:

$$\begin{aligned}
P_{q\to qg}(z) &= C_F\,\frac{1+z^2}{1-z}\ ,\\
P_{g\to gg}(z) &= N_C\,\frac{(1-z(1-z))^2}{z(1-z)}\ ,\\
P_{g\to q\bar{q}}(z) &= T_R\,(z^2+(1-z)^2)\ ,\\
P_{q\to q\gamma}(z) &= e_q^2\,\frac{1+z^2}{1-z}\ ,\\
P_{\ell\to \ell\gamma}(z) &= e_\ell^2\,\frac{1+z^2}{1-z}\ ,
\end{aligned} \tag{2.13}$$

with $C_F = 4/3$, $N_C = 3$, $T_R = n_f/2$ and e_q^2 and e_ℓ^2 the squared electric charges. These equations can be combined to allow for the successive emission in several steps, where e.g. a $q \to qg$ branching is followed by further branchings of the daughters.

2.4.3 Definition of the Sudakov form factors

As already mentioned, Q^2 or t can be interpreted as a kind of time for the shower evolution. For initial state showers, t will increase during successive branchings, while in final state showers just the opposite is true. Starting from equation 2.12, one can, for a given value of t and a specific branching, define the probability for that branching as an integral over the allowed z-values:

$$\mathcal{I}_{a\to b\,c}(t) = \int_{z_-(t)}^{z_+(t)} dz\, \frac{\alpha_{abc}}{2\pi}\, P_{a\to b\,c}(z)\ . \tag{2.14}$$

The probability that a branching occurs during a small range of t values, δt, is given by

$$\mathcal{P}_{branching}(t, t+\delta t) \ = \ \sum_{b,c} \mathcal{I}_{a\to b\,c}(t)\, \delta t, \tag{2.15}$$

and thus the probability for no emission by $\mathcal{P}_{no-branching} \ = \ 1 - \mathcal{P}_{branching}$.

If the evolution of parton a starts at a ‘time’ t_0, the probability that the parton has not yet branched at a ‘later time’ $t > t_0$ is given by the product of the probabilities that it did not branch in any of the small intervals δt between t_0 and t. In other words, letting $\delta t \to 0$, the no-branching probability exponentiates:

$$\mathcal{P}_{no-branching}(t_0, t) = \exp\left\{ - \int_{t_0}^{t} dt' \sum_{b,c} \mathcal{I}_{a\to b\,c}(t') \right\} = S_a(t)\ . \tag{2.16}$$

Thus the actual probability that a branching of a occurs at t is given by

$$\frac{d\mathcal{P}_a}{dt} = -\frac{d\mathcal{P}_{no-branching}(t_0,t)}{dt} = \left(\sum_{b,c} \mathcal{I}_{a\to b\,c}(t) \right) \exp\left\{ - \int_{t_0}^{t} dt' \sum_{b,c} \mathcal{I}_{a\to b\,c}(t') \right\}. \tag{2.17}$$

The first factor denotes the total branching probability for a given value of t, while the exponential suppression factor ensures the conservation of total probability: if a parton has already branched at a $t' < t$, it can no longer branch at t [3]. This latter factor is often referred to as the ‘Sudakov form factor’ [69]. Since, from convention, this Sudakov factor is defined from a lower cut-off value t_0, it is appropriate to use in the description of initial state showering. For final state showering however, a parton typically starts at a t_{max} and evolves to a $t < t_{max}$. The no-branching factor in this case becomes $\mathcal{P}_{no-branching}(t_{max}, t) \ = \ S_a(t_{max})/S_a(t)$.

In the case of initial state parton showering, the evolution of the shower is implemented ‘backwards’. For a parton b entering the hard event, the code tries to reconstruct what happened before by the definition of a conditional probability: if parton b is present at scale t, what is the probability that it will turn out to come from a branching $a \to b\,c$ at some infinitesimally smaller scale.

---

[3] This exponential term can perfectly be compared with the exponential factor arising in the description of radioactive decay. All atoms already decayed can no longer decay.

parameter	nominal value	negative shift	positive shift
Λ_{QCD}	0.25 GeV	0.15 GeV	0.35 GeV
${Q^2_{max}}^{ISR}$	1	0.25	4
${Q^2_{max}}^{FSR}$	4	1	16

Table 2.1: Most relevant parton shower `PYTHIA` parameters, with their respective nominal value and expected uncertainty at 14 TeV as suggested in [70].

A last comment then on the infrared singularities ($z \to 1$) of the splitting functions defined in eq. 2.13. To get rid of these divergences, an explicit infrared cut-off value must be introduced, $z < 1 - \epsilon(t)$. A Sudakov form factor defined with such a cut-off is the probability of evolving from t_0 to t without any 'resolvable' branching. Hence, the no-branching probability is sum of the virtual parton loop contributions and the real, but unresolvable contributions: both terms are divergent, but their sum is finite.

2.4.4 Initial and Final State Radiation in `PYTHIA`

The parton-showering available in the `PYTHIA` Monte-Carlo generator will be applied in the simulation of all datasets used in the presented top quark analyses. The following three parameters are of main importance in the description and treatment of ISR and FSR by `PYTHIA`:

- Λ_{QCD} [`PARP`(61), `PARP`(72), `PARJ`(81)]:
  This parameter was introduced in the definition of t (Eq. 2.11) and refers to the general QCD scale parameter used in the evaluation of the running α_s and consequently in the DGLAP-equations. `PARP`(61) is the parameter value for space-like showers, `PARP`(72) the value for time-like showers, except in the decay of resonances. In this last case `PARJ`(81) should be set.

- ${Q^2_{max}}^{ISR}$ [`PARP`(67)]:
  The Q^2 scale of the hard scattering, Q^2_{hard}, is multiplied by this parameter to define the maximum parton virtuality allowed in space like showers.

- ${Q^2_{max}}^{FSR}$ [`PARP`(71)]:
  Analogue, this parameter scales Q^2_{hard} for the maximum allowed scale to start from in time-like showers. The default is taken 4 times ${Q^2_{max}}^{ISR}$, extracted from the fact that in a 2×2-scattering process, $Q_{hard} = p_T < Q/2$. This equation however expects massless particles produced in the final state.

For the above parameters, the `PYTHIA` nominal values and their uncertainties as suggested in [70] are listed in Table 2.1. These values were cross-checked with the ones adopted by the `CDF` experiment. When studying the systematic effect arising from the uncertainty on the parton-showering method, these three parameters should be varied in the listed interval.

2.5 Matching parton level events with a parton shower

Whether one uses a LO matrix element generator or a NLO calculation for the simulation of the hard events, a meaningful event description is only possible if a removal of double counted jet configurations is possible. Hard emissions during the parton shower (PS) evolution, as well as extra partons generated in the matrix element (ME) might both result in the same n-jet final state. Solutions are available, and known as PS/ME matching algorithms.

While the LL matching of LO ME calculations cannot be expected to accurately reproduce the inclusive NLO cross-section and its stability w.r.t. scale variations, the exact LO description of higher multiplicity partonic final states will give a better accuracy for the distributions of the extra jets. ME are good in describing the kinematics of well separated jets, while PS is more appropriate for the structure inside a jet. In this respect the two approaches are complementary, both in goals and in expected performance.

Depending on the choice to start from a LO matrix element generator or a NLO calculation, two different procedures are developed to match the parton shower evolution to the simulated parton final states.

2.5.1 Vetoed parton showers

The objective of this matching strategy is to define a generic approach to combine several orders of emission multiplicity in a consistent way. The trick in ME/PS matching is to use the information of the Sudakov form factors (Eq. 2.16) to ensure a detailed balance between the different jet multiplicity cross-sections. Every shift in a jet resolution resulting in a drop of an n-jet multiplicity cross-section should be compensated by an increase in cross-section in e.g. the $n+1$-jet cross-section.

A few alternative algorithms exist, of which the two famous ones are the CKKW [71] and the MLM algorithm [72]. As this MLM matching procedure is employed in all `AlpGen` generated event samples studied in this work, its main principles are briefly touched:

1. Pick up an unweighted event, which was generated in a phase-space region constrained by the ME ‘generation cuts’:

$$p_T^{lp} > p_T^{min} \ , \quad |\eta_{lp}| < |\eta_{max}| \ , \quad \Delta R > \Delta R_{min} \ , \tag{2.18}$$

   where p_T^{lp} and $|\eta_{lp}|$ are the transverse momentum and the pseudorapidity of the light final-state partons, and ΔR is their minimal separation in (η, ϕ)-space. These cuts are needed to stay away from the singular collinear and infrared phase-space regions.

2. Reconstruct an imagined shower history that describes how the event could have evolved from the ‘core’ 2×2 process to the n-jets final state. That provides an ordering of emissions by whatever shower-evolution variable is intended (e.g. the t-variable defined in eg. 2.11).

3. Apply parton showering on all events of each n-light parton sample (with $n = 0 \ldots N$).

4. For each event, perform a jet reconstruction on the showered particles (before hadronisation). For this jet clustering, `AlpGen` uses the GetJet cone algorithm [73], with a jet cone opening angle ΔR^{clus}. Clusters are accepted if their transverse energy exceeds a threshold value E_T^{clus}.

5. Associate each parton from the parton level event with a reconstructed cluster, using as criterion $\Delta R_{parton-cluster} < \Delta R_{match}$. If each parton has its match, the algorithm continues.

6. If $n < N$ and the number of clusters exceeds the number of partons ($N^{clus} > n$), the event is skipped.

7. In the case of the highest multiplicity sample ($n = N$), all well matched events are kept if the extra non-matched clusters are softer than the matched ones.

8. Combine all exclusive samples with partonic multiplicity $n = 1 ... (N-1)$ with the inclusive event sample with $n = N$ to obtain a fully inclusive high multiplicity sample.

It should be remarked that the presence in this approach of the extra generation and matching parameters compared to the usual parton level or shower-only approaches is not necessarily bad. The ultimate use of LL event generators is not to incorporate and enable high-precision predictions of QCD, but rather to provide the most faithful representation of the data, so that the experimental searches for and studies of new phenomena can be built on a solid foundation. The uncertainties introduced by the possibility to change the matching and generation parameters should therefore be seen as an opportunity to optimise, via their fitting, the agreement between the generator and the data.

2.5.2 The `MC@NLO` generator

`MC@NLO` [74] in some respects is the most ambitious approach: it aims to get not only real but also virtual contributions correctly included, so that cross-sections are accurate to NLO, and that NLO results are obtained for all observables when formally expanded in powers of α_s. Thus hard emissions should again be generated according to ME, while soft and collinear ones should fall within the PS regime. In simplified terms, the scheme works as follows:

1. Calculate the NLO ME corrections to an n-body process, including $n+1$-body real corrections and n-body virtual ones.

2. Calculate analytically how a first branching in a shower starting from a n-body topology would populate $n+1$-body phase space

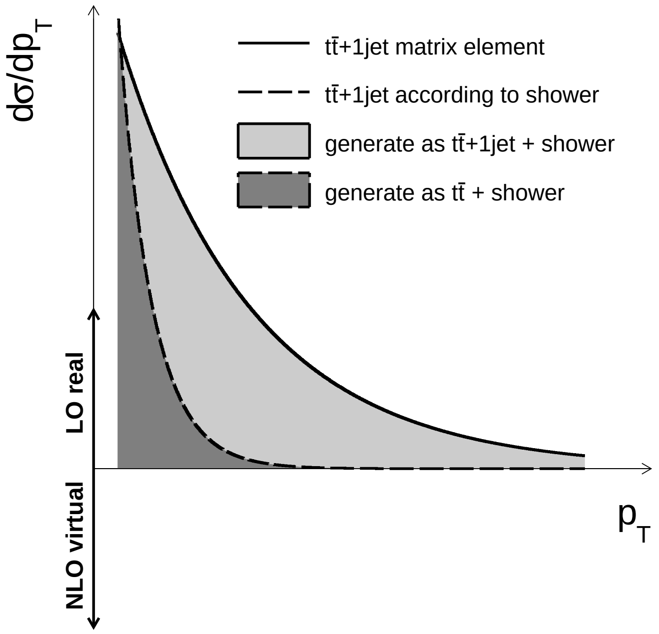


Figure 2.6: Schematic representation of MC@NLO applied to $\mathrm{t\bar{t}} + 1\mathit{jet}$.

3. Subtract the shower expression from the $n+1$ ME one to obtain the 'true' $n+1$ events, and consider the rest as belonging to the n-body class. The PS and ME expressions agree in the soft and collinear limits, so the singularities there cancel, leaving finite cross-sections both for the n- and $n+1$-body event classes. This is illustrated in Figure 2.6.

4. Now add showers to both kinds of events.

A technical problem is that, although ME and PS converge in the collinear region, it is not guaranteed that the ME contribution is everywhere above PS one. This is solved by having a small fraction of events with negative weights. In summary, `MC@NLO` is superior in that it does provide the total cross-section for a process to NLO accuracy, and hence includes virtual terms. It is therefore an essential tool for precision tests, or to rescale the LO cross-sections with a constant K-factor. In principle this rescaling is only meaningful if this K-factor is indeed flat with respect to the various event observables. As we will illustrate in the paragraph 2.6.2, this requirement is well approximated when using multi-parton LO ME generators like `AlpGen` compared to pure PS LO productions such as `PYTHIA`.

2.6 Study of several event generators for top quark physics

The main objective of this section is to motivate the choice for the `AlpGen` event generator for the simulation of all data samples used as an input to the top quark analyses presented in the next chapters. The main argument may even be the most logic one: experiments at Tevatron show a very good agreement between the simulated observable distributions and the data. This will be illustrated in paragraph 2.6.3. Compared to a LO event generator like `PYTHIA` or `TopRex`, `AlpGen` is also expected to produce spectra closer to the `MC@NLO` simulation, on which we will come back in paragraph 2.6.2. Firstly, the adequate performance of the MLM matching technique is demonstrated.

2.6.1 Evaluation of the MLM matching method

The ability of the `AlpGen` event generator to give an accurate description of the observed event kinematics strongly depends on the method applied to merge the different jet multiplicity subsamples. In principle, a change in the sample generation or matching definition only result in a different distribution of events over the $\mathrm{t\bar{t}}$ jet multiplicity subsamples. As an example, an increase of the minimal light jet p_T^{min} parameter defined in the MLM algorithm (cfr. paragraph 2.5.1) would lead to a higher number of events in the $\mathrm{t\bar{t}}0j$ bin. To compensate this loss of radiation, relatively more jets will be produced during the parton showering. However, overall quantities like the $\mathrm{t\bar{t}}$ production cross-section should stay invariant under these choices.

An important request of the ME/PS matching algorithms is the need for smooth spectra for all physical observables. No knicks due to the transition from one multiplicity to another are allowed. This requirement was checked on the `AlpGen` sample to be used as signal in the upcoming $\mathrm{t\bar{t}}$ analyses. This sample consist of several subsamples, defined according to the number of light partons n present in the ME, and whether the extra jet veto is applied: n_{excl} (n_{incl}). For example, 1_{excl} denotes the event sample obtained after the showering, matching and extra-jet veto of a set of $\mathrm{t\bar{t}}$ plus 1 extra parton events. The nominal `AlpGen` production used in this work combines the following such subsamples: $0_{excl} + ... + 3_{excl} + 4_{incl}$, and will be defined as an S_4 sample. Hence, up to four extra partons were included in the matrix element calculations. All relevant generation cuts as well as all MLM algorithm parameters used in this production are listed in Table 2.2. Furthermore, the `CTEQ5L` fit is chosen as parton distribution functions, while the factorisation and renormalisation scheme is set to

$$\mu^2 = \frac{1}{2}\left(p_{T,t}^2 + m_t^2 + p_{T,\bar{t}}^2 + m_{\bar{t}}^2\right) . \tag{2.19}$$

The overall and subprocess cross-sections are given in Table 2.3, and compared with the inclusive LO cross-section obtained with the `PYTHIA` and `TopRex` generator. The difference in inclusive cross-section between `AlpGen` and the LO generators is observed to be somewhat higher than the fluctuations expected due to the choice of generation or matching cuts [25]. However, it should be noted that the choice of p_T^{min} at 20 GeV/c results in more statistics in the 1_{excl} sample compared to in the 0_{excl} one. This indicates that this cut value is too low to observe a proper QCD scaling or perturbative expansion in α_s when moving from one multiplicity to another. According to the results documented in [75], this difference in inclusive cross-section does indeed decrease for `AlpGen` samples produced with higher p_T^{min} values.

The addition of the different contributions in the construction of the S_4 sample is illustrated in Figure 2.7 for six physical observables. The upper plots show the

generator cuts	matching cuts
$p_T^{light\ parton} > 20\,GeV/c$	$E_T^{clus} > 15\,GeV$
$\lvert\eta\rvert^{light\ parton} < 5$	$\lvert\eta\rvert^{clus} < 5$
$min\,\Delta R(j_i, j_j) = 0.7$	$min\,\Delta R(p_i, j_i) = 1.5 \times 0.7 = 1.05$

Table 2.2: Default generation and matching cut values applied in the simulation of the `AlpGen` samples.

	0_{excl}	1_{excl}	2_{excl}	3_{excl}	4_{incl}	S_4	`PYTHIA`	`TopRex`
Cross-section (pb)	190	170	100	40	61	561	492	488

Table 2.3: Inclusive and exclusive (sub)sample cross-sections for the `AlpGen` S_4 $\mathrm{t\bar{t}}$-production (in pb), compared with the LO `PYTHIA` prediction. All generation and matching cuts used in the simulation of this sample are listed in Table 2.2.

transverse momentum (p_T^{top}) and pseudorapidity (η^{top}) of the (anti-)top quark, while on the second row the transverse momentum of the $\mathrm{t\bar{t}}$-system ($p_T^{\mathrm{t\bar{t}}}$) and the azimuthal angle between the t and $\bar{t}$ quarks ($\Delta\phi^{\mathrm{t\bar{t}}}$) is plotted. The generation of extra light quarks in the matrix element requires on average a harder Q^2 scale. This explains the harder p_T^{top} spectrum observed for the higher jet multiplicity bins. Also the disappearance of the dip at $\eta = 0$ can be understood, as the Lorentz-boosted longitudinal p_z^{top} component gets less impact on the particle's direction. An observable which is even more sensitive to the change in scale is the transverse momentum of the $\mathrm{t\bar{t}}$-system. From this plot, a first hint is given that the matching algorithm succeeds well in the merging of the different subsamples, demonstrated by the smooth overall S_4 curve. The fourth plot of the azimuthal angle between the top and anti-top quark illustrates the impact of initial state radiation in the different subsamples. Where for events in the 0_{excl} sample only relatively soft initial state radiation is allowed as a result of the parton showering, the other bins suffer from hard radiation produced in the ME. The higher the multiplicity the higher the probability for such a hard radiation to occur. The possible assumption that the top and anti-top quark are produced back to back, even in the transverse plain, is hence highly smeared due to the presence of ISR.

Where all previous plots illustrate the kinematics of the several subsamples on parton level, the bottom two plots in Figure 2.7 are constructed after the hadronisation. By means of an iterative cone jet algorithm with opening angle 0.5 all stable final state particles are clustered into jets. Only jets with a transverse energy exceeding $20\,GeV$ and a pseudorapidity $|\eta| < 5$ were selected. To be more sensitive to jets arising from the extra light parton radiation, all jets clustered close to top decay products are rejected ($\alpha_{jp} < 0.2$). In the bottom left plot the number of remaining jets is shown, while in the right plot the transverse energy of the jet with the second highest E_T-value is drawn.

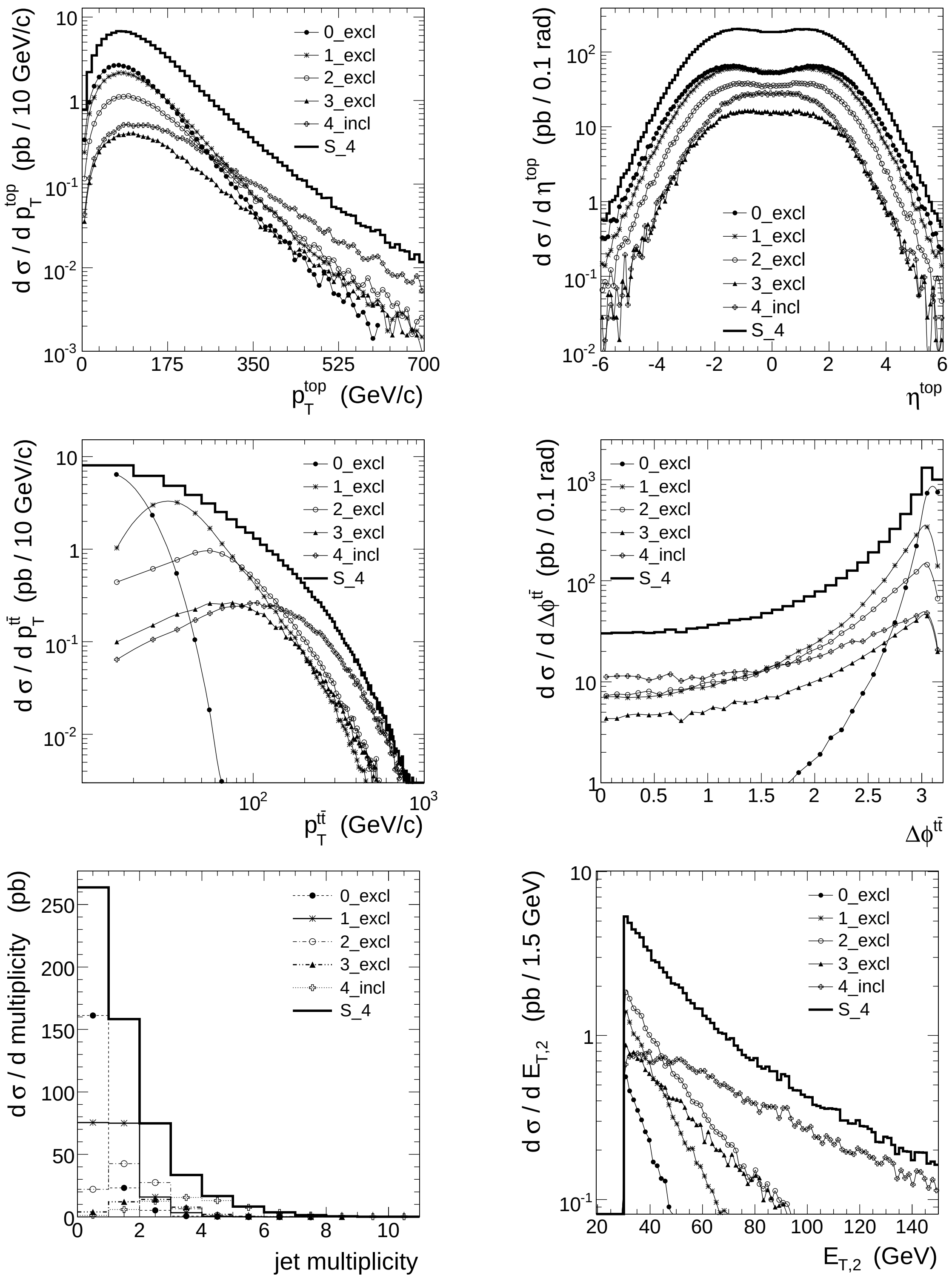


Figure 2.7: The differential cross-sections of the different $t\bar{t} + nj$-subprocesses are plotted as a function of some kinematic observables.

Again, no significant knicks can be identified in the distributions, indicating a smooth transition between the subsamples. These results give confidence that the matching algorithm allows to merge different jet multiplicity subsamples with a proper removal of double counting.

2.6.2 The `AlpGen` $t\bar{t}$ S_1 sample versus `MC@NLO` and `TopRex`

We will now compare in detail the description of $t\bar{t}$ events as provided by `AlpGen` and `MC@NLO`. For consistency with the `MC@NLO` approach, where only the $\mathcal{O}(\alpha_s^3)$ ME effects are included, we have constructed an inclusive `AlpGen` sample by summing the 0_{excl} and 1_{incl} contributions[4]. It is expected that the overall shapes of physical observable distributions agree much better for both programs, compared to pure Born Level process generators such as `TopRex` [76]. This, due to the fact that both `AlpGen` and `MC@NLO` give a correct LO description of $\mathcal{O}(\alpha_s^3)$, i.e. the emission of one extra light parton.

To approximate the `MC@NLO`'s default configuration, the same top mass (175 GeV/c^2), and factorisation and normalisation scale $\mu^2_{R/F}$ are used in both the `MC@NLO` and the `TopRex` generator:

$$\mu^2_{R/F} = m^2_{top} + p^2_{T,top}. \tag{2.20}$$

Additionally, the `CTEQ6M` NLO PDF set was chosen in both the `MC@NLO` and `AlpGen` simulation, while the `CTEQ5L` fits were employed in the `TopRex` sample. Furthermore, it should be mentioned that for both the `AlpGen` and `TopRex` generators the hadronisation is performed by the `PYTHIA` program, while currently only `HERWIG` can be interfaced to `MC@NLO`.

The results of the comparison are summarised in Figure 2.8 and 2.9. Where the former figure shows inclusive properties of the $t\bar{t}$-system, the latter illustrates kinematics of the jets produced in semi-muonic decaying $t\bar{t}$ events. In both sets of figures, the lower insets represent the relative difference between the `MC@NLO` and `AlpGen` results, $[d\sigma(NLO) - d\sigma(S_1)]/d\sigma(S_1)$.

The upper row of plots in Figure 2.8 shows the transverse momentum and pseudorapidity of the top and anti-top quark, p_T^{top} and η^{top}. For both plots the overall agreement is good and no large discrepancy is seen between the three spectra. This is no longer the case for the $p_T^{t\bar{t}}$ and $\Delta\phi^{t\bar{t}}$ distributions, which are both highly sensitive to a recoiling jet against the $t\bar{t}$ system. For generators capable to incorporate the description of such an extra jet production in the ME calculation, a harder jet spectrum, and consequently $p_T^{t\bar{t}}$ spectrum is expected and observed in Figure 2.8. Another manifestation of this jet recoil is the broader $\Delta\phi^{t\bar{t}}$ distribution. Where in the case of `TopRex` only the initial state radiation and the underlying event are responsible for a change in azimuthal angle between the $t\bar{t}$-pair, the extra outgoing parton produced in the $\mathcal{O}(\alpha_s^3)$ `MC@NLO` or `AlpGen` simulation will further smear this angle. When we focus on the comparison of the `MC@NLO` and `AlpGen` description, there is excellent agreement for $p_T^{t\bar{t}}$ and $\Delta\phi^{t\bar{t}}$ as soon as we move away from the regions dominated by Sudakov effects ($p_T^{t\bar{t}} = 0$ and $\Delta\phi^{t\bar{t}} = \pi$). Although both the programs should be able to include these effects correctly, a relative important difference in event rate (of the order of about 25%) is observed. Based on the results described in [25] a larger suppression

---

[4] Both subsamples were simulated with a generator cut value $p_T^{min} = 40$ GeV/c. Cross-sections of 250 and 251 pb were obtained for respectively the 0_{excl} and 1_{incl} sample.

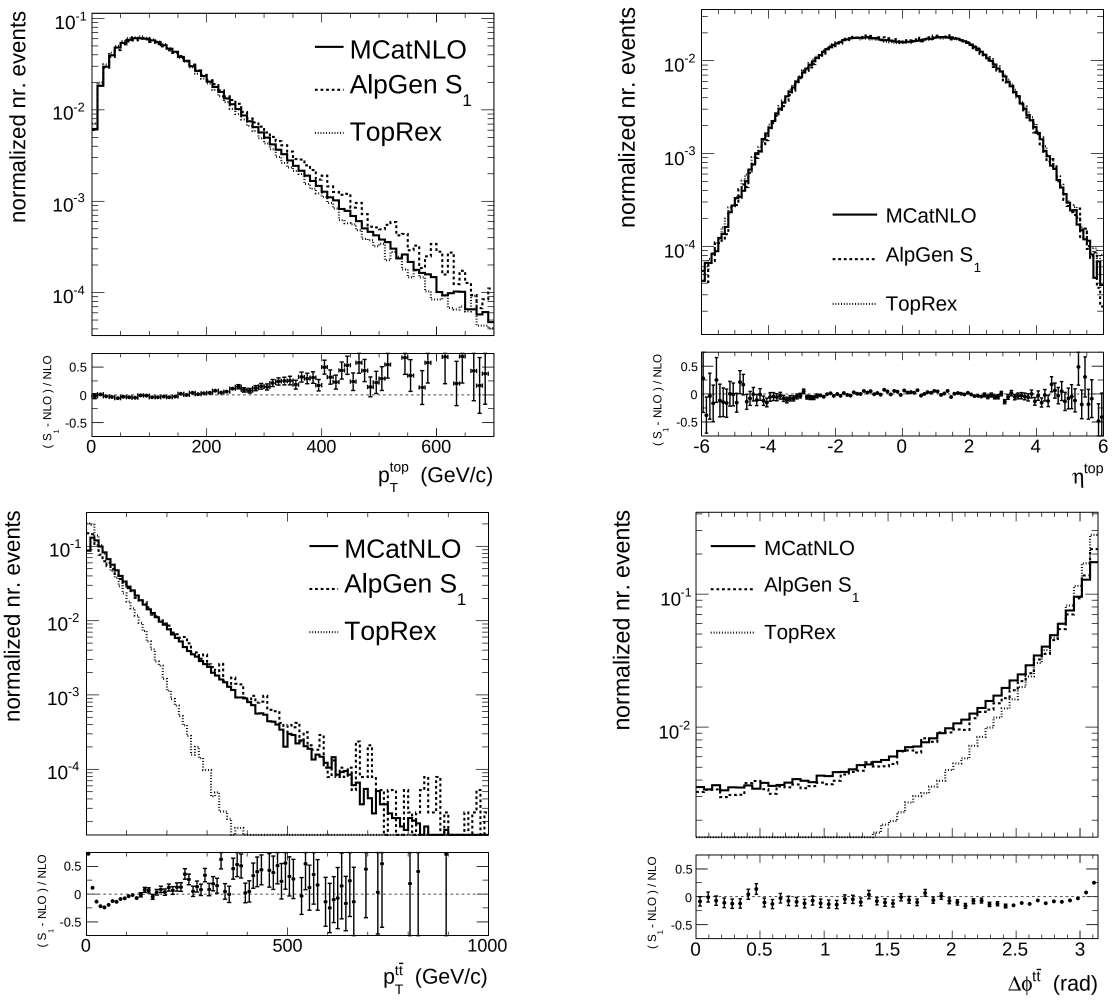


Figure 2.8: The kinematics of the $t\bar{t}$-pairs produced in the `AlpGen` S_1 sample is compared to the NLO `MC@NLO` and parton level `TopRex` predictions. The lower insets correspond to the relative difference between the `MC@NLO` and `AlpGen` results, $[d\sigma(NLO) - d\sigma(S_1)]/d\sigma(S_1)$.

of the Sudakov dominated regions is obtained for the `MC@NLO` generation. A plausible reason for this discrepancy might be the different PS description (`HERWIG` vs. `PYTHIA`).

In Figure 2.9 some results are shown of observables defined after the hadronisation. By means of an iterative cone algorithm with opening angle $R = 0.5$, all stable generated particles[5] are clustered in jets. Only jets with a transverse energy exceeding 30 GeV and pseudorapidity in $|\eta| < 2.4$ are selected in order to concentrate on those kinematic regions that are selected in the upcoming top quark analyses. Additionally, all jets with a well matching top quark decay product ($\alpha_{jp} < 0.2$) are removed from the list. This, to be more sensitive to jets arising from the hard parton radiation in the ME or PS. After this filtering, the number of reconstructed jets is illustrated for all three generators in the upper left plot of Figure 2.9. As expected, the fraction of

---

[5] For a definition of the Iterative Cone algorithm we refer to Section 4.1.2.

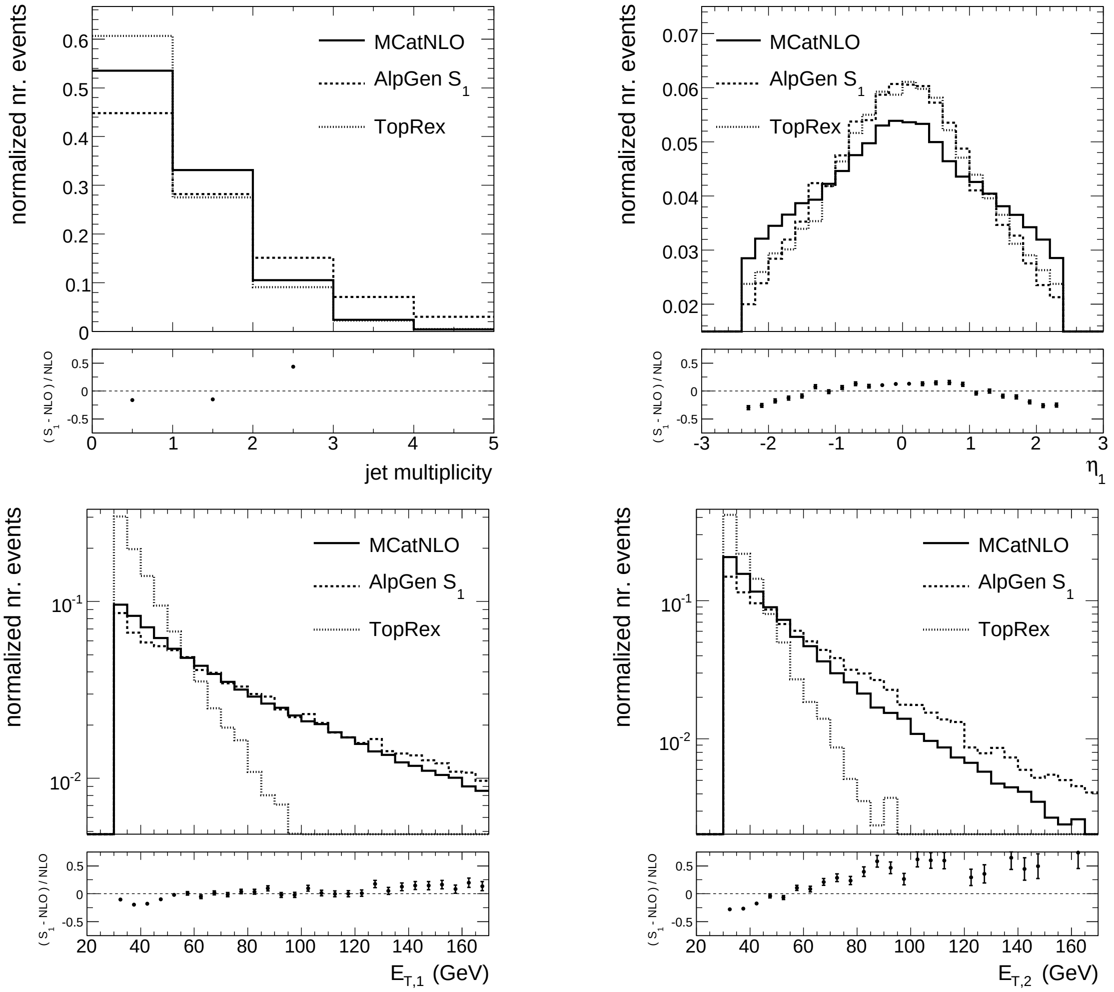


Figure 2.9: Comparison between jet observables distributions in the inclusive `AlpGen` S_1 sample and the NLO `MC@NLO` and parton level `TopRex` prediction.

events without any additional hard E_T jet clustered in the event is significantly higher for the LO parton level generator. However, more difficult to understand is the important difference between the `AlpGen` and `MC@NLO` description. Nevertheless, if we focus on the kinematics of the first two leading E_T jets among the remaining jets in the event, the agreement is acceptable, especially what concerns the transverse energy spectrum. Apparently, the leading jet is produced on average more central with the `AlpGen` program, what might explain the higher jet multiplicity after the $|\eta| < 2.4$ cut compared to the `MC@NLO` prediction.

In general the above observations motivate the use of the `AlpGen` program over a LO parton level generator like `TopRex` or `PYTHIA`, because the description of physics observables offered by `AlpGen` matches much better the NLO result. Additionally, these observations strengthen the belief that the observable shapes obtained after the merging of higher parton multiplicities subsamples (like e.g. in the `AlpGen` S_4 sample) approximate well the higher order corrections. Also a rescale of the inclusive cross-section of a certain process towards the most accurately calculated theoretical value will be more reliable knowing that the observable's distributions have the correct shape, and hence the so-called K-factor can be assumed constant over phase-space.

2.6.3 Validation of `AlpGen` with CDF data

The measurement of the cross-section for the inclusive W-boson production offers a suitable way to study the extra jets produced at a hadron-hadron collider. In the CDF experiment at Fermilab, a new measurement of the $W + jet$ cross-section as a function of relevant jet kinematic variables was presented for $320 \pm 18\,\mathrm{pb}^{-1}$ of Tevatron Run-II data [77]. Figure 2.10 shows the differential cross-section as a function of the transverse energy of the first, second, third and fourth jet if existing in the event. These cross-sections were corrected to the parton level jets, and defined in a limited W -boson decay phase space to ensure an accurate theoretical reproduction, minimising the model dependence that can enter a correction back to the full W-cross-section. Only $W \to e\nu$ events were selected with ${E_T}^e > 20\,\mathrm{GeV}$, $|\eta^e| < 1.1$ and $E_T^{miss} > 30\,\mathrm{GeV}$[6]. Jets were clustered using an iterative seed-based cone algorithm, with cone radius $R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.4$. Only jets with a transverse energy exceeding $E_T^{jet} > 15\,\mathrm{GeV}$ and pseudorapidity $|\eta^{jet}| < 2.0$ were selected. The process backgrounds considered in this study are QCD, $W \to \tau\nu$, $Z \to ee$, WW and top quark production. To extract the background fraction in each $W + N\,jet$ sample the E_T^{miss} distribution of the candidates is fitted to a Monte-Carlo template with signal and background contributions. The uncertainty of this template is accounted for in the systematics.

After normalising the total simulated $W + njet$ cross-section to the data, for each jet multiplicity the obtained jet spectrum is reasonably well described by the `AlpGen`+`PYTHIA` event samples. For the $W+ \geq 2\,jet$ sample, the shape of the cross-section as a function of the 2-leading jets' angular correlation points out to be well modelled too (Figure 2.11).

In can hence be concluded that also the data indicate the multi-parton ME generators such as `AlpGen` perform well in the description of the observed physics. At the Large Hadron Collider (cfr. Chapter3), proton-proton collisions at an increased centre-of-mass energy will produce harder initial and final state radiation spectra. Consequently, the call for a generator that is able to reproduce the observed jet distributions is even louder.

2.7 Non-perturbative QCD-evolution of the showers

After the parton shower described in Section 2.4.1 has terminated, a set of partons are left with virtualities around the cut-off scale t_0 (cfr. Eq. 2.11). At this time-scale, corresponding with Q^2-values in the order of Λ^2_{QCD}, the couplings become strong, and perturbative QCD is no longer valid. The most important effect is the conversion

---

[6] The pseudorapidity η is defined in Eq. 3.1, while the missing transverse energy E_T^{miss} is introduced in Section 4.3.

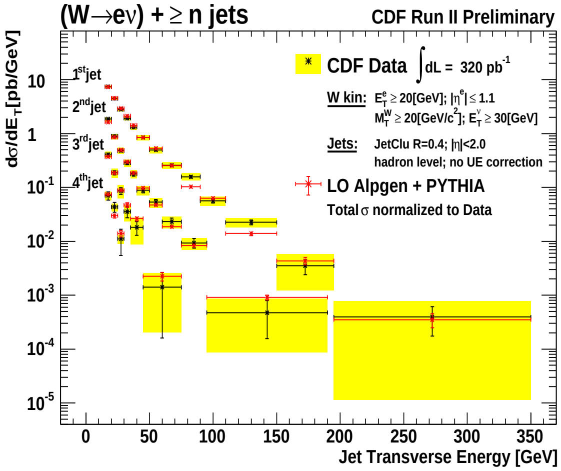


Figure 2.10: Differential cross section $d\sigma(W \rightarrow e\nu + N\,jets)/dE_T^{jet}$ for the 1st, 2nd, 3rd and 4th inclusive jet sample. Data are compared to `AlpGen`+`PYTHIA` predictions.

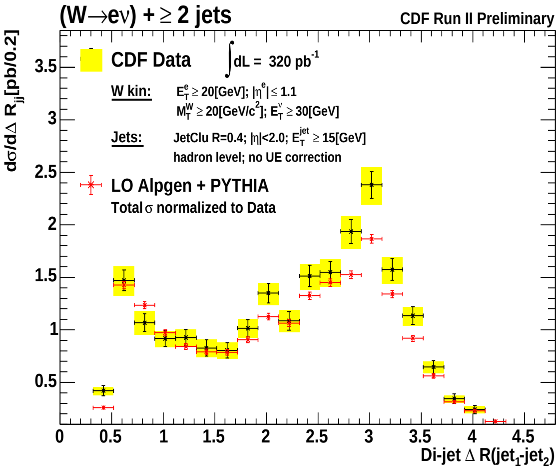


Figure 2.11: Differential cross section $d\sigma(W \rightarrow e\nu + N\,jets)/dR_{j1j2}$ as a function of the angular separation of the 2 leading jets in $W+ \geq 2\,jet$ events.

of partons into stable, experimentally observable hadrons. This process, denoted as hadronisation or fragmentation, has yet to be understood from first principles. As a consequence, several phenomenological models are developed and implemented as extra modules in the general purpose MC generation programs. In the `PYTHIA` application, the 'Lund' string fragmentation model is chosen as default scheme [78].

In this widely used string hadronisation model colour singlet partons are converted into colour neutral objects via an iterative string break-up scheme. It is based on the assumption of linear confinement, according to which the energy stored in the colour dipole field between a charge and an anti-charge increases linearly with the separation between the charges, if the short-distance Coulomb term is neglected. Typically, the string's rest energy density estimated to be $\kappa \simeq 1\,\mathrm{GeV/fm}$. In this confinement picture, the virtual quark-antiquark pair can be seen as a flux-tube stretched between both opposite colour charges via a number of intermediate gluons. As the q and $\overline{\mathrm{q}}$ move apart, the potential energy stored in the string increases, and the string may break by the production of a new $\mathrm{q}'\overline{\mathrm{q}}'$ pair, so that the system splits into two colour-singlet systems $\mathrm{q}\overline{\mathrm{q}}'$ and $\mathrm{q}'\overline{\mathrm{q}}$. If the invariant mass of either of these string pieces is large enough, further breaks may occur. In the Lund string model, the string break-up process is assumed to proceed until only on-mass-shell hadrons remain, each hadron corresponding to a small piece of string with a quark in one end and an antiquark in the other.

The production of baryons is still a poorly understood area. In the simplest possible approach, a diquark in a colour antitriplet state is just treated like an ordinary antiquark, such that a string can break either by quark-antiquark or antidiquark-diquark pair production. A more complex scenario is the 'popcorn' one, where diquarks as such do not exist, but rather quark-antiquark pairs are produced one after the other.

In order to generate the quark-antiquark pairs $\mathrm{q}'\overline{\mathrm{q}}'$ which lead to string break-ups, the Lund model invokes the idea of quantum mechanical tunnelling. Since the string is assumed to have no transverse excitations, any $p_\perp$ is locally compensated between the quark and the antiquark of the pair. In terms of the mass m and the transverse momentum $p_\perp$ of the created (anti)quark, the tunnelling probability is given by

$$\mathcal{P}^{\,tunnelling} = \exp\left(\frac{-\pi m^2}{\kappa}\right) \exp\left(\frac{-\pi p_\perp^2}{\kappa}\right) \tag{2.21}$$

This leads to a flavour-independent Gaussian spectrum for the $p_\perp$ of $\mathrm{q}'\overline{\mathrm{q}}'$ pairs. Some contribution of very soft perturbative gluon emission are also included in this description, as a possible explanation why the experimentally measured average $p_\perp$ is somewhat higher than obtained with the formula above. In the `PYTHIA` program, this tunnelling is regulated by one parameter, corresponding to the root-mean-square $p_\perp$-value. Varying this parameter therefore tunes the width of the Gaussian $p_\perp$ spectrum.

Furthermore, the mass term in Eq 2.21 implies a suppression of heavy-quark production, $u : d : \mathrm{s} : c \approx 1 : 1 : 0.3 : 10^{-11}$. Charm and heavier quarks hence are not expected to be produced in the soft fragmentation, but only in perturbative parton-shower branchings $\mathrm{g} \rightarrow \mathrm{q}\overline{\mathrm{q}}$, as is qualitatively confirmed by experiment. Once the flavours q_{i-1} and $\overline{\mathrm{q}}_i$ have been selected, a random choice is made between the possible multiplets, based on experimentally measured data.

Longitudinal hadron momenta are determined by means of the symmetric Lund functions, based upon the general principles such as causality, Lorentz covariance and confinement. These functions express the probability that a hadron consumes a given fraction z of the available energy-momentum and are described by the parameters a and b for light quarks (u,d,s):

$$f(z) \propto \frac{(1-z)^a}{z} \exp(\frac{-bm_\perp^2}{z}), \tag{2.22}$$

where $m_\perp^2 = m^2 + p_\perp^2$ is the transverse mass squared of the meson. As parameter b is strongly anti-correlated with a, only one of both should be varied in its uncertainties to check the sensitivity of a given measurement towards the tuning of this Lund string fragmentation model.

Charm and bottom data clearly indicate the need for a harder fragmentation function for heavy flavours, that result in on average higher $f(z)$ values. The best known of these is the Peterson/SLAC formula [79]

$$f(z) \propto \frac{1}{z\left(1 - \frac{1}{z} - \frac{\epsilon_q}{1-z}\right)^2}\ , \tag{2.23}$$

where ϵ_q is a free parameter, expected to scale between flavours like $\epsilon_q \propto 1/m_q^2$.

The effect of this difference in hadronisation function between light (u,d,s) and heavy (c,b) quark flavours on experimental observables is visualised in Figure 2.12. Both plots are constructed using only jets with a well matching initial quark direction ($\alpha_{jq} < 0.2$) in a selected kinematic window: the jet is required to have a transverse energy between 50 and 80 GeV, and a pseudorapidity $\eta < 0.5$. For the reconstruction of the jets the iterative cone algorithm was employed with an opening angle $R = 0.5$,

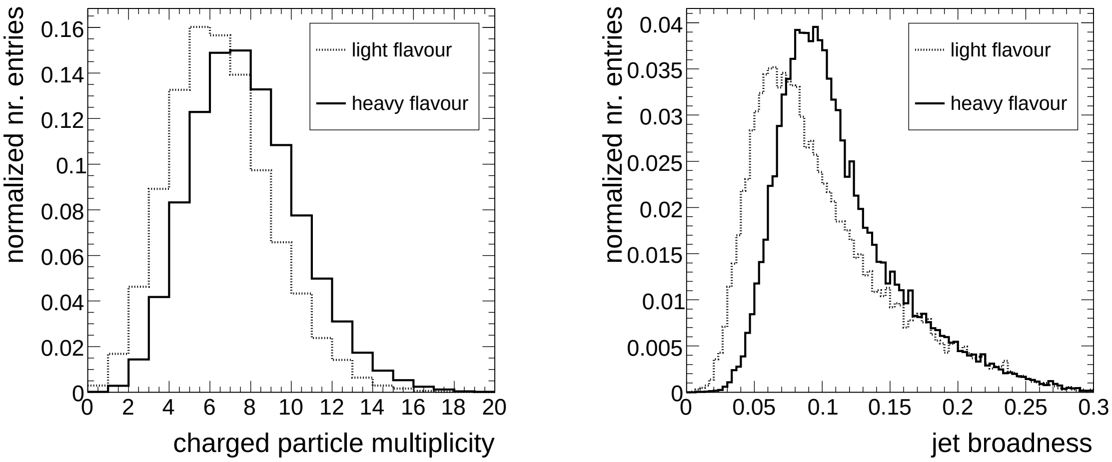


Figure 2.12: Difference in charged particle's multiplicity in a jet(left) and the jet broadness (right) due to the harder heavy quark flavour hadronisation function.

using as input all stable hadron-level Monte-Carlo particles[7] produced in semi-muonic decaying $t\bar{t}$ events. For all selected jets, the charged particle's multiplicity is illustrated in the left plot, while the jet broadness[8] is drawn on the right, both for light and heavy quark flavour jets. Only jet constituents with a transverse momentum exceeding 1 GeV/c are accepted in the left plot, while a minimum of three jet constituents was required in order to be filled in the broadness histograms. Clearly, heavy quark jets have on average more charged particles produced during their fragmentation. Where for light quark jets a mean multiplicity of 6.5 charged particles per light quark jet was found, this value increases to 7.8 in the case of heavy flavour jets. The difference in fragmentation also results in on average broader heavy flavour jets. This effect can be seen as a direct manifestation of the harder Peterson fragmentation function compared to the Lund fragmentation function.

In the `PYTHIA` application, the hadronisation process is hence configured via four main parameters [70], of which the default values and uncertainties are given in Table 2.4. The listed values correspond to the values obtained from the fit performed to the `OPAL` data [80].

---

[7] For more information on the definition and performance of the iterative cone jet algorithm we refer to Section 4.1.

[8] The broadness of a jet is defined as:

$$B^{jet} = \frac{\sum_{i=0}^{n} p_{\perp}{}^{i}}{\sum_{i=0}^{n} p^{\,i}} = \frac{\sum_{i=0}^{n} |\,p^{\,i} \sin\left(acos\left(\vec{p}^{\,i}.\vec{p}^{\,jet}\right)\right)|}{\sum_{i=0}^{n} p_i}, \tag{2.24}$$

with $p_\perp$ the transverse momentum of a object with respect to the jet axis.

parameter	`PYTHIA`	nominal value	uncertainty
rms $p_\perp$	`PARJ[21]`	0.40 GeV/c	0.03 GeV/c
Lund b	`PARJ[42]`	$0.52\,\mathrm{GeV^{-2}c^2}$	$0.04\,\mathrm{GeV^{-2}c^2}$
Peterson ϵ_c	`PARJ[54]`	-0.031	0.011
Peterson ϵ_b	`PARJ[55]`	-0.0041	0.0004

Table 2.4: Most relevant `PYTHIA` hadronisation parameters, with their respective nominal value and expected uncertainty at 14 TeV as suggested in [70].

2.8 Underlying event

In a hard proton-proton collision, only a fraction of energy of the incoming particles is consumed by the partons which participate to the hard event, leaving behind a highly energetic proton remnant. This remnant is colour connected to the hard interaction, and forms part of the same fragmenting system. Also, the composite nature of the two incoming beam particles implies the possibility that several parton pairs undergo separate hard or semi-hard scatterings during the same collision, which is denoted as 'multiple interactions'. Both contributions build up the 'underlying event' structure. Finally, in high-luminosity colliders like the LHC, several collisions between beam particles in the same or consecutive beam crossing, referred to as pile-up events, will contribute to the particle production activity that is to be observed by the detectors.

Huge progress in the phenomenological study of the underlying event in jet events has been achieved by the `CDF` experiment at the Tevatron collider, using the multiplicity and transverse momentum spectra of charged particles measured in regions away from the leading jets. Therefore only contributions are included from the underlying event and pile-up collisions. The average charged multiplicity per unit of pseudorapidity in these regions turns out to be significantly higher in QCD dijet events with respect to the one measured in minimum bias events. This effect, referred to as the pedestal effect, is well reproduced by accounting for a varying impact parameter for the two colliding hadrons. Indeed, the average number of interactions at a given proton-proton collision is expected higher for more central collisions. Because also the probability for a hard interaction increases if the total number of interactions goes up, the underlying event activity is more pronounced in harder dijet productions compared to minimum bias events. Analogue, an even higher energy flow due to the underlying event is expected in $\mathrm{t\bar{t}}$.

Despite the important progress in the understanding of the underlying event, and the possibility to tune the underlying event structure to energies reached at the Tevatron collider, an extrapolation to the LHC is not straightforward. The main problem of extrapolating the predictions of the multiple interactions models to the LHC is that some of the parameters are explicitly energy dependent, in particular the colour screening p_T cut-off $p_\perp^{UE}$. This parameter is used in the simulation of the multiple interactions and corresponds to the minimum momentum transfer in the $2 \rightarrow 2$ processes, to prevent that the differential cross section becomes divergent (for $p_\perp^{UE} \rightarrow 0$). For a sensible

parameter	`PYTHIA`	nominal value	3σ-uncertainty
$p_\perp^{UE}$	`PARP[82]`	2.9 GeV/c	0.5 GeV/c

Table 2.5: The most relevant `PYTHIA` underlying event parameter, with its respective nominal value and uncertainty.

estimation of the theoretical uncertainty arising from the underlying event, this model parameter is advised to be varied with three standard deviations around its nominal value [70]. The numerical values are given in Table 2.5, and were obtained after a fitting of the minimum bias `CDF` and `UA5` data and an extrapolation to the LHC energy scale.

Chapter 3

The CMS experiment at the Large Hadron Collider

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [81] is a 27 km circumference particle accelerator spanning the Swiss-French boarder like illustrated in Figure 3.1. The project is carried out at CERN (Centre Européenne pour la Recherche Nucléaire). Since November 2007 the collider's circle has completed installation and the first physics beam is expected in summer 2008. Once fully operational, the LHC will provide collisions between proton beams at a luminosity exceeding $10^{33}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ and a centre-of-mass energy of 14 TeV, which is an order of magnitude higher than the previous generation of colliders. Also, it will operate with heavy ion (Pb-Pb collision) beams.

There are four main experiments which will take place at the LHC: two with general purpose detectors, ATLAS [82] and CMS [83], and two with dedicated detectors, ALICE [84] and LHC-b [85] which will study heavy ion physics and B-physics respectively. Figure 3.2 shows the four experimental sites along the LHC ring. The CMS detector will be situated approximately 100 m underground and will be taking data from both the p-p and Pb-Pb collision runs.

3.1.1 Physics Motivation and Programme

One of the primary objectives of the LHC is to detect the Higgs boson which is predicted by the Standard Model of particle physics and is believed to be responsible for the electroweak symmetry breaking (EWSB) like described in Section 1.1.3. However, this is far from trivial, considering its small production cross-section that decreases with an increasing Higgs boson mass as is illustrated in Figure 3.3. Typically, around 10^5 Higgs bosons are expected to be produced during a year of running at a luminosity of $2\times10^{33}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ (with M_H =130 GeV/c^2), which is two orders of magnitude lower than the rate of $\mathrm{t\bar{t}}$ production.

From Figure 3.4 it can be observed that the Higgs boson may decay via a number of channels with branching ratio depending on the Higgs boson mass. It can be seen that the decays of the Standard Model Higgs boson into a pair of photons ($H_{SM} \rightarrow \gamma\gamma$) is

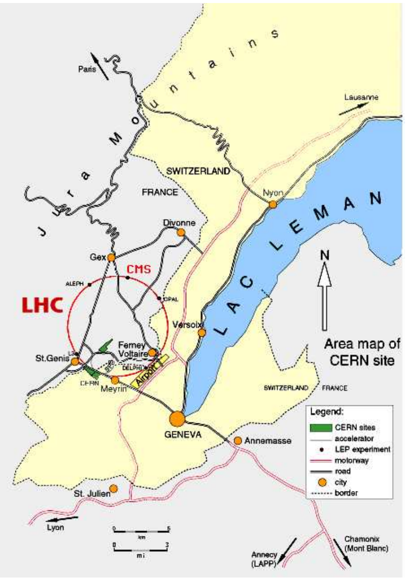


Figure 3.1: The geographical location of the Large Hadron Collider.

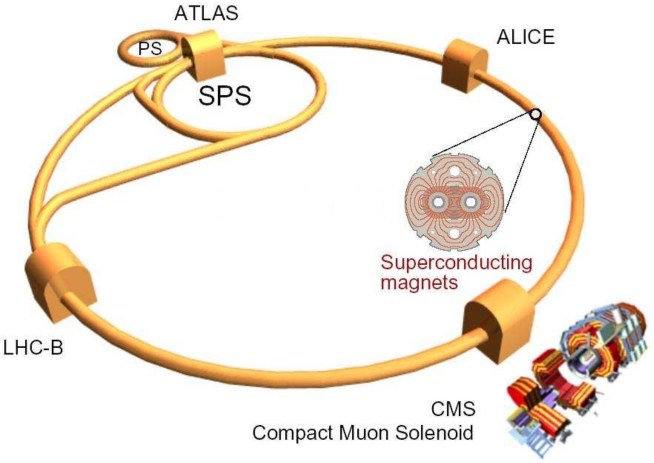


Figure 3.2: The main four high energy experiments located in the various interaction points of the Large Hadron Collider.

significant if the Higgs boson mass is below 150 GeV/c^2. Given the current upper bound of 144 GeV/c^2 (cfr. Section 1.2.4), this channel is expected to be the most probable for the detection of the SM Higgs boson. However, the measurement of m_H via this decay channel will pose stringent requirement on the detector's performance. We will come back to this in the next section. Decays to more massive vector bosons (e.g. $H_{SM} \rightarrow ZZ^{(*)}$) dominate the high, and less probable, mass range. In total around 10^4 Higgs bosons are expected to decay to two photons for a year of running at a luminosity of 10^{34}cm^{-2}s^{-1} (with M_H =130 GeV/c^2), which is 12 orders of magnitude lower than the rate of inelastic collisions and 4 orders of magnitude less than the $t\bar{t}$ production.

Apart from detecting the Higgs boson, the LHC aims for the discovery of new physics signals. As already introduced in Section 1.1.5 a model such as the Minimal SuperSymmetric Model (MSSM) predicts the production of new particles in the LHC energy acceptance. In the case R-parity is conserved and a Lightest Supersymmetric Partner (LSP) exists, the manifestation of SUSY might be visible as an important source of missing transverse energy in the detectors. Therefore, in order to be sensitive for these new physics, the detector should have a very hermetic coverage.

Another important task of the LHC experiments is to provide further understanding of the physics of already known particles. The LHC will produce large quantities of Standard Model particles, enabling studies of the properties of the most recently discovered top quark and CP violation, which are two of today's highly active research domains in elementary particle physics. With the high energy Pb-Pb collisions, the observation of a new state of matter called 'Quark-Gluon plasma' is envisaged, which is a deconfinement of hadrons into individual quarks and gluons predicted by the theory of Quantum Chromodynamics (QCD) at high temperature. The `ALICE` experiment will therefore complement the current experiments at the Relativistic Heavy Ion Collider

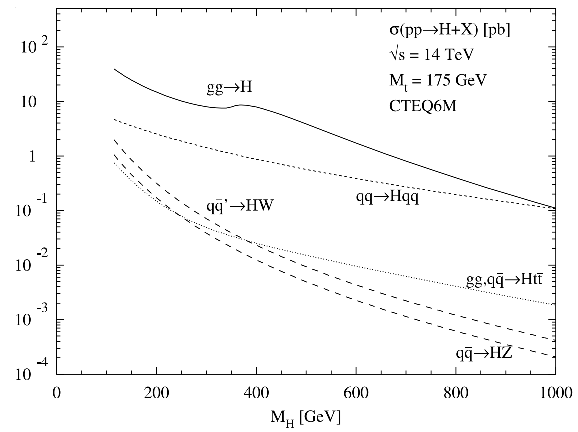


Figure 3.3: The production cross-sections of the SM Higgs boson as a function of its mass [87].

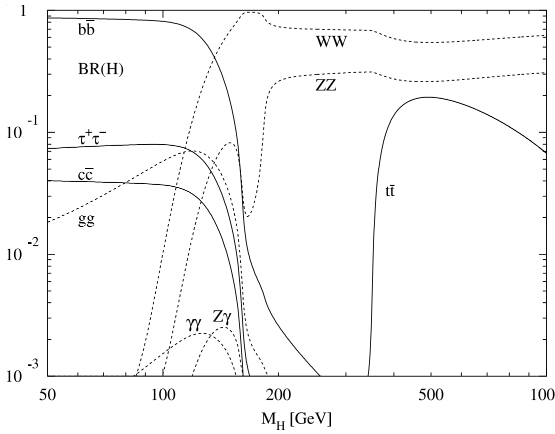


Figure 3.4: The branching ratio of the SM Higgs boson as a function of its mass [87].

(RHIC) located at the Brookhaven National Laboratory [86].

3.1.2 Design and operation of the LHC Collider

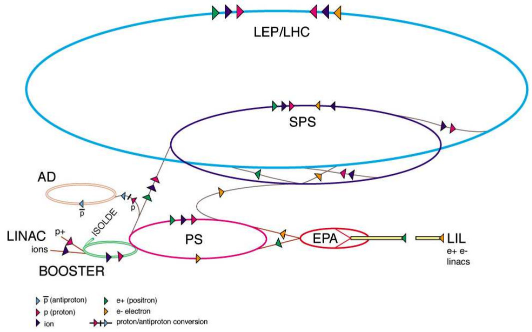
Before being injected into the LHC, proton beams will be prepared by CERN's existing 'accelerator complex', illustrated in Figure 3.5. Two linear accelerators or LINAC's will be used for the initial acceleration of protons and lead ions respectively. The Proton Synchrotron (PS) will be used to provide a low energy beam (25 GeV) with the final bunch structure. Next, the beams are pre-accelerated using the Super PS (SPS), and finally are injected into the LHC, where particles will be accelerated from 450 GeV to the nominal energy of 7 TeV (for proton beams).

Inside the LHC accelerator, the particles circulate in opposite directions in two separate beam pipes. The 1232 superconducting dipoles and more than 2500 other magnets will guide the LHC beams. They range from small, normally conducting bending magnets to large, superconducting focusing quadrupoles. These magnets allow to reduce the diameter of the beam to $\simeq 16\mu$m before the beam enters every experimental interaction point and consequently to achieve higher luminosities. A schematic cross-section of the LHC dipole is shown in Figure 3.6. The two beam pipes are surrounded by shells of superconducting coils creating the magnetic field which guides the beams to follow a circular path. The entire dipole rests inside a vessel filled with liquid Helium to enable operation at temperature below 2 K. In total, 40 000 cryogenic units will be placed around the 27 km circumference LHC ring.

The LHC is designed to ultimately reach a luminosity of $\mathcal{L} = 10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$, referred to as high luminosity, after a first few years of operation at the so-called low luminosity of $\mathcal{L} = 2\times 10^{33}\mathrm{cm}^{-2}\mathrm{s}^{-1}$. To accomplish such high luminosities, bunches with about 10^{11} protons are collided each 25 ns. As a consequence of the high luminosity, on average several interactions will happen during the same bunch crossing [1]. While for high lumi-

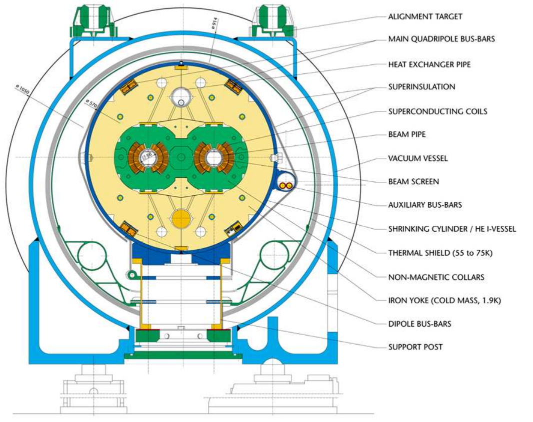
---

[1] At the LHC start-up the beams will contain fewer protons with possibly larger bunch spacing, enabling detector calibration/alignment and physics studies in a clean environment.



CENR AC _HF205_ V2/2/1998

Figure 3.5: The CERN accelerator complex.



CERN AC/DI/MM - HE107 - 30 04 1999

Figure 3.6: Schematic view of the cross-section of the LHC dipole.

nosity around 22 collisions per bunch crossing are expected to overlap[2] this number of 'pile-up' collisions decreases with a factor of five during the low luminosity. Additional to this 'in-time' pile-up collisions, also 'out-of-time' pile-up should be accounted for. These extra pile-up collisions can arise when the detector readout time is shorter than the bunch crossing frequency.

During the low luminosity phase approximately 10 fb^{-1} of integrated luminosity is expected to be collected each year, compared to 100 fb^{-1} per year at high luminosity. This first period will be important in thoroughly analysing and understanding the current knowledge of particle physics, and may also be sufficient for the discovery of new physics.

As the energy flux and collision rate aimed for at the LHC are much higher than in previous experiments, there are a number of strict requirements for the implementation of the detector layout and performance. These comprises:

- radiation hardness,
- fast readout electronics & small charge collection time,
- high detector granularity.

3.2 The CMS experiment

The Compact Muon Solenoid (CMS) is one of the two general-purpose detectors that almost finished construction at CERN. The main features of the CMS detector are

---

[2] For high luminosity, the collision rate per bunch crossing becomes:

$$\begin{aligned}
\#collisions &= \sigma^{inelastic} \times \mathcal{L} \times Bunch\,separation \times Fraction\,filled\,bunches \\
&= 70\,10^{-27} cm^2 \times 10^{34} cm^{-2} s^{-1} \times 25\,10^{-9} s \times 2808/3564 \\
&\simeq 22
\end{aligned}$$

summarised by its name. It is very compact compared to the other general purpose detector at the LHC, ATLAS, which has more than twice the volume, it has a good muon system assisted by an excellent central tracking detector, and the solenoidal magnet produces a magnetic field which is twice as strong as the one used by ATLAS and which allows an accurate particle momentum measurement. Furthermore, the design of the CMS detector has focused on the hermicity and electromagnetic calorimeter (ECAL) performance. Detector hermicity is crucial to be fully sensitive to manifestations of new physics through the produced non-interacting high energetic particles such as neutrinos. An outstanding ECAL energy resolution is required for a possible detection of the Higgs boson in the two photon decay channel.

In Figure 3.7 an illustration of the CMS detector is shown. As most of the high energy detectors, it has a multi-layered cylindrical barrel structure enclosing the beam pipe and endcaps to complete the hermetic coverage. Starting at the beam interaction point, the constituent compartments are the silicon tracker, the electromagnetic calorimeter, the hadronic calorimeter (HCAL) and the superconducting solenoidal magnet, followed by layers of muon chambers supported by the iron return yoke of the magnetic field lines. In the following paragraphs each of these sub-systems is introduced.

The coordinate convention is such that the z-direction is parallel to the beam line, the y-direction is vertical, and the x-direction is horizontal and points to the centre of the LHC ring. The centre of the detector is taken the origin of the coordinate system. Additionally, the azimuthal angle, ϕ, is measured around the beam axis with respect to the x-axis, and the polar angle, θ, is the angle of inclination from the beam axis. Often however, the polar angle is expressed in terms of the pseudorapidity, η, defined as:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \tag{3.1}$$

3.2.1 The Silicon Tracker

The CMS silicon tracker [88] has been designed to be able to cope with the high particle flux expected in the LHC proton collisions. In order to ensure an effective pattern recognition, it has been designed such that the occupancy of the silicon sensor channels is small, ranging from 10^{-4} for pixel sensors to the percent level for strip sensors. Hence, since the energy flux decreases with the distance to the interaction point, also the radial spacing between the layers increases. In order to achieve such a small hit occupancy, the CMS tracker consists of 20 000 silicon sensors which have altogether a surface of $210\,\mathrm{m}^2$. It is the largest silicon tracker ever build, having a diameter of 2.4 m and a length of 5.4 m. The tracking system is subdivided into five main parts: the Tracker Outer Barrel (TOB), the Tracker Inner Barrel (TIB), the Tracker Inner Disks (TID), the Tracker End Caps (TEC) and the pixel detector, illustrated in Figure 3.8. The barrel-endcap transition occurs at $0.9 < |\eta| < 1.4$, and the entire tracker provides rapidity coverage up to $|\eta| \simeq 2.5$.

The pixel detector must be situated as close to the interaction region as possible in order to identify relatively long-lived particles such as bottom and charm hadrons and τ's. Therefore the three pixel layers are positioned at a radial distance of 4.4, 7.3 and 10.2 cm from the beam axis. The layers are composed of ladders, which include 8

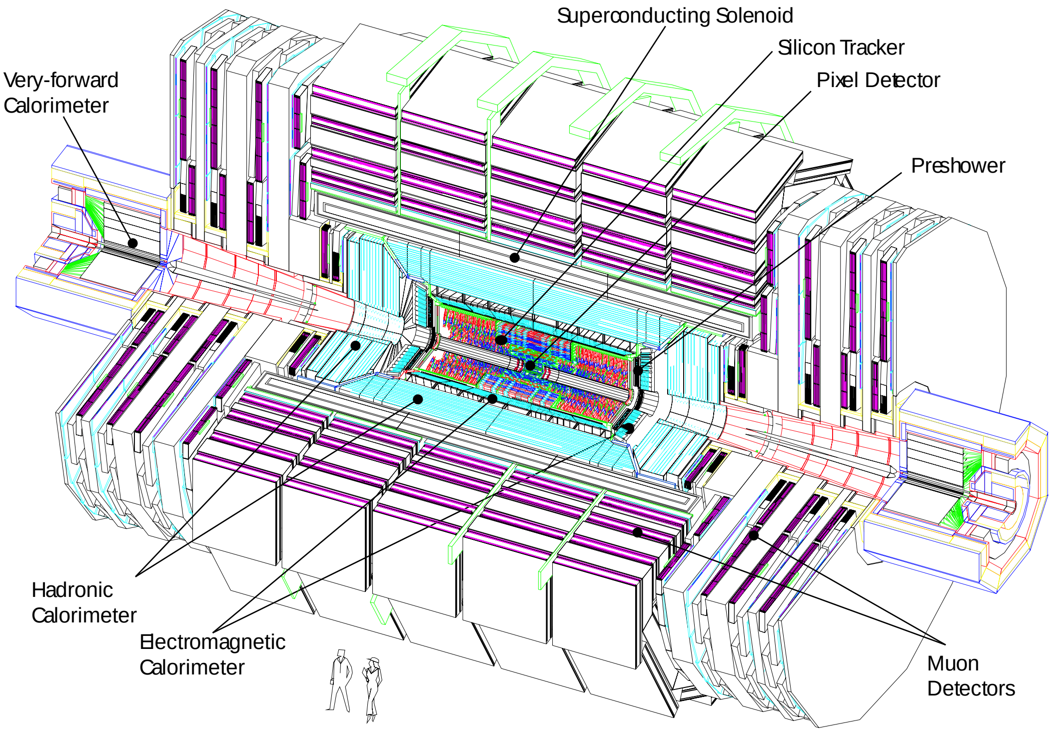


Figure 3.7: The Compact Muon Solenoid or CMS detector layout.

pixel sensors each. The pixel endcaps consist of two end disks, which extend from 6 to 15 cm in radius and are placed on each side at $|z| = 34.5$ cm and 46.5 cm. Each disc blade holds 24 pixel modules. In order to optimise the resolution of the pixel detector in both the $r\phi$ and z-directions, each module in the pixel end disks is rotated to take advantage of the electron Lorentz drift angle in the sensor layer due to the 4 T magnetic field. The hit resolution of the pixel detector is anticipated to be approximately $10\,\mu$m in $r\phi$ and 15-20 μm in z.

The barrel part of the strip detector consists of four layers of TIB modules and six layers of TOB modules. The modules are equipped with chips to read out energy deposition from the strips. Depending on the radii, the rectangular barrel modules have different sizes, and contain different numbers of strips, in order to keep the occupancy low, while sustaining the separation of strips to a 100-200 μm level to ensure a good hit position resolution. The inner two layers of TIB and TOB modules are made double-sided with two back-to-back sensors at a relative angle of 100 mrad, providing high precision two dimensional hit detection.

Due to its multi-layered complex structure, the amount of material a particle traverses and the fraction of energy it may lose until it exits the tracker varies with position. One of the aims in the tracker design is to keep this so-called material budget minimal. In the case of the CMS silicon tracker the radiation length of the material crossed by a particle is very dependent of the particle's direction. For the central region of the detector the radiation length is about $0.4\,X_0$, but this number increases rapidly when moving to forward regions, as can be seen in Figure 3.9. A maximum of $1.4\,X_0$

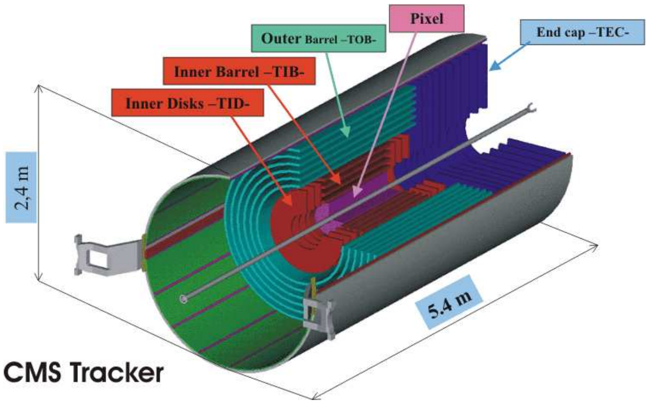


Figure 3.8: The different sub-systems of the CMS silicon tracker.

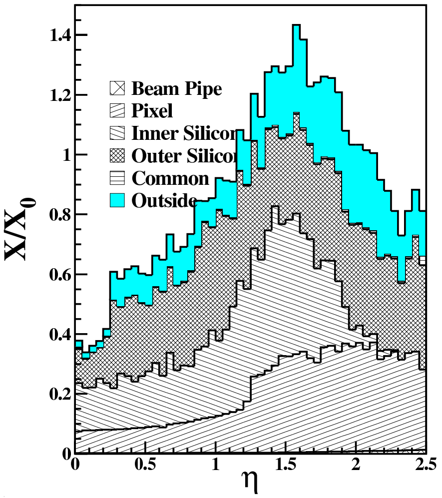


Figure 3.9: Tracker material budget as a function of the pseudorapidity in units of radiation length.

is found for the barrel-endcap transition region. This rather large amount of material in the tracker leads to significant energy loss, multiple scattering for electrons, and to photon conversions.

3.2.2 The Electromagnetic Calorimeter

Because the ECAL [88] plays an important role in reconstructing the di-photon decay of the Standard Model Higgs boson as well as in identifying electrons from vector boson and τ-lepton related channels, CMS has chosen for a very compact homogeneous scintillating crystal calorimeter designed for precision measurements of electron and photon energies and directions. The ECAL barrel is positioned just outside the tracking system at a radius of only 120 cm from the interaction point. Consequently, to distinguish energy depositions from different sources and to achieve high position resolution a material is needed that can provide small lateral and longitudinal spread of the shower.

Due to its high density, small Molière radius, fast response and radiation hardness, lead tungstate ($PbWO_4$) crystals were selected as the scintillation material to realise a very compact, high performance ECAL (e.g. with an energy resolution of 0.5% for 50 GeV particle). The light produced in the crystals is gathered with silicon avalanche photo-diodes. Around 80% of the light is emitted in the first 25 ns.

Similar to the other sub-detectors, also the ECAL consists of a barrel and two endcap substructures, as illustrated in Figure 3.10. The ECAL barrel, which is symmetric around η =0 contains a total of 61200 $PbWO_4$ crystals with each a surface of 22 mm×22 mm, which is similar to the Molière radius of the scintillation material. This surface translates in a 1° or 0.0174 rad coverage in η and ϕ, while its depth of 23 cm corresponds to a radiation length of 25.8 X_0. All the barrel crystals approximately point toward the interaction point with an offset of 3° in order to minimise the energy

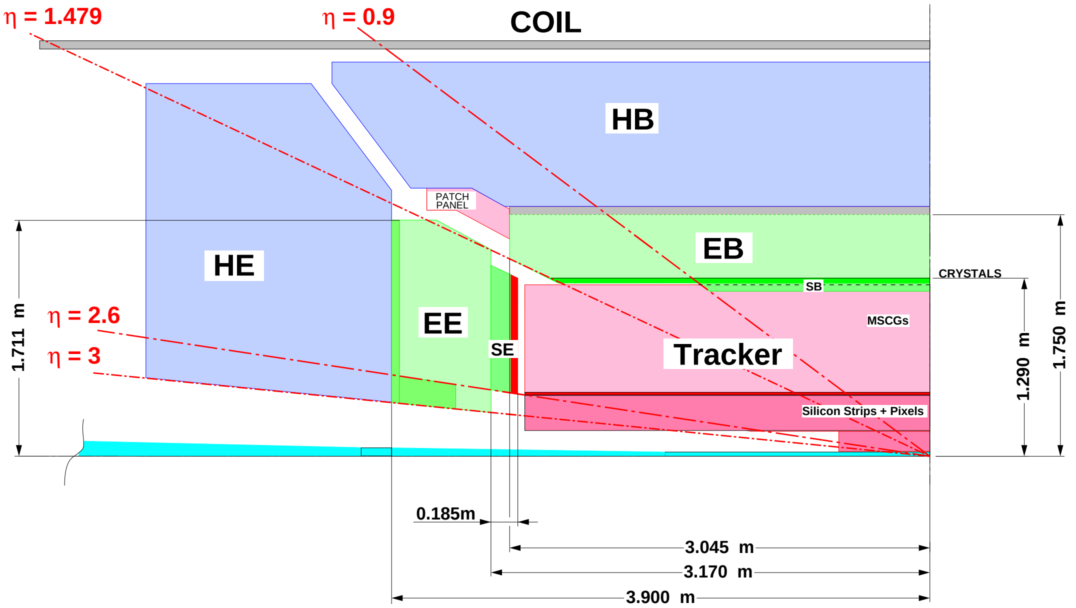


Figure 3.10: An rough overview of the calorimeter layout inside the magnet. Outside of the solenoid, the HO and HF respectively improve the number of radiation lengths and the acceptance of the calorimeter system.

loss if a particle transverses exactly between two crystals.

The endcap crystals have different dimensions from those used in the barrel, having a larger front face of size 29 mm×29 mm and a shorter length of 22 cm corresponding to $\sim 25X_0$. Additionally, pre-shower detectors are inserted between the tracker and the ECAL endcaps. The role of the pre-shower is to provide position measurement of the electromagnetic shower to high accuracy and is used to discriminate e.g. photons produced in a Higgs boson decay from photons produced in $\pi^0 \rightarrow \gamma\gamma$. The pre-shower is located just in front of the ECAL endcap, where thin lead radiators are used to initiate the shower (providing 3 X_0 in total). After each radiator layer, silicon strip sensors are placed to measure the hit position of the shower.

3.2.3 The Hadronic Calorimeter

Surrounding the ECAL is the HCAL [88] that is responsible for energy measurements of hadrons and their products with an energy resolution (when combined with the ECAL) of $\sigma_E/E = 120\%/\sqrt{E} \oplus 6.9\%$, where E is measured in GeV. Hadronic calorimeters rely on nuclear interactions which result in both hadronic and electromagnetic showers. As the probability of a nuclear interaction is small but the energy deposited in the calorimeter is large, there are significant fluctuations in the measured energy in hadronic calorimeter showers, descreasing the overall precision of the detector.

The CMS hadron calorimeter design was strongly influenced by the decision to place the calorimeter inside the solenoid, leading to little space for the detector. The HCAL consists of three main parts: the Hadron Barrel (HB) and the Hadron Endcap (HE), which extend to $|\eta| = 3$, and the Hadron Forward (HF), which are located around the beam pipe outside the muon system at $|z| = 10.9\,\mathrm{m}$, and to complete the coverage up to $|\eta|$=5.3 . The large pseudorapidity coverage ensures to collect most transverse energy in order to determine with a relative high precision the missing transverse energy, which will be crucial for the search of many topologies of physics beyond the Standard Model.

Both the HB and HE are built in a classical sandwich like sampling calorimeter structure with brass as absorber material and plastic scintillator with wavelength-shifting fibres. Scintillators behind the coil of the magnet built the Hadron Outer (HO) detectors and effectively increase the calorimeter thickness to 10 interaction lengths. The scintillators are sub-divided into tiles of size $\Delta\eta \times \Delta\phi$ =0.087×0.087 for $|\eta| <2$ and larger divisions at high pseudorapidity. These are to be used as the basic building blocks in the reconstruction of the hadronic showers.

For the construction of the Hadron Forward (HF) steel is selected as absorber. The collection of the energy is accomplished using quartz fibre emitting Cerenkov light, a technology which is preferred to plastic because of its greater radiation tolerance. Additionally, relatively faster phototubes are selected in the HF compartment instead of hybrid photodiodes, that collect the light in the HB and HE.

3.2.4 The Muon Spectrometer

The muon system is a tracking device in the outermost region of the CMS detector. Only muons and non-interacting particles such as neutrinos manage to pass through the calorimeters without depositing a large fraction of their energy. An important challenge of the muon system is to provide fast recognition and efficient reconstruction, which are required for triggering purposes, as will be further elaborated upon in Section 3.2.6. There are four layers of muon stations in the barrel and the endcaps interleaved with the iron return yokes, like illustrated in Figure 3.11. Each of the stations is meant to provide track segments reconstructed from a few distributed hits. These will later be combined with information from the inner silicon tracking system to form the complete muon tracks.

Because of the varying muon and neutron radiation flux and the inhomogeneous magnetic field in forward detector regions, the decision for different technologies for barrel and endcap region is taken. This led to the choice of Drift Tubes (DT) in the barrel stations and Cathode Strip Chambers (CSC) in the endcap stations. Both types of detectors are assisted by Resistive Plate Chambers (RPC), as shown in Figure 3.11. Each of the stations contain respectively 12 DT layers for the barrel and 6 CSC layers in the endcap. Within a single station, the DT layers provide a final spacial resolution of 100 μm and the CSC's provide 80-450 μm depending on the location of the station. The DT and CSC detectors are used to obtain a precise position measurement, while the RPC's are, due to their very fast response and time resolution of the order of 1 ns, dedicated for the trigger purpose. The RPC plates in the endcaps extend to $|\eta| < 1.6$ and the reconstruction of muon tracks in the region outside is carried out using the CSC's alone.

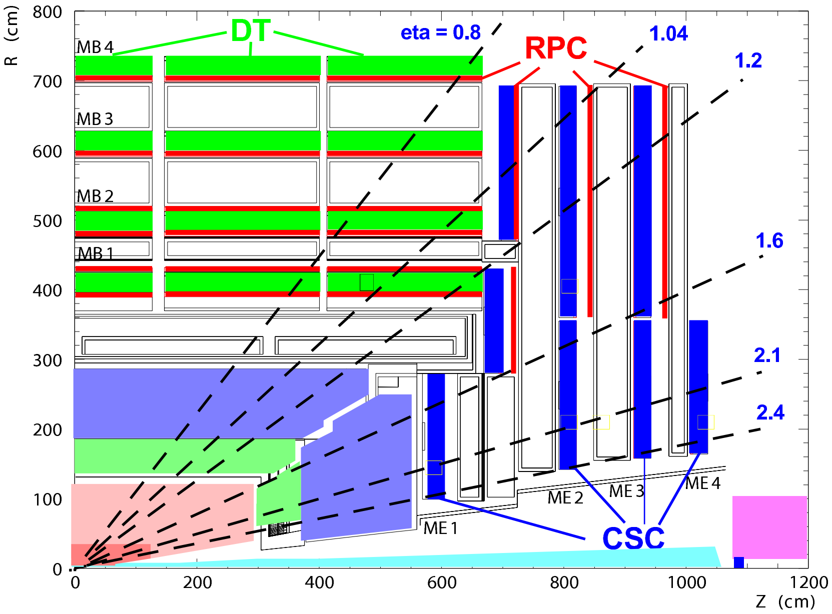


Figure 3.11: A longitudinal view of the muon system indication the location of the three detector types contributing to the muon spectrometer.

3.2.5 Alignment and calibration of the CMS detector

At the beginning of CMS data taking, the alignment of the tracking system (silicon tracker and muon spectrometer) can be determined only from survey measurements during construction, the Laser Alignment System and cosmic muon tracks [89, 90]. Because of this, the tracking system is subject to substantial uncertainties, due to an imprecise mounting and movement of detector components after cooling and commissioning of the magnetic field. Misalignment will affect the measurements and associated uncertainties of hits, tracks and all reconstructed objects depending on them. With the accumulation of collision data, track-based alignment of the tracking system will become applicable and reduce the alignment uncertainties. We will focus on the precision that is claimed to be achievable once $100\,\mathrm{pb}^{-1}$ of data is collected [91], as this accuracy will be relevant in the study of the systematic uncertainties on the physics estimators introduced in the next chapters.

With $100\,\mathrm{pb}^{-1}$ of data, not only minimum bias events and low mass resonances such as J/Ψ will be available for the alignment of the silicon tracker, also high mass resonance decays from W and Z bosons will become a valuable tool to help in the extraction of the alignment constants. The precision expected for the pixel detector is of the order of 20 μm, while for the strip tracker 30-50 μm should be reachable. Additionally the whole muon system should be positioned with respect to the inner tracker with a precision of 1 mm and 250 μrad. For the muon stations and the interleaving chambers on their turn an alignment precision of respectively 1 mm and 200 μm is foreseen.

Apart from the alignment of the tracking device, also the calibration constants employed in the various CMS sub-detector are determined only up to a certain level of accuracy. Therefore, as part of the CSA07 exercise [92], Monte-Carlo simulated event samples were produced in which, in addition to the described misalignment shifts, several of the calibration constants were varied in their uncertainties [93]. Examples are the intercalibration precision of the ECAL crystals (1.4-4%, depending on their location in η) and the gains for the HCAL (which are smeared by 5% in the 100 pb^{-1} scenario).

3.2.6 Online Selection and Data Acquisition

General goal and strategy

The LHC bunch crossing frequency for proton-proton interactions is 40 MHz, what results in a total event rate of the order of 10^9 Hz. Since the raw event size is of the order of 1 MB and storing and processing the resulting amount of data would exceed the current data-storing and process-time limits, a large fraction of the total event rate needs to be discarded by means of an online event selection system. The goal of the CMS trigger system is to reduce the event rate to the order of 100 Hz, the maximum rate that can be archived by the online computer farm. Hence, typically only one out of 10^7 events can be stored, which fortunately leaves enough freedom to accommodate all signal channels of interest at the LHC.

Another aspect of the online selection is its speed to process the events. Since the input rate is 40 MHz, a decision must be taken every 25 ns. However, this time is too small even to read out all raw data from the detector. The accept/reject decision is therefore taken in several steps (levels) of increasing refinement, where each level takes a decision using only part of the available data. The CMS trigger will consist of only two physical levels. The Level-1 (L1) trigger [94] is implemented on dedicated hardware and has only access to data from the calorimeters and the muon detectors with coarse granularity. Hence, information of the central tracking system is not yet used in the L1 trigger decision. On the basis of this limited information it has to reduce the input rate up to a level acceptable for the Data Acquisition system (DAQ). At LHC startup, the DAQ system will be able to handle an event rate of up to 50 kHz[3], which will be increased to 100 kHz when the full LHC design luminosity is reached. Once accepted, the L1 triggered objects are passed to the High Level Trigger (HLT) [95], which will be implemented in software running on a single farm of commercial processors. A software-based HLT allows full flexibility of the algorithms and offers the possibility to benefit from future improvements in the offline reconstruction software and in computing technologies.

The online event selection at the LHC is generally based on the fact that most proton-proton collisions produce soft hadrons with transverse momenta (p_T) of few GeV/c, or so-called 'minimum bias' events, while signal events (e.g. decays of heavy objects like the Higgs boson and the top quark) typically contain high-p_T leptons, jets or missing transverse energy in the final state. Hence, the trigger decisions are made based upon the identification of the following physics objects:

- electrons/photons

---

[3] In fact, to be robust against statistical fluctuations and simulation uncertainties, a safety margin of a factor of three has been included to ensure that the total acceptance rate does not exceed 16 kHz.

- jets
- muons
- missing transverse energy

where each object is reconstructed in a specific detector sub-system or with combined information.

In the next two paragraphs, both trigger levels will be described, with the focus on the filtering of events which contain an isolated single muon in the final state. Because all analyses and results presented in this work are based on the study of semi-muonic decaying $\mathrm{t\bar{t}}$-events, $\mathrm{t\bar{t}} \rightarrow q\,\bar{q}\,b\,\mu\bar{\nu_\mu}\bar{b}$, this is the only trigger stream that will be applied.

The Level-1 Trigger

The Level-1 trigger is implemented on custom programmable hardware. In order to run dead-time free, and consequently to take an accept/reject decision at a rate equal to the bunch crossing frequency, a synchronous pipelined structure of processing elements is required, each taking less than 25 ns to complete. The complete detector data are stored in pipeline memories, whose depth is limited to 128 bunch crossings. The total Level-1 latency is therefore fixed to $128{\times}25\,\mathrm{ns}{=}3.2\,\mu\mathrm{s}$. This time includes the transmission time between the detector and the counting room and, in the case of Drift Tube detectors, the drift time. The time available for calculations can therefore be as low as $1\,\mu$s.

The Level-1 trigger is divided into three subsystems: the calorimeter trigger, the muon trigger and the global trigger. The muon trigger is further subdivided into three independent systems for the DT, CSC and RPC detectors, respectively. A schematic view of the components of the Level-1 trigger system and of their relationships is shown in Figure 3.12.

The calorimeter and muon triggers do not perform any selection themselves. They only identify trigger objects of different types, such as muons with an isolation flag. The four best candidates of each type are selected and sent to the global trigger, together with the measurement of their position, direction, transverse energy or momentum and a quality word. Furthermore, the global trigger receives the total and missing transverse energy measurement and counters of the number of jets above programmable thresholds from the calorimeter trigger. It is this global trigger that actually selects the interesting events according to programmable trigger conditions, which can include requirements on the presence of several different objects with energies or momenta above predefined thresholds. Up to 128 of these trigger conditions can be tested in parallel, and each can be pre-scaled to accept only a fraction of selected events.

The High Level Trigger

The role of the HLT is to further reduce the data rate to ~100 Hz using more detailed information and sophisticated reconstruction algorithms, and to make the final selection of events to be stored for offline analyses. The HLT is implemented as software running on a single farm of commercial processors, each processing one full event at a time. The

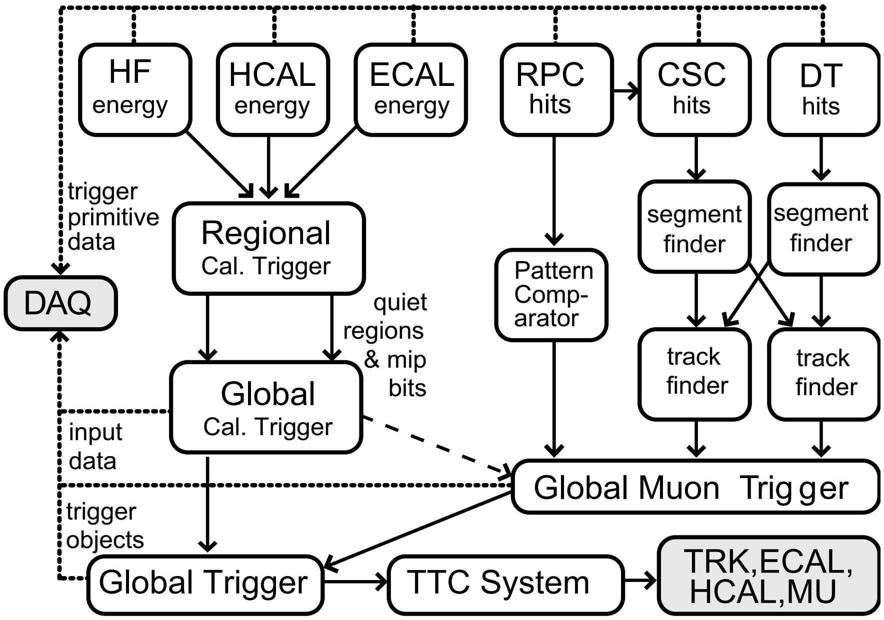


Figure 3.12: Structure of the Level-1 trigger system.

advantages of this choice are multiple: a fully programmable system allows complete flexibility of the algorithms, only limited by the maximum available CPU time and data bandwidth. Also, the use of standard commercial components leads to significant economies in the costs of hardware, development and maintenance. Because of the choice to implement this HLT software analogue to the offline reconstruction software in C++, the same algorithms and techniques can be used.

The HLT is carried out on one event per processor basis, hence in order to minimise the processing time background events should be discarded as soon as possible. It is therefore useful to organise the selection in a chain of logical steps that consist of progressively more sophisticated and CPU-time consuming algorithms. Selection criteria are applied at the end of each step, in order to reduce the event rate to a level acceptable for the following one. It is customary to identify two steps, the Level-2, where the data of the calorimeters and muon detectors are used, and the Level-3, which also utilises the full information from the inner tracker system, for which reconstruction is more time consuming. Additional rate rejection is obtained at both levels by the use of isolation algorithms.

The muon selection in the HLT proceeds in three steps. First, muons are reconstructed in the muon chambers as described in Section 4.5, starting from the Level-1 trigger objects, resulting in a refined p_T-measurement. Then the muon track candidates are propagated to the central silicon tracker, another time improving the uncertainties on the track parameters. Finally, HLT isolation algorithms are used to discriminate events on the basis of the energy deposits in the calorimeters and the transverse momentum of tracks in a cone around the direction of the muon. This isolation requirement is added to separate muons from b and c decays from those from heavy object decays like $W \to \mu\nu_\mu$.

Calorimeter isolation is based on the measurement of the transverse energy in the

trigger level	single muon p_T-threshold	event rate
L1	14 GeV/c	2.5±0.2 kHz
HLT	19 GeV/c	25.8±0.8 Hz

Table 3.1: Low luminosity trigger thresholds and expected rates for the single muon stream relevant for this work.

towers of the hadronic and electromagnetic calorimeters. This algorithm is very fast and can be applied already after the Level-2 selection. However, since it is based on the calorimeters, this algorithm is rather sensitive to the influence of pile-up collisions, particularly important at high luminosity.

Track isolation algorithms are based on the measurement of the sum of the transverse momenta of tracks reconstructed in the central tracker around the muon. Tracks are reconstructed using regional reconstruction in the silicon tracker by defining a tracking region corresponding to the isolation cone[4]. Only tracks originating from the same collision vertex as the muon are used in the sum, so that the algorithm is less sensitive to pile-up collisions than calorimeter isolation.

The isolated single-μ trigger thresholds

Although previously we motivated that the CMS trigger implementation still leaves much room for new trigger object definitions or improved reconstruction algorithms, all trigger streams have been appointed a maximum event rate in order to achieve the infrastructural limits. This exercise resulted in the thresholds on the reconstructed transverse momentum of the isolated muon candidate as given in Table 3.1 for low luminosity conditions. These are the thresholds as applied to all reconstructed and selected semi-muonic $t\bar{t}$-events used as an input to the analyses presented in this work.

In the meantime however, improved trigger studies have been performed resulting in updated trigger tables [96]. Important is the creation of several L1 single muon streams with each their own p_T-threshold, often accompanied with a proper pre-scale factor to limit the total bandwidth. For the HLT two streams are now defined: a 'Single Isolated Muon' stream with a threshold at only 11 GeV/c and a 'Single Relaxed Muon' stream with a threshold at 16 GeV/c. This relaxation of the trigger requirements will be crucial in order to help in the understanding of the detector response and to study for example QCD events over a wide energy range.

3.2.7 The CMS software environment

CMSSW and the Event Data Model

At the end of 2004 CMS decided to redesign its offline software framework. The new software now includes a completely revisited event data model and is fully integrated with a database infrastructure for handling calibration and alignment data. The primary goal of this decision is to facilitate the development and deployment of reconstruc-

---

[4] For a description of the track reconstruction, we refer to Section 4.4.1.

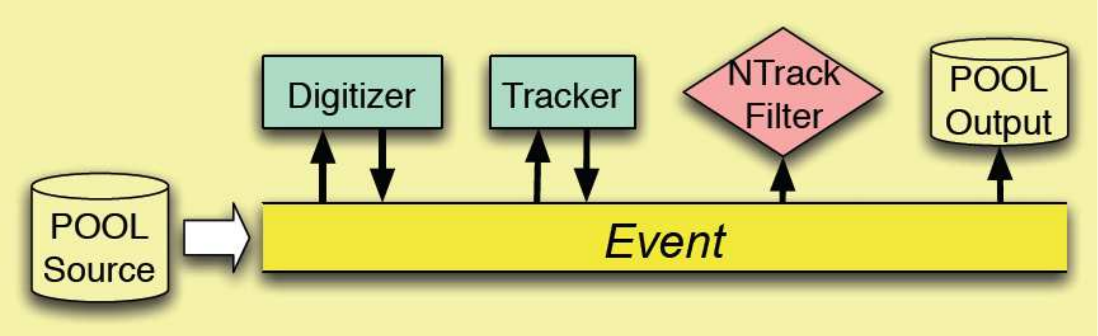


Figure 3.13: Illustration of the EDM 'Event' concept.

tion and analysis software and to make the software ready and flexible for data-taking. In the meantime, this 'CMSSW' software framework has had its first severe and successful test in an experimental environment during the Magnet Test Cosmic Challenge which took place the first months of 2006.

In contrast to the more traditional approach to have several executables for the different tasks that can be isolated while processing an event, only one executable, 'cmsRun', is known in the CMSSW framework. This executable is able to run both on Monte-Carlo simulated and detector data.

The whole Event Data Model (EDM) [88] is based on the concept that all data processing should pass through a single data structure called the 'Event'. This is illustrated in Figure 3.13. It is the user who defines in a job-specific configuration file which data to start from, what tasks to perform and what information finally needs to be stored in a ROOT-file. For each task in the processing of the data (e.g. digitisation of the drift tube hits), a framework module needs to be implemented. The module defines which information is needed from the Event to execute its module-specific task. Depending on the type of the module, the output of the task is added to the Event ('EDProducer'), the event is selected or discarded for further analysis ('EDFilter') or the output is only used to produce e.g. histograms ('EDAnalyser'). For every module, parameters can be implemented, of which the values are passed to the executable through the configuration file.

To be able to fully process an event, one has to take into account potentially changing and periodically updated information about the detector environment and status (e.g. calibrations, alignments, geometry descriptions, ...). This information is not tied to a given event, but rather to the time period for which it is valid. This 'Event Setup' is also passed to the cmsRun executable via the CMSSW configuration file.

Apart from the reconstructed data added to the Event after each executed producer module, the event also keeps track of the configuration of the software used for the reconstruction of each contained data object as well as the conditions and calibration data employed during the reconstruction. This provenance information, which is stored as part of the output ROOT-file, guarantees the reproducibility of any given physics result, and allows to trace back and browse through the module configurations and event setup at any later time.

Simulation of the Detector Response

In order to understand and design the flow of electrical pulses as will be detected by CMS due to the interaction of the particles with the detection material, adequate Monte-Carlo simulations are required. An exact simulation of the trajectory of a particle can only be delivered if the position and chemical composition of every little detector compound is archived in a database and a map of the magnetic field is available. Starting from this information, a dedicated software package known as GEANT-4 [97] is used to predict the trajectory of every single Monte-Carlo simulated particle through the detector material. Next, the hits in the various detection systems are determined and digitised, resulting in a data-stream as is expected from the real detector with real data. Test beam data is shown very useful in understanding and optimising this detector simulation.

One of the important disadvantages of the GEANT based detector response simulation is the fact that all necessary computations are highly CPU intensive. However, in view of doing physics analyses, develop and tune reconstruction algorithms, design detector upgrades or test the influence of systematics in a given physics analysis, a much faster alternative is needed with still a high accuracy in the detector effect description. To achieve this, a number of simplifying assumptions were made, a number of dedicated parameterisations were used, and some optimised reconstruction algorithms were developed (e.g. for tracking)[5]. This exercise resulted in a fast simulation program called FastSim, which is no longer depending on the GEANT software. Additionally, the output of the fast simulation is based on the same data formats as the output provided by the complete GEANT reconstruction software. As a consequence, the same modules for higher level reconstruction tasks or physics analyses can be applied on ROOT-files produced with full and fast detector simulation.

The Top Quark Analysis Framework

Many analyses are to be performed on the same top quark events (single top or top quark pairs) and are often based on the same reconstruction tools and final states. To avoid the duplication of work and rationalise the task-sharing in the group, we started the development of the ‘Top Quark Analysis Framework (TQAF)’ in Spring 2007 [98, 99]. This analysis code is structured in three layers, and is illustrated in Figure 3.14:

- **Layer-1:** reconstruction of general-purpose TopObjects starting from the central CMS reconstructed data objects (final state and analysis independent);
- **Layer-2:** construction of event ‘solutions’ from these TopObjects (final state dependent but analysis independent);
- **Layer-3:** the actual analyses to perform the actual physics measurements (final state and analysis dependent).

---

[5] For more information on the difference between full and fast simulation, and a comparison of their performance, we refer to Section 2.6 in [88].

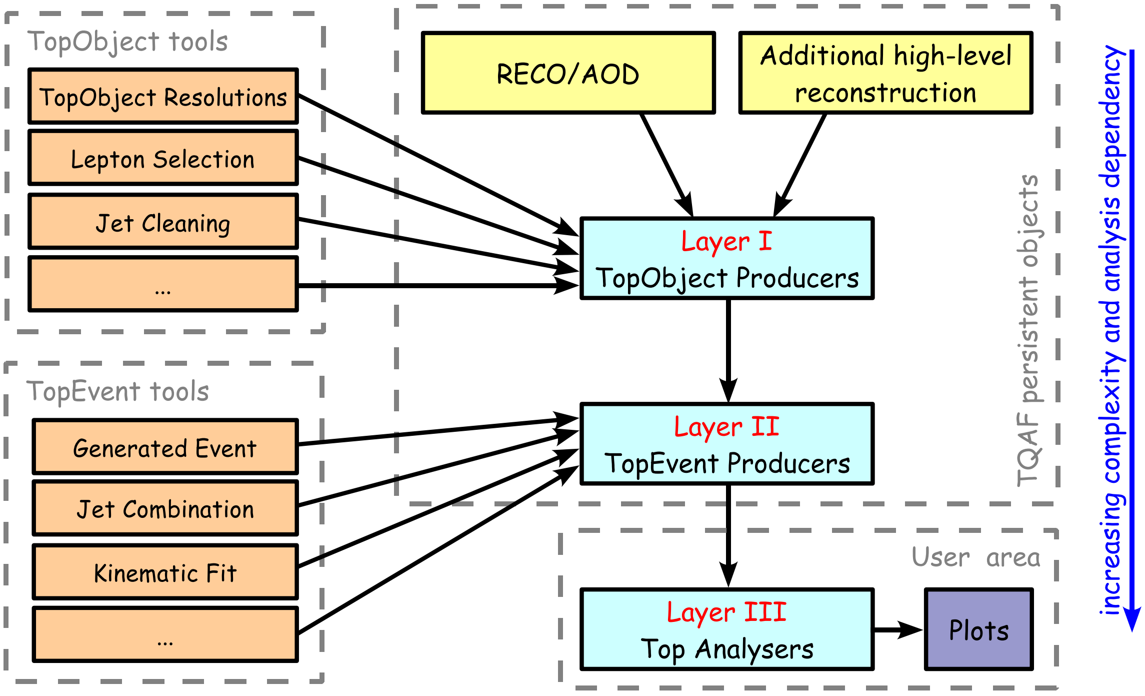


Figure 3.14: Schematic representation of the TQAF layered design.

This structure will allow the use of the same framework modules in a whole range of physics analyses. All analyses presented in this work are implemented in this TQAF.

3.2.8 The CMS Computing Model

The new generation of collider experiments presents challenges not only in terms of the physics to discover and the detector to build and operate, but also in terms of the data volume and the necessary computing resources. Data sets and resource requirements are at least an order of magnitude larger than in previous experiments. For both technical and funding reasons, it was decided to connect computing services and resources available in the CMS groups all around the world via the GRID computing network and to construct the CMS computing environment as a fully distributed system. All the computing, storage and connectivity resources will be merged in one infrastructure managing the data processing, data archiving and Monte Carlo event generation. The system will be available to all CMS collaborators, independently of their physical locations.

The CMS computing model [100] defines a hierarchy of the computing centres available to CMS. Three levels or 'Tier's' are defined and interconnected:

- **the Tier-0 Centre at CERN**
  All data flowing out of the CMS Online Trigger and Data Acquisition System (TriDAS) will be both archived on tape and distributed to the Tier-1 centres in order to have two copies of all the 'RAW' data detected in the experiment. In LHC down-times the centre will contribute to the reprocessing of the data.

- **the Tier-1 Centres**
  Currently eight centres are each responsible for the safe keeping of a share of the (second copy of the) RAW and reconstructed data. These centres should have enough CPU power for large-scale reprocessing steps. Their data-product are then distribution to Tier-2 centres, while shares of the simulated data produced in the Tier-2's are stored in the Tier-1 centres too.

- **the Tier-2 Centres**
  A more numerous set (∼49) of smaller Tier-2 centres, but with substantial CPU resources, provide capacity for user analyses, calibration studies, and Monte Carlo production.

Hence, the input ROOT-files for a physics analysis are to be found distributed over the Tier-2 sites. In order to run on the data, jobs should be created, consisting mainly of a CMSSW configuration file, and send to the sites where the data is located. On a fair share basis, the job is offered a computer node at the given Tier-2 where the job is executed. The resulting output ROOT-file is then sent back to a storage element selected by the user and can be used for the production of the final physics results. This procedure was followed to access the simulated top quark event data used as an input to the analyses presented in this work.

3.2.9 Overview of the produced Monte-Carlo samples

A list of all centrally produced data-sets accessed for this thesis, with their statistics, corresponding process cross-section and integrated luminosity is given in Table 3.2. All these samples are produced using the full GEANT-4 detector simulation and without the inclusion of pile-up collisions. Apart from this list, fast simulation samples were employed in the estimation of the systematic effects on the various top quark estimators presented in the upcoming chapters, and for a check of the linearity of the top quark mass estimators. All these additional samples are listed in Table 3.3.

Sample	generator	σ^{LO} (pb)	size sample	$\mathcal{L}$ (fb^{-1})
$\mathrm{t\bar{t}}$ (Spring07 production)				
$\mathrm{t\bar{t}}$0j excl.	`AlpGen`	190	492 k	2.59
$\mathrm{t\bar{t}}$1j excl.	`AlpGen`	170	434 k	2.55
$\mathrm{t\bar{t}}$2j excl.	`AlpGen`	100	257 k	2.57
$\mathrm{t\bar{t}}$3j excl.	`AlpGen`	40	96 k	2.40
$\mathrm{t\bar{t}}$4j incl.	`AlpGen`	61	98 k	1.60
$W(bb)$ (Spring07 production)				
W4j excl.	`AlpGen`	174	73 k	0.42
W5j excl.	`AlpGen`	45	55 k	1.22
W6j incl.	`AlpGen`	31	20 k	0.65
Wbb2j excl.	`AlpGen`	17	148 k	8.71
Wbb3j excl.	`AlpGen`	7	23 k	3.28
QCD di-jet($\hat{p_T}$ bins)				
0-15 GeV/c	`PYTHIA`	$5.52\,10^{10}$	54 k	$9.81\,10^{-10}$
15-20 GeV/c	`PYTHIA`	$1.46\,10^{9}$	54 k	$3.71\,10^{-8}$
20-30 GeV/c	`PYTHIA`	$6.32\,10^{8}$	92 k	$1.46\,10^{-7}$
30-50 GeV/c	`PYTHIA`	$1.63\,10^{8}$	188 k	$1.16\,10^{-6}$
50-80 GeV/c	`PYTHIA`	$2.16\,10^{7}$	119 k	$5.54\,10^{-6}$
80-120 GeV/c	`PYTHIA`	$3.08\,10^{6}$	40 k	$1.33\,10^{-5}$
120-170 GeV/c	`PYTHIA`	$4.94\,10^{5}$	27 k	$5.53\,10^{-5}$
170-230 GeV/c	`PYTHIA`	$1.01\,10^{5}$	25 k	$2.52\,10^{-4}$
230-300 GeV/c	`PYTHIA`	$2.45\,10^{4}$	53 k	$2.18\,10^{-3}$
300-380 GeV/c	`PYTHIA`	$6.24\,10^{3}$	50 k	$8.06\,10^{-3}$
380-470 GeV/c	`PYTHIA`	$1.78\,10^{3}$	29 k	$1.67\,10^{-2}$
470-600 GeV/c	`PYTHIA`	$6.83\,10^{2}$	82 k	$1.21\,10^{-1}$
600-800 GeV/c	`PYTHIA`	$2.04\,10^{2}$	3 k	$1.47\,10^{-2}$
800-1000 GeV/c	`PYTHIA`	$3.51\,10^{1}$	29 k	$8.55\,10^{-1}$

Table 3.2: Overview of the centrally produced full GEANT-4 simulated samples employed for the various physics analyses presented in this thesis.

Sample	generator	σ^{LO} (pb)	size sample	$\mathcal{L}$ (fb^{-1})
Nominal $\mathrm{t\bar{t}}$ (CSA07 production)				
$\mathrm{t\bar{t}}$0j excl.	`AlpGen`	335	1509 k	4.51
$\mathrm{t\bar{t}}$1j excl.	`AlpGen`	90	365 k	4.05
$\mathrm{t\bar{t}}$2j excl.	`AlpGen`	19	96 k	5.05
$\mathrm{t\bar{t}}$3j excl.	`AlpGen`	3.2	15 k	4.74
$\mathrm{t\bar{t}}$4j incl.	`AlpGen`	0.82	5 k	6.10
MA/MC $\mathrm{t\bar{t}}$ (CSA07 production)				
$\mathrm{t\bar{t}}$0j excl.	`AlpGen`	335	1525 k	4.57
$\mathrm{t\bar{t}}$1j excl.	`AlpGen`	90	353 k	3.92
$\mathrm{t\bar{t}}$2j excl.	`AlpGen`	19	88 k	4.69
$\mathrm{t\bar{t}}$3j excl.	`AlpGen`	3.2	15 k	4.74
$\mathrm{t\bar{t}}$4j incl.	`AlpGen`	0.82	5 k	6.10
Privately produced semi-μ $\mathrm{t\bar{t}}$ events				
Nominal	`PYTHIA`	72.9	200 k	2.74
Pile-Up collisions	`PYTHIA`	72.9	200 k	2.74
PDF CTEQ6M	`PYTHIA`	72.9	200 k	2.74
Underlying Event (+/-)	`PYTHIA`	72.9	200 k	2.74
Λ_{QCD} (+/-)	`PYTHIA`	72.9	200 k	2.74
${Q^2_{max}}^{ISR/FSR}$ (+/-)	`PYTHIA`	72.9	200 k	2.74
Light quark fragm. (+/-)	`PYTHIA`	72.9	200 k	2.74
Heavy quark fragm. (+/-)	`PYTHIA`	72.9	200 k	2.74
M_{top} =165 GeV/c^2	`PYTHIA`	72.9	100 k	1.37
M_{top} =170 GeV/c^2	`PYTHIA`	72.9	100 k	1.37
M_{top} =180 GeV/c^2	`PYTHIA`	72.9	100 k	1.37
M_{top} =185 GeV/c^2	`PYTHIA`	72.9	100 k	1.37

Table 3.3: Overview of the fast simulated samples employed in this work. 'MA/MC' denotes the 100 pb^{-1} misalignment/miscalibration scenario introduced in Section 3.2.5, while '(+/-)' indicates that for the given systematic effect a sample was produced for both the positive and negative parameter shift. For more details on the chosen `PYTHIA` parameters in the private productions we refer to Chapter 2 and the sections describing the systematic effects on the different estimators.

Chapter 4

Object Reconstruction and Selection

Following the introduction to the CMS detector, its different sub-detectors and the data handling in Chapter 3, we will now emphasise on the reconstruction of the physics objects starting from the data detected in the different detector subsystems. In this chapter only those techniques and algorithms are discussed to reconstruct the high level objects used as building blocks in the analyses presented in Chapter 5 till Chapter 7. Because only 'semi-muonic' decaying $\mathrm{t\bar{t}}$-events, $\mathrm{t\bar{t}} \rightarrow bW\bar{W}\bar{b} \rightarrow bj\bar{j}\mu\bar{\nu}_\mu\bar{b}$, are used as signal events in these analyses, the objects of particular interest are light and b-quark jets, missing transverse energy and muons. For details on the reconstruction of other objects like electrons, photons and τ's we refer to [88].

In a first section, the complex issue of reconstructing the initial parton's four-vector from a co-linear flow or jet of stable and detectable particles, the so-called jet reconstruction, is discussed. Also the different techniques and strategies to calibrate the reconstructed jet energy are introduced. Section 4.2 summarises the results obtained from a method we developed to compare the jet clustering performance of several jet algorithms and their respective configuration via some pre-defined quality markers. A third section describes the reconstruction of the missing transverse energy, while Section 4.4 will elaborate on track and vertex reconstruction, and the use of these to tag the flavour of the reconstructed jet. The reconstruction of muons in the CMS detector is addressed in Section 4.5. Finally, in Section 4.6, an overview is given of all selection cuts and their efficiencies that are applied on the reconstructed objects.

4.1 Jet Reconstruction

The robust and accurate reconstruction of the initial four-vector of quarks and gluons is one of the most challenging tasks in the analysis of the final state of hadron collisions. Because of the QCD confinement described in Section 1.1.1, colour charged particles cannot exist in a free form and will fragment into colour neutral hadrons before direct detection is possible. This hadronisation process results in a collimated flow of hadrons referred to as a jet. These reconstructed objects will need to be measured and studied in a particle detector in order to determine the properties of the original parton. For the jet reconstruction task, several algorithms are developed to cluster jet input objects, such as the energy deposits in the calorimeters. The definition of the jet input should however be optimised for effects like detector noise, the underlying event and pile-up collisions. Also the choice of jet clustering algorithm and the tuning of its parameters is crucial for the jet reconstruction performance. Depending on the clustering algorithm, a reconstructed jet can be more robust to the influence of the magnetic field, gluon radiation, an overlap in the hadronisation phase-space of the initial partons, … Furthermore, issues such as an imperfect theoretical description of the non-perturbative QCD in the event simulation, the intrinsic non-linear detector response and the production of invisible particles in the hadronisation process will strengthen the need for a well-studied, data driven jet calibration. A good knowledge of the jet energy response will not only provide a benchmark for understanding the detector and to offer a first window on possible new physics, but is also crucial for the precision achievable in many physics analyses.

In this section, we will start with a description of the jet input objects and the jet clustering algorithms. In paragraph 4.1.3 and 4.1.4 we focus respectively on the main jet calibration techniques already studied in the CMS collaboration, and the jet resolutions in semi-leptonic $\mathrm{t\bar{t}}$-events.

4.1.1 Jet Input

Jet reconstruction methods cluster objects in the observed final state with a pre-defined algorithm. These objects could be either tracks, energy deposits in the calorimeters or other reconstructed objects parametrised as four-vectors. The optimal choice of input objects will depend on the understanding of the performance of the CMS detector. While at start-up only calorimeter or track information may be used, the combination of both sub-detectors' information into particle flow objects is the ultimate goal[1]. Because these objects are still under development [101], calorimeter towers are used as input for the jets in the analyses presented in this thesis.

EcalPlusHcalTowers

The most commonly used jet input at hadron colliders are the calorimeter towers, that merge the information of one or more hadronic calorimeter cells with several much finer electromagnetic calorimeter cells covering the same (η,ϕ) space regarded from the origin of the detector. For the CMS detector, a total of 4320 such towers are distributed over the (η, ϕ)-space with a granularity plotted in Figure 4.1 and symmetric in phi. The cells in the barrel ($|\eta| < 1.4$) region have a segmentation of $\Delta\eta \times \Delta\phi = 0.087{\times}0.087$, becoming progressively larger in the endcaps ($1.4 < |\eta| < 3.0$) and the forward regions ($2.9 < |\eta| < 5.2$), due to the increasing HCAL cell size. The overlap between the endcap and the forward HCAL region at $|\eta|$=3.0 is clearly visible.

A tower is treated as a massless particle, with a direction defined by the centre of

---

[1] Particle flow reconstruction corresponds to the identifying and reconstructing of each particle in an event or in a jet, such as photons, electrons, muons, charged hadrons and neutral hadrons, followed by the best possible determination of its energy and direction.

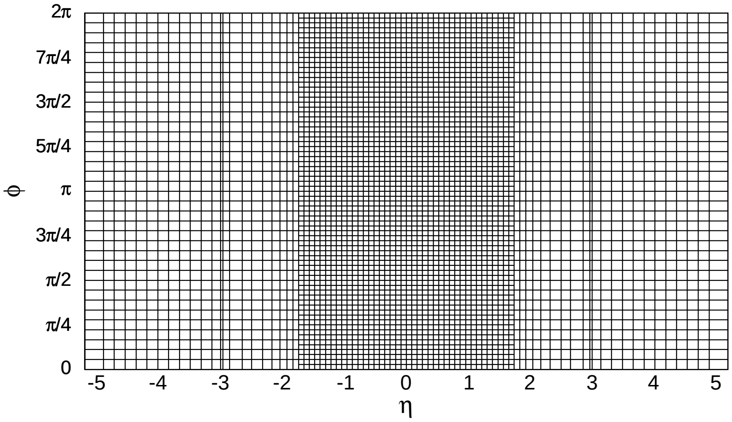


Figure 4.1: Granularity of the 4320 towers in the (η, ϕ)-plane.

Calorimeter cell thresholds (GeV)				
HB	HO	HE	ΣEB	ΣEE
0.9	1.1	1.4	0.2	0.45

Table 4.1: Calorimeter cell energy thresholds for the different HCAL and ECAL subsystems introduced in Chapter 3.

the detector and the position of the tower, and an energy that equals the sum of all the contributing read-out cells. A calorimeter cell will only be selected if its energy deposit passes the offline noise rejection energy thresholds, summarised in Table 4.1 for the different kinds of calorimeter compartments [102]. ΣEB (ΣEE) denotes the sum of all EB (EE) crystal energies contribution to a tower[2]. The relative higher thresholds for HCAL cells reflect the intrinsic higher noise level of this detector compared to the ECAL. After applying these cuts, the calorimeter noise only contributes on average for 0.3 GeV of the total energy of an iterative cone jet[3] with opening angle $R = 0.5$ clustered in the barrel region. These offline calorimeter cell thresholds will also result in an important reduction of the number of fired towers in an event. Without any offline threshold almost all towers show activity. Figure 4.2 shows that on average only ~10% of the towers pass these noise reduction cuts in semi-muonic decaying $\mathrm{t\bar{t}}$-events without pile-up.

For the same events, Figure 4.3 shows the average transverse energy deposit in a tower versus the pseudorapidity of the tower position. All towers passing the offline thresholds are included in the average. An increasing average tower E_T is observed when moving from the forward to the central region of the detector, showing a clear discontinuity at each transition of the HCAL subsystem. This behaviour is explained by the η-dependence of the E_T-observable: the same energy deposit will have a smaller E_T-value in the endcaps compared to the barrel. Additionally, in the forward regions numerous low transverse energy deposits are detected that are produced in the under-

---

[2] These various calorimeter sub-detectors where introduced in Chapter 3. EB and EE stand respectively for ECAL barrel and endcap.

[3] For a definition of this jet clustering algorithm we refer to the next paragraph.

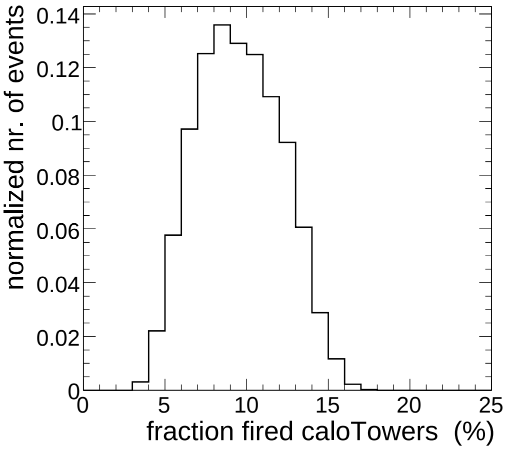


Figure 4.2: The fraction of calorimeter towers passing all cell thresholds as defined in Table 4.1.

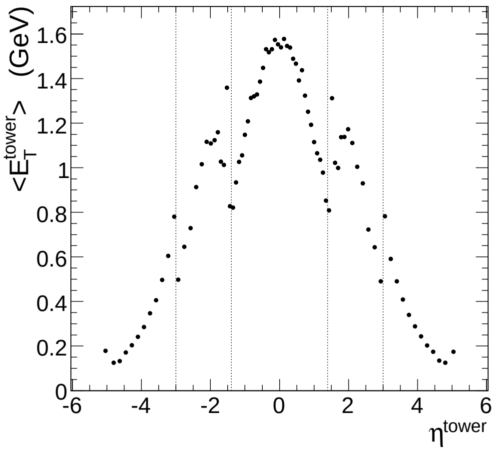


Figure 4.3: The average transverse energy deposit in a calorimeter tower versus the pseudorapidity of the tower position.

lying event, leading to smaller average tower E_T-values. The larger tower granularity in the endcaps somehow weakens the observed η-dependence, as more energy is collected in one tower in the forward region of the detector.

Furthermore, a cut of 0.5 GeV is applied on the transverse energy of all calorimeter towers [102]. To illustrate the importance of this cut, Figure 4.4 shows the distribution of the transverse energy of a tower over the pseudorapidity range. While in the right plot the result is plotted for the towers belonging to $t\bar{t}$ signal jets, only towers away from these jets where included in the left plot. Clearly, the underlying event activity results mainly in low transverse energy deposits in the forward detector region, while jets arising from the hard event are built from towers with much higher transverse energies that are more directed into the barrel region. Consequently, the extra requirement for a tower to have a minimum E_T of 0.5 GeV will reduce the influence of the underlying event significantly.

Monte-Carlo Generated Particles

To study the effects of the jet algorithms independent of the detector, the stable particles in the Monte-Carlo generated final state are often used as an input for the jet clustering algorithms. The momentum of these particles is taken at the vertex, so no deflection of the magnetic field is taken into account. In general, particles invisible for detection like neutrinos are excluded from the list. Additionally, to further mimic the calorimeter input described above, muons may be rejected from the list of input objects because of their relatively small energy deposits in the calorimeter. With this generator input the best reachable resolution on the jet kinematics can be estimated, taking into account the intrinsic fragmentation and hadronisation effects in the formation of a jet.

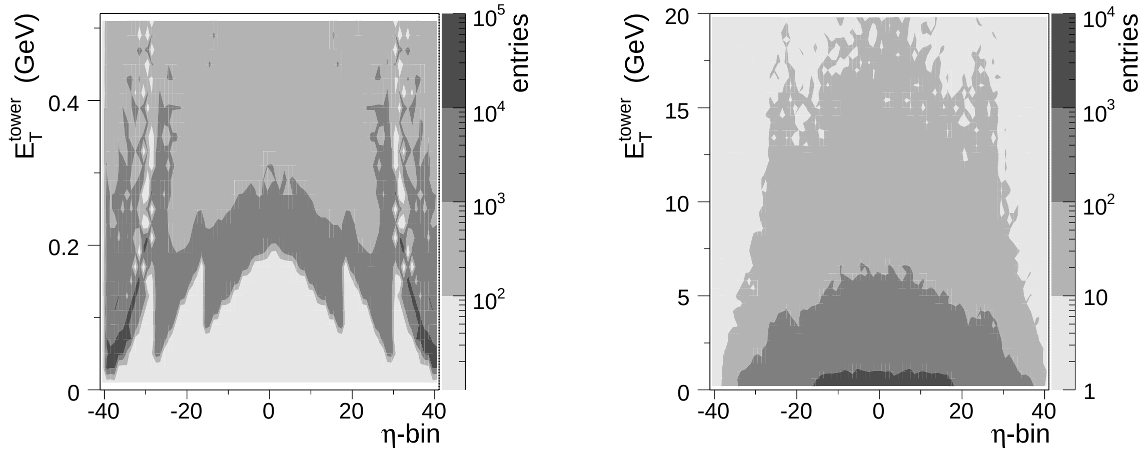


Figure 4.4: The distribution of the calorimeter towers over the (E_T^{tower},η)-plane for the underlying event (left) and $t\bar{t}$ signal jets (right). The pseudorapidity bin number corresponds to the 82 η-bins illustrated in Figure 4.1.

4.1.2 Jet Clustering Algorithms

The full jet reconstruction allows to estimate the primary parton's energy and direction from the four-vectors of the jet input objects. To achieve this, predetermined merging algorithms are defined that accumulate the four-momenta of all input objects that are in each others vicinity according to some algorithm specific metric.

Over the last two decades, these clustering techniques have greatly improved in sophistication [103]. In this chapter, three principal jet clustering algorithms are compared: the iterative cone [104], the midpoint cone [105] and the Inclusive k_T [106] jet clustering algorithm. All three methods are implemented in the CMS software. The iterative cone algorithm is simple and fast, and therefore commonly used in online trigger systems, while the more CPU intensive midpoint cone and k_T algorithms are more frequently applied in offline data analyses. While cone algorithms use the Lorentz-boost invariant[4] ΔR metric to compare the distance between two input objects, i and j,

$$\Delta R = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2} \; , \tag{4.1}$$

the k_T-based algorithms uses an E_T-weighted ΔR metric. Both cone and k_T-based algorithm are used in the Run-II of the Tevatron experiments [107, 108]. Furthermore, all algorithms are configured to use the energy recombination scheme to calculate the jet four-vector, i.e. all four-vectors of the jet constituents are vectorially summed, resulting in massive jets.

- **Iterative Cone algorithm**
  This algorithm starts with an E_T-ordering of the input objects. A cone with opening angle R in (η, ϕ)-space is then cast around the object with the highest transverse energy, if this value exceeds a threshold value $E_{T,min}^{seed}$. All constituents

---

[4] This is only true for a boost in the beam or z-direction.

in the cone are combined, and a proto-jet direction is calculated, which results in a new seed for a next iteration. This continues until a stable proto-jet energy and direction is reached. Next, the proto-jet is promoted to a jet and all objects in this jet are removed from the list of input objects. The same procedure is repeated for all remaining input objects until only objects are left with an $E_T < E^{seed}_{T,min}$. To summarise, this algorithm has to be configured via two parameters, the opening angle R and the minimal E_T of an input object to seed the clustering of a jet, $E^{seed}_{T,min}$

- **Midpoint Cone algorithm**
  With the Iterative Cone methods the cones of the reconstructed jets might overlap. The midpoint-cone algorithm is designed to facilitate the splitting and merging of such overlapping jets, and is implemented in three steps. First, all stable cones are searched, but in contrast to the iterative cone case, no input object are removed from the list. As a consequence the same object might belong to several proto-jets. In a second step, a midpoint is calculated for each overlapping pair of proto-jets as the direction of the combined momentum. The midpoints will then serve as extra seeds to find more proto-jets. Finally, in a third step, the splitting and merging procedure is applied, starting from the proto-jet with the highest E_T. If a proto-jet does not share any objects with another proto-jet, the proto-jet is added to the list of jets. Otherwise, the ratio of the transverse energy shared with the highest E_T neighbour to the total E_T of that neighbour is calculated. If this fraction is higher that a certain overlap threshold the two jets are merged. In the other case, the two proto-jets are splitted, i.e. each constituent is added to the closest jet in (η, ϕ)-space. This algorithm is hence dependent on three main parameters: the cone opening angle, the seed E_T threshold and the jet overlap threshold.

- **Inclusive k_T algorithm**
  This cluster based algorithm starts with the computation of the following distances for each input object i and each pair of objects (i, j):

$$d_i = E^2_{\mathrm{T},i} R^2 \qquad \text{and} \qquad d_{ij} = \min\left\{E^2_{\mathrm{T},i}, E^2_{\mathrm{T},j}\right\} \Delta R^2_{ij}, \tag{4.2}$$

  where R is a dimensionless parameter, and ΔR^2_{ij} is the metric defined in Eq. 4.1. The algorithm now searches for the smallest distance. If a value of type d_{ij} is found, the two input objects are merged, and the resulting object substitutes the original objects. In case a d_i-value is the minimal, the object i is removed from the list, and added to the collection of final jets. The procedure continues till all objects are included in a jet. From the definition it follows that ΔR^2_{ij} will be larger than R for all final jets. For this algorithm, the R-parameter has to be tuned[5].

---

[5] Also an exclusive mode of this k_T algorithm exists, in which contributions are kept or thrown away into a beam jet depending on their d-value with respect to the beam axis [109]. Instead of specifying the stopping scale d_{cut}, one can choose to stop merging when a given number of jets is reached.

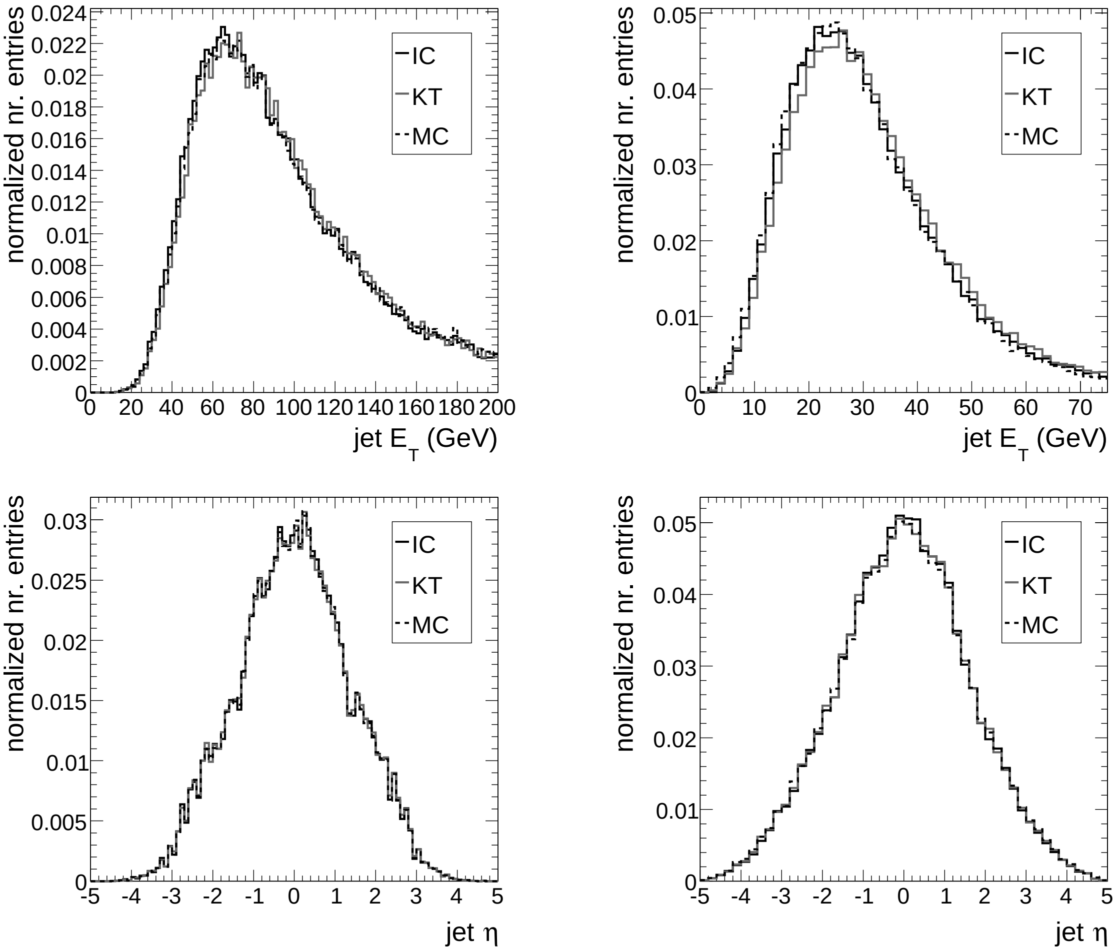


Figure 4.5: The transverse energy (top) and pseudorapidity (bottom) spectrum for both the leading (left) and the 4th highest (right) E_T-jet reconstructed in semi-muonic $\mathrm{t\bar{t}}$-events, comparing the iterative cone (IC), the k_T (KT) and the midpoint cone (MC) jet clustering algorithm.

A comparison of the jet kinematics obtained after a clusterisation with the three above defined algorithms is shown in Figure 4.5 for semi-muonic $\mathrm{t\bar{t}}$ `AlpGen` events. Both for the leading and fourth highest E_T jet the jet transverse energy and pseudorapidity spectrum are plotted for the iterative cone, midpoint cone and k_T-algorithm. For the cone algorithms an opening angle $R = 0.5$ and a minimum seed $E_{T,min}^{seed}$ of 1 GeV is used, while for the k_T-algorithm the R-parameter is set to 0.6. Additionally, for the midpoint cone algorithm, the overlap threshold for splitting and merging is set to 75%. The plots show a very good inclusive agreement for the different jet algorithms. Nevertheless, on an event-by-event basis important differences may be observed. In Section 4.2 all three algorithms and their respective configuration will be thoroughly compared via some carefully defined quality markers.

4.1.3 Jet Energy Scale Corrections

In many physics analyses the main systematic uncertainty will result from an imperfect knowledge of the jet energy scale. Dedicated studies are performed to extract precise correction factors to calibrate the reconstructed jet energy scale back to the initial parton level, as is needed for the top quark analyses in this thesis. Generally this correction happens in two main steps: a correction of the reconstructed jet to the Monte-Carlo generated particle level, and a particle-level correction to the parton level. Hereby, the Monte-Carlo generated particle-level is defined as the energy of all stable particles in a jet resulting from the hard scatter, whereas the parton-level corresponds to the energy of the parton originating the jet. In some data-driven calibration techniques however, both steps are combined.

After an introduction to the proposed jet calibration chain to be applied to the data, we will describe the main calibration tools implemented in CMS.

The Jet Calibration Chain

To well understand each step and influence in the total jet energy scale calibration, in both the Tevatron experiments the problem is decomposed into (semi-)independent factors [110, 111]. Although the ultimate sequence is still under discussion, this multi-level jet correction strategy will also be applied in the CMS experiment [112]. The following correction factors should be evaluated for each jet algorithm configuration:

1. **Offset Correction**
   In this first step the jet energy contribution from pile-up and any residual electronic noise is corrected for. Consequently, all following corrections will be luminosity independent.

2. **Relative η-Correction**
   Relative to the jet response in a control region, an η-dependent correction is applied to cancel the differences in calorimeter response among the different η-regions.

3. **Absolute p_T-Correction**
   A third factor, most sensitive to systematic effects, will apply a p_T-dependent correction to obtain the final particle level corrected jet energy scale. The main origin for this dependency are the non-linear response of the calorimeter to hadrons, the influence of the magnetic field, and undetected energy in the calorimeters.

4. **Flavour Correction**
   This factor corrects for the different response of the different types of jets (gluon, light and heavy quark).

5. **Underlying Event Correction**
   Optionally, a dedicated correction factor to reduce the influence of the underlying event may be applied.

6. **Parton Correction**
   The parton-level calibration accounts for particles from the originating parton

> that are not included in the particle-level jet. When these correction factors are estimated using simulated Monte-Carlo event, their precision depends on the hadronisation model used and the type of parton considered.

Dedicated samples and data-driven techniques will be developed to measure these calibration factors. However, Monte-Carlo calibration techniques together with the radioactive source calibration and test beam measurements will provide a starting point for understanding the initial jet calibration.

Monte-Carlo Jet Correction

A calibration method based on Monte-Carlo simulated QCD events in a range of $0 < \hat{p_T} < 4000$ GeV/c[6] is developed to correct the reconstructed jet energy response to the level of Monte-Carlo generated particle jets [113]. For several jet clustering algorithm configurations, reconstructed jets are matched to particle level jets if both are closer than 0.2 in ΔR-space. To construct the particle level jets, only stable generated particles are used, including neutrinos and muons. The particle jets are divided into (E_T^{GEN}, η^{GEN})-bins, and in each bin the ratio of the reconstructed jet transverse energy (E_T^{REC}) to the particle level jet transverse energy (E_T^{GEN}) is fitted with a Gaussian. Figure 4.6 shows for the different E_T^{GEN} subsamples the E_T^{REC}/E_T^{GEN} fitted mean values as a function of the pseudorapidity of the particle level jet. In this plot iterative cone jets with an opening angle R=0.5 are used. For all bins, the transverse energy response is underestimated, due to low E_T-particles not arriving at the calorimeter surface because of the magnetic field, invisible particles such as neutrinos and energy leakage resulting from e.g. dead material and cracks. The fraction of transverse energy lost because of particles curved outside of the cone will increase for lower E_T^{GEN}-bins. Furthermore, a similar behaviour in the response is observed for the different E_T^{GEN}-bins when moving from the barrel to the endcap region, what motivates the strategy to factorise the η-correction from the E_T correction.

In Figure 4.7 and Figure 4.8 the E^{JET}/E^{GEN}-ratio is plotted for semi-leptonic $t\bar{t}+2j$ `AlpGen` events, respectively before (E^{REC}) and after (E^{CAL}) applying the Monte-Carlo jet corrections. In these plots only jets with a reconstructed transverse energy above 10 GeV and a pseudorapidity lower than 2.4 are used, as all other jets will not be considered in the later analyses. The decomposition of the jet sample in the different parton flavours illustrates the much lower jet response for gluon jets and b-quark jets with respect to the light quark jets. After applying the correction factors, we observe a jet response close to unity for the gluon jets, but an overall overcalibration, especially for the light quark jets. This is explained by the high gluon jet fraction in the QCD events, used to produce the MC jet correction factors. The leptonic decay of the b-quark, $b \rightarrow c\, l\, \nu_l$ in ~20% of the b-decays, results in undetected energy in the calorimeter and hence a lower response with respect to the light quark jets. Furthermore, we should take into account that this MC jet calibration only corrects the reconstructed jet energies to the particle-level. Figure 4.9 shows for the different parton flavours the jet energy response as a function of the reconstructed transverse energy

---

[6] Hereby, $\hat{p_T}$ is defined as the transverse momentum in the rest frame of a 2→2 processes such as a QCD dijet production.

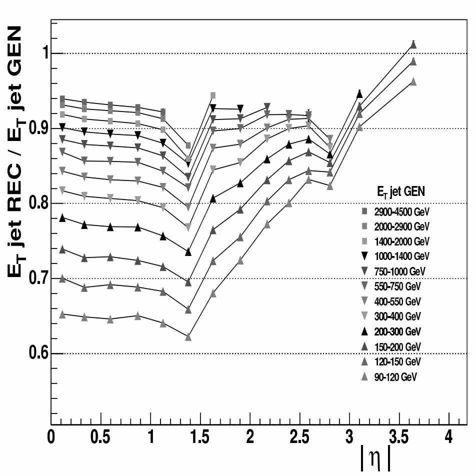


Figure 4.6: Monte-Carlo jet energy correction curves as a function of η^{GEN} for several bins of E_T^{GEN}, based on iterative cone clustered QCD jets with an opening angle $R = 0.5$.

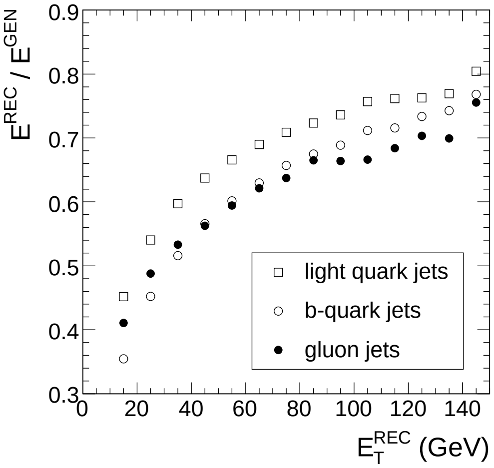


Figure 4.7: The E^{REC}/E^{GEN} ratio for the different jet-types as a function of E_T^{REC}. Only jets from semi-leptonic $t\bar{t} + 2j$ `AlpGen` events with a $E^{REC} > 10\,\text{GeV}$ and $|\eta|^{REC} < 2.4$ are used.

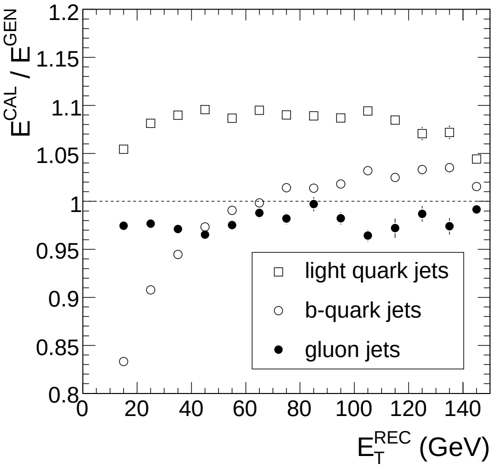


Figure 4.8: The MCJet corrected jet response, E^{CAL} / E^{GEN}, versus E_T^{REC}, using the same jet sample as for Figure 4.7.

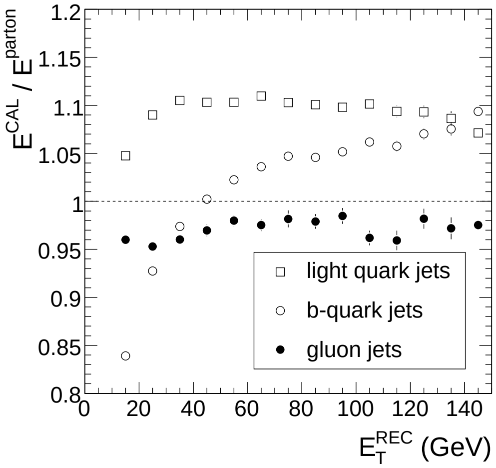


Figure 4.9: The MCJet corrected jet response with respect to the original parton energy, versus E_T^{REC}, using the same jet sample as for Figure 4.7.

with respect to the original parton level. This plot indicates an overcalibration of the light quark jet energy of ∼9% and an undercalibration of almost 4% for gluon jets. In the case of b-quark jets the jet energy reconstruction is shown to be very dependent on the transverse energy of the jet. Where for a reconstructed transverse energy of 20 GeV the b-quark energy scale is more than 10% underestimated, for high energetic b-quark jets an overestimate of more than 5% is observed.

Data driven techniques

To complement the radioactive source calibration at the tower level as well as to extract the correct jet calibration factors defined above, dedicated data samples will be used. Three main calibration procedures are studied in CMS:

- **p_T-balance in dijet events** [114]
  Posing the demand of an equal transverse energy of both jets reconstructed in QCD dijet events is a powerful method to obtain the relative η-correction factors, introduced as the level-2 correction in the description of the calibration chain. In this technique one of both jets, the reference jet, should be reconstructed in a reference η-region, usually taken the barrel region. Several η-bins are constructed corresponding to the pseudorapidity of the second jet, denoted as the probe jet, and for each of these bins a relative correction is estimated from the $(p_T^{probe} - p_T^{ref})$ / p_T^{ref} distribution. Due to the large QCD dijet production cross section, this method has the advantage of being applicable almost immediately after startup. Figure 4.10 illustrates the precision that is reachable on the η^{probe}-dependent relative jet response with only one hour of data-taking at $\mathcal{L}$=2 10^{33}cm^{-2}s^{-1}, both before and after applying the relative η-calibration. On average the corrected jet response is in agreement with the jet response in the $|\eta| < 1$ reference region over the whole η^{probe}-range.

- **p_T-balance in γ+jet events** [115]
  Similarly to the case of QCD dijet events, the transverse momentum balancing technique can be applied to $\gamma + jet$ events produced via the $qg \to q\gamma$ or $q\bar{q} \to g\gamma$ process. At leading order, i.e. without any initial state radiation, the transverse momentum of the jet is equal to the photon transverse energy, p_T^{γ}. The excellent energy resolution of the electromagnetic calorimeter to measure the photon energy allows the extraction of absolute p_T-calibration factors, that will calibrate the reconstructed jets back to the initial parton level. Consequently, in terms of the different calibration factors described in the calibration chain, this technique combines a Level-3 and Level-6 correction. For different p_T^{γ}-bins, the spectrum of $k_{jet} = p_T^{jet}/p_T^{\gamma} \simeq p_T^{jet}/p_T^{Parton}$ is fitted, resulting in a p_T^{γ}-dependent calibration curve, illustrated in Figure 4.11. In this plot the calibration curve extracted from the $\gamma + jet$-events is compared with the MC true calibration factors for a sample of only quark jets and a QCD-sample that contains mainly gluon jets. Analogue to Figure 4.7, an important shift in jet response is observed between quark and gluon jets. As both jet flavours are present in the $\gamma + jet$ final state, the corresponding calibration curve lies intermediate. Also this method will show itself useful from the first days of data-taking.

- **W-mass constraint in tagged $\mathrm{t\bar{t}}$ events** [116]
  In this method, a pure semi-leptonic $\mathrm{t\bar{t}}$ sample is selected from the data, delivering for each event two light quark jets that are expected to have an invariant mass equal to the W-mass. This constraint allows a robust extraction of an inclusive calibration constant. Chapter 5 will further elaborate on this very powerful and promising jet calibration method, resulting in a way to extract a parton-level

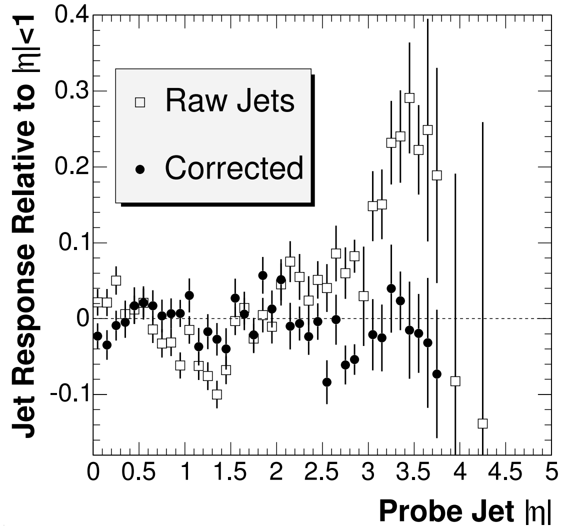


Figure 4.10: Jet response as a function of the $|\eta|$-value of the probe jet. The graphs compare jets before and after the relative η-correction is applied.

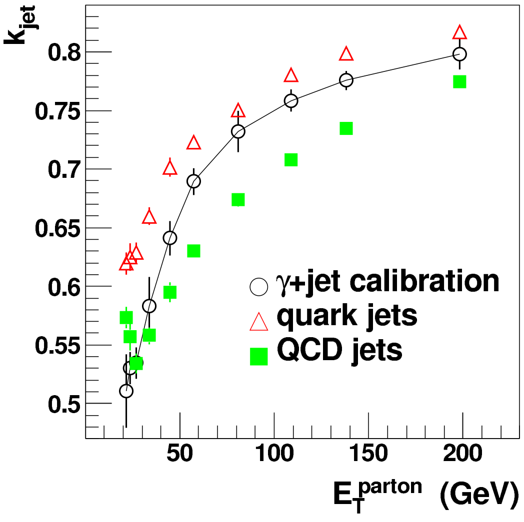


Figure 4.11: Comparison of the outcome of the γ + *jet*-calibration method with the true values for quarks and QCD jets (mainly gluon-jets), for iterative cone jets with $|\eta| < 1.5$.

correction factor for light quark jets (corresponding to the correction factors 3 and 6 in the jet calibration chain).

From the dijet and $\gamma + jet$ calibration techniques, a 5% overall uncertainty on the jet energy scale can be achieved for the first 1-10 fb^{-1} [88, 117]. Including the W-mass constraint will further reduce this uncertainty to 1% for light quark jets with a p_T > 30 GeV/c.

4.1.4 Jet Resolutions in Semi-leptonic t$\bar{\text{t}}$-events

Analogue to the determination of the Monte-Carlo calibration curves, jet resolutions may be obtained via the matching of reconstructed jets with particle level jets. As an example, in Figure 4.12 the E^{JET}/E^{GEN}-distribution is shown for light quark jets in semi-leptonic t$\bar{\text{t}}$-events with a reconstructed transverse energy between 80 GeV and 90 GeV, both before and after applying the Monte-Carlo jet corrections. For this plot, only jets in the tracker acceptance with a well matching parton ($\Delta R < 0.2$) are used. After the MC jet calibration, not only the light quark jet response is shifted, also the RMS of the distribution gets larger. However, on the contrary, the relative spread of the E^{JET}/E^{GEN}-distribution decreases when applying the MC jet correction. This is illustrated in Figure 4.13, where the relative jet resolution curves are drawn for both reconstructed and calibrated jets. This improvement in relative jet resolution is most pronounced for small E_T^{REC}-values and is the result of the E_T- and η-dependency of the MC jet correction factors (cfr. Figure 4.6). Consequently, the whole jet calibration chain and methods discussed in the previous section will not only result in a reduced average bias of the correct jet energy scale, but will as well improve the relative resolution and hence uncertainty on the energy measurement.

Another interesting observation is the difference in resolution for the different jet flavours, illustrated in Figures 4.14 and 4.15. The first plot shows the E^{CAL}/E^{GEN}-distribution for the different jet flavours, using the same jets as in Figure 4.12, while in the second plot the respective relative jet resolution curves as a function of the reconstructed jet E_T are plotted. A worse jet energy resolution is observed for b-quark jets, explained by the leptonic decay of the b-quark in around 20% of the b-quark jets, and consequently the undetected neutrino and muon energies in the calorimeter cells.

For the top quark analyses presented in the next chapters, an estimate is needed of the resolution on the measured jet kinematics. Because in these analyses the calibrated jet four-vectors will be treated as the initial quark four-vector, also this resolution should be defined with respect to the parton level instead of to the generated particle level jet. For three jet variables the absolute resolution curves are plotted in Figure 4.16 as a function of the calibrated jet transverse energy, and for different bins of the pseudorapidity. In the left plots only light quark jets are used from semi-muonic decaying $\mathrm{t\bar{t}}$-events, while b-quark jets are selected as input for the right plots. The upper left plot shows the resolution on the transverse calibrated energy measurement E_T^{CAL}. A first observation is the increasing absolute, but decreasing relative, resolution for an increasing calibrated transverse energy. This is explained by the lower fraction of out of cone energy because of the higher boost, smaller relative contributions from the underlying event, ... Another observation is the improvement of the uncertainty on this energy related variable for high pseudorapidity bins. This is explained by the relatively higher jet energies corresponding to the same E_T^{CAL}-bin for jets in more forward directions, what leads to a better jet energy reconstruction. The last two plots on the left in Figure 4.16 illustrate the absolute resolutions on the angles θ^{CAL} and ϕ^{CAL}. Another time an improving resolution for increasing jet transverse energy is observed. When comparing the θ and ϕ results, a worse resolution is obtained for the reconstruction of the polar angle θ. This is due to the magnetic field that bends the jet constituents only in the transverse plane of the detector, resulting in an extra smearing of the ϕ^{CAL}-angle. Furthermore, the θ resolution is observed to improve for increasing η-values. This effect is a direct result from the applied ΔR jet-parton matching criterion. Requiring jets close to an initial parton in ΔR-space puts more stringent demands on the θ reconstruction in forward regions, what leads to better resolutions on the polar angle for increasing η-values. To summarise, a light quark jet with $E_T^{CAL} = 40\,GeV$ reconstructed in the barrel has an uncertainty on its E_T^{CAL}-measurement of around 8 GeV, and uncertainties on θ^{CAL} and ϕ^{CAL} of respectively 0.06 and 0.08 rad. When comparing the results for both quark flavours, only an increase in energy uncertainty is observed for the b-quark jets, as is expected from the previous plots.

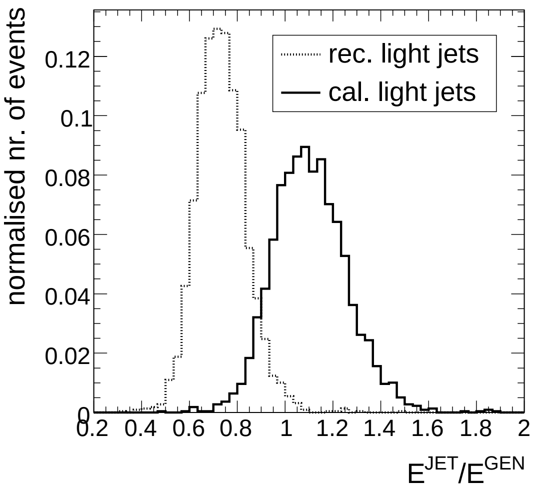


Figure 4.12: The fraction of the particle level jet energy that is reconstructed, before and after applying the MCJet correction. For this plot only light quark from semi-leptonic $t\bar{t}$-events, reconstructed in within the tracker acceptance with a E_T^{REC} between 80 GeV and 90 GeV are considered.

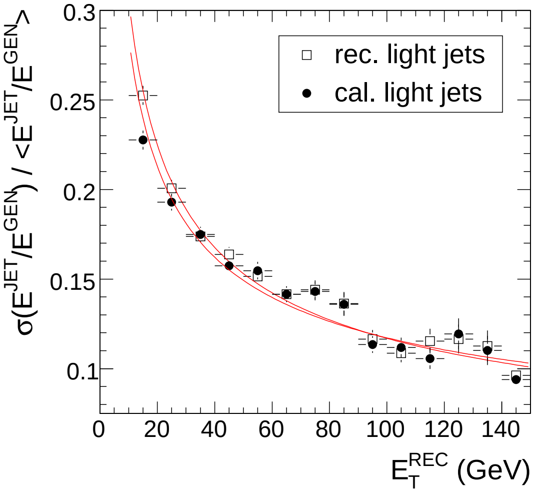


Figure 4.13: For calibrated and uncalibrated jets, the relative jet energy resolution curve versus the transverse energy of the light quark jet.

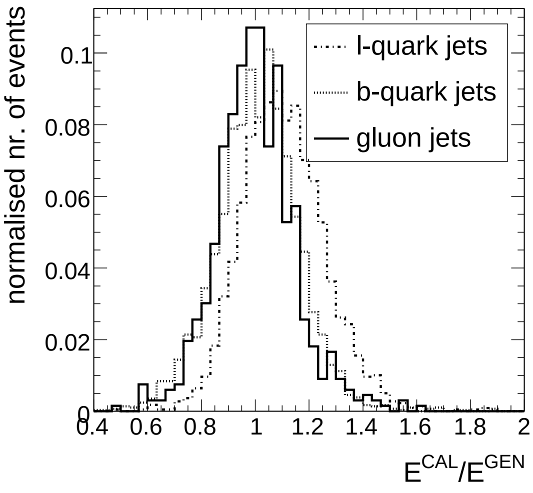


Figure 4.14: The fraction of the particle level light quark, b-quark or gluon jet energy that is reconstructed. The plot shows the result for the 80 GeV < E_T^{REC} < 90 GeV bin.

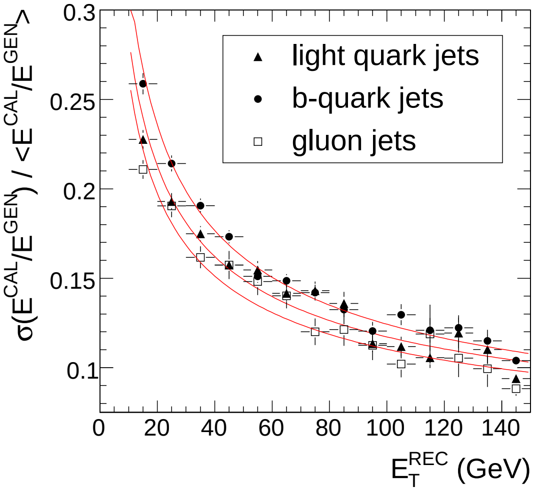


Figure 4.15: For the different jet-types, the jet energy resolution curve versus the transverse energy of the reconstructed jet.

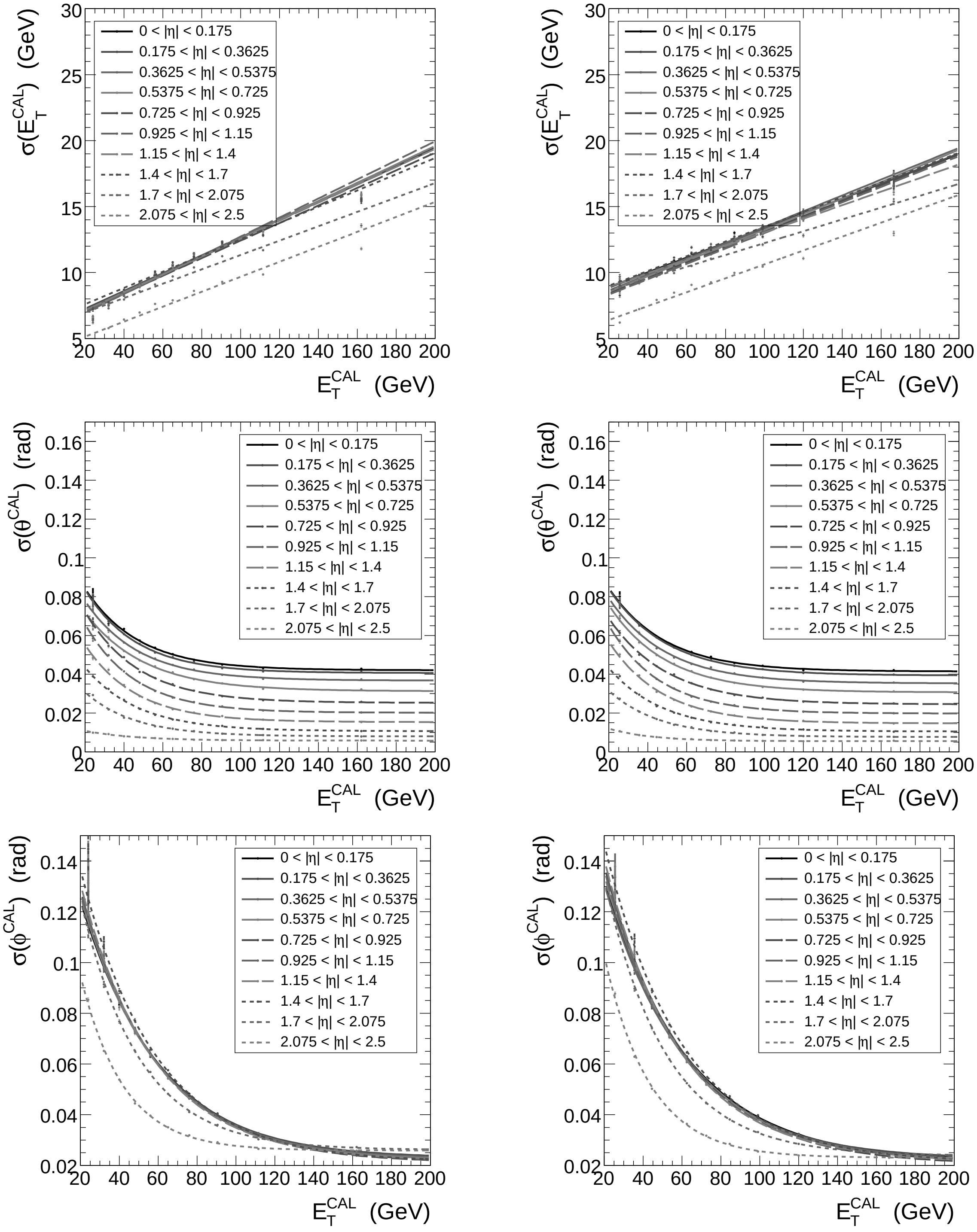


Figure 4.16: Absolute light quark (left) and b-quark (right) jet resolutions on E_T^{CAL}, θ^{CAL} and ϕ^{CAL} for different bins in η, using iterative-cone jets with $R = 0.5$ reconstructed in semi-muonic $\mathrm{t\bar{t}}$ events in the $|\eta| < 2.4$ range.

4.2 Search for Optimal Jet Clustering Algorithm Configurations

In many physics analyses studying events with hard partons produced in the final state, the performance and efficiency of the jet clustering algorithms will be crucial for the precision and physics potential achievable in the analysis. Therefore, a study is performed to define a method able to compare different jet clustering algorithms and their corresponding configurations. We started from an analysis perspective, which means that the jet clustering is considered to be optimal if the reconstruction efficiency of the complete kinematics of the primary quark event topology is maximised.

This reconstruction efficiency will be defined in terms of some quality markers, quantifying a distance between the generated primary partons and the reconstructed jets, and therefore the error of the jets, in both energy and angular space. Physics effects like the underlying event and hard gluon radiation enlarge this mean error. The scope of this study is to find the most efficient configuration for jet clustering algorithms in the presence of these effects, in order to maximise the fraction of events for which all final state quarks in the leading order process have smaller errors than some predefined criteria. Consequently, events suffering from hard gluon radiation will degrade the overall reconstruction efficiency.

The resulting efficiency does not only depend on the configuration of the jet clustering algorithm but also on the event topology. In this section however, we will only present the results for semi-leptonically decaying $\mathrm{t\bar{t}}$-events, which have four quarks produced in the final state. For the comparison of several jet multiplicities we refer to our contribution in [118].

4.2.1 Jet Selection

For this study to be meaningful and relevant for a physics analysis, a realistic jet selection should be applied. Therefore, we ask the jet direction to be within the tracker acceptance to allow proper b-tagging ($|\eta| < 2.4$), and the reconstructed transverse energy to exceed 10 GeV in the case when calorimeter towers are used as jet input, or 20 GeV for a MC generated particles input. We consider only semi-muonic $\mathrm{t\bar{t}}$ decays, to avoid calorimeter deposits from an electron or a tau decay.

4.2.2 Description of the Quality Markers

In order to obtain an efficient reconstruction of the kinematics of the primary partons, the selected jets should match both in energy and direction the primary partons. Quality markers are defined to quantify the goodness of the event reconstruction from that perspective.

Event selection efficiency, ϵ_{sel}

The event selection efficiency ϵ_{sel} is defined as the fraction of events in which four jets are selected that pass the above mentioned jet selection criteria.

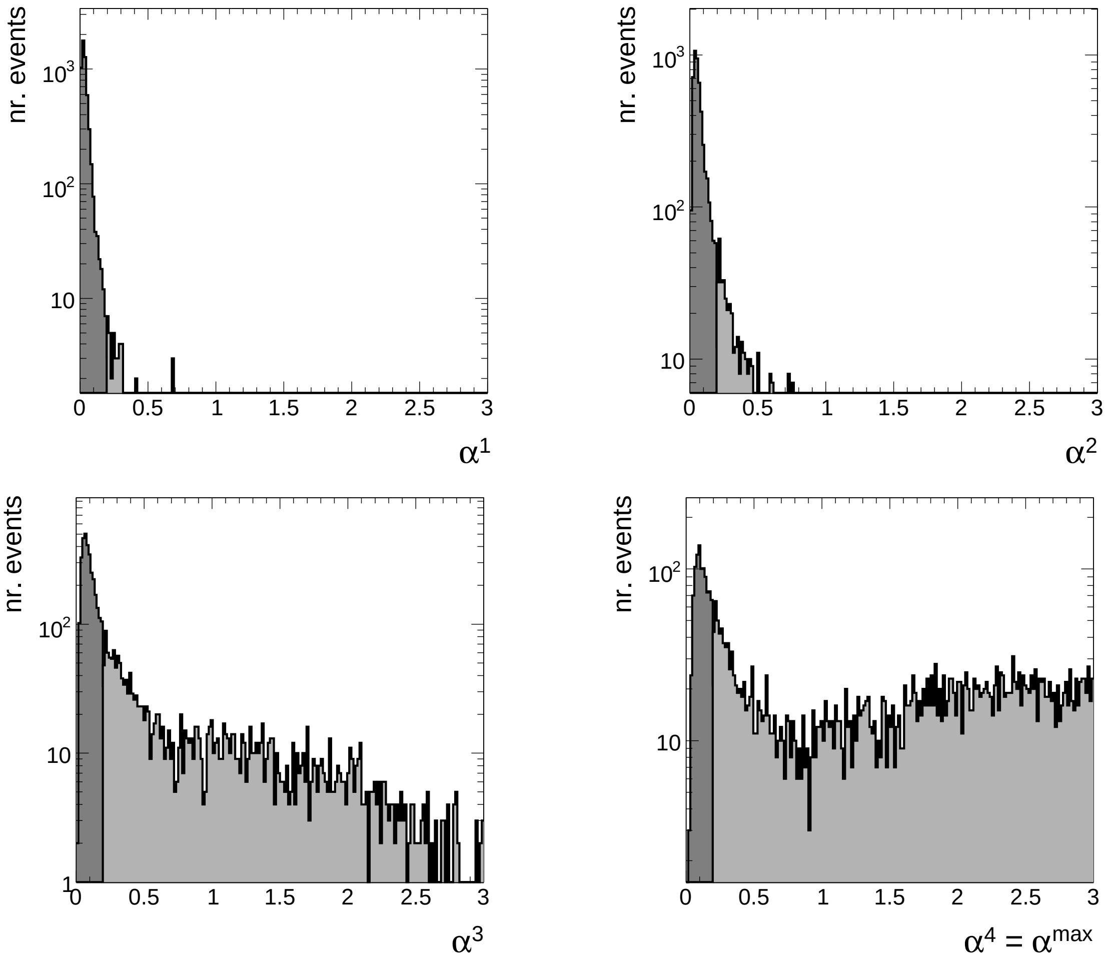


Figure 4.17: Distributions of α^i in increasing order for iterative cone jets with a jet radius $R = 0.5$ and calorimeter towers selected as jet input. The ΔR <0.2 rad criteria as discussed in the text is indicated.

Jet direction related quality marker, $Frac\,(\alpha^{max})$

A jet is considered to be well reconstructed, if the ΔR distance between its direction and the best matching parton direction, α, is sufficiently small. For each semi-muonic decaying $t\bar{t}$-event, this results in a list of four increasing α^i-values, $\{\alpha^1, \ldots, \alpha^4 = \alpha^{max}\}$, of which an example of their distribution is shown in Figure 4.17 for iterative cone jets with a jet radius $R = 0.5$. It is shown that on average two out of the four jets are almost always well reconstructed from an angular point of view, but that the demand to have as well the third and fourth jet matched well to a parton often rejects the whole event from being well reconstructed. To quantify the angular reconstruction performance of the events using a particular jet definition, a quality marker is defined as the fraction of events with a α^{max} value lower than 0.2 rad, denoted as $Frac\,(\alpha^{max})$. This cut-off value is illustrated in the example distributions, and results for the iterative cone $R = 0.5$ jets in a $Frac\,(\alpha^{max})$ value equal to 18.3%. This relatively low efficiency is the result of effect like hard gluon radiation, underlying event activity and overlapping jets.

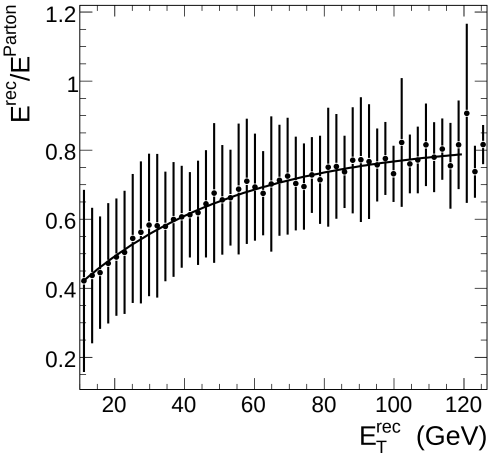


Figure 4.18: Example of a $\frac{E^{REC}}{E^{Parton}}$ vs. E_T^{REC} calibration curve constructed from iterative cone jets with a jet radius $R = 0.5$, using calorimeter towers as jet input. The vertical error-bars illustrate the resolution.

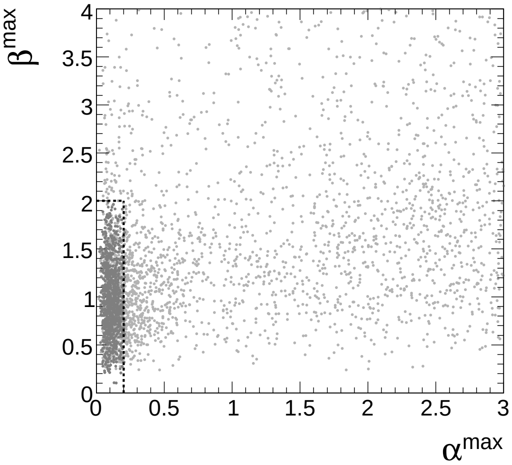


Figure 4.19: Scatter plot of β^{max} vs. α^{max} for iterative cone jets with a jet radius $R = 0.5$ and calorimeter towers selected as jet input. The rectangular area represents the events in which all four jets are well clustered from energy and angular point of view.

However, the choice of considering only to the worst reconstructed jet is motivated by the requirement that the direction of all primary quarks in the event should to be well determined in order to use an event in a physics analysis.

Jet energy related quality marker, $Frac\,(\beta^{max})$

The raw reconstructed energy of the primary quarks is biased and has a broad resolution. As described in the previous section, it is the aim of jet calibration studies to determine the average correction factors to be applied to the reconstructed jet energies. In this study, we define jets badly clustered from energy point of view, if its E^{REC}/E^{Parton}-value is more than two standard deviations away from the average fraction in a certain E_T^{REC} bin. Hence, in a first step a resolution curve is constructed for each studied jet definition, like illustrated in Figure 4.18 for iterative cone $R = 0.5$ jets, in which the error bars indicate the jet resolution in a given E_T^{REC}-bin. This curve is then used as a reference to calculate the number of standard deviations a jet response is away from the curve. This number, denoted β, is calculated for all four jets in the event. For each selected event, the primary quark with the highest β^i value, called β^{max}, is considered to be the one with the worst energy reconstruction. An energy related quality marker is then defined as the fraction of events with a β^{max} lower than 2 standard deviations, denoted as $Frac\,(\beta^{max})$. An example for these β^i distributions is shown in Figure 4.20 for iterative cone $R = 0.5$ jets, in which the 2σ cut-off value is illustrated. Analogue to the observations for the α-distributions, these plots show that also from reconstructed energy point of view in most events minimum two jets are well clustered. In 49% of the events even the worst jet still has a β-value lower than 2σ.

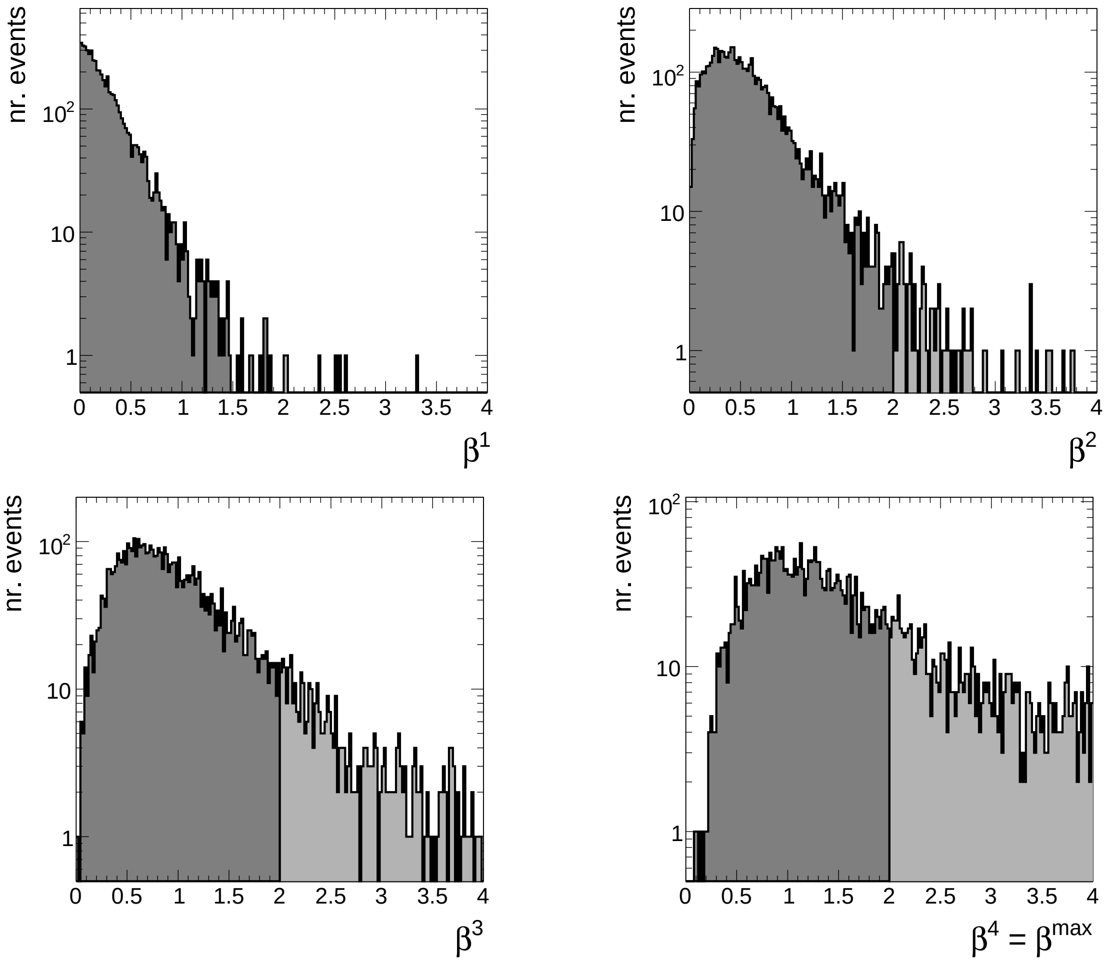


Figure 4.20: Distributions of β^i in increasing order for the iterative cone jets with a jet radius $R = 0.5$ using calorimeter towers as jet input. The 2σ cut-off value is illustrated.

Jet direction and energy combined quality marker, $Frac\,(\alpha^{max} + \beta^{max})$

A combined variable, $Frac\,(\alpha^{max} + \beta^{max})$, is defined as the fraction of events in which both the direction and the energy of the four primary quarks are well reconstructed following the above criteria. The correlation between α^{max} and β^{max} is shown in Figure 4.19 for iterative cone R =0.5 jets and found to be 40%. Both quality criteria define a rectangular area in which the kinematics of the primary quarks are sufficiently well reconstructed. In the given example a $Frac\,(\alpha^{max} + \beta^{max}) = 16.8\%$ is found.

Overall quality marker, $\epsilon_{sel} \times Frac\,(\alpha^{max} + \beta^{max})$

The fraction of selected and well reconstructed events, i.e. the selection efficiency ϵ_{sel} multiplied by the combined quality marker $Frac\,(\alpha^{max} + \beta^{max})$, can be interpreted as an estimate for the reconstruction efficiency of the kinematics of the primary quarks of the complete event. This overall quality marker may therefore be used to compare different algorithms and configurations. However, although this variable gives a powerful indication of the goodness of a certain jet definition, it is sometimes useful to consider the partial information of the individual quality markers. Depending on the priorities of a specific physics analysis, one could be interested to optimise the selection efficiency rather than the full reconstruction performance of the event.

4.2.3 Results

In this section the most important observations for each jet clustering algorithm introduced in section 4.1.2 are summarised for two types of jet input: Monte-Carlo generated particles (excluding neutrinos and muons) and calorimeter towers.

Iterative cone algorithm

The optimal configuration for the iterative cone algorithm is searched for by varying the jet cone radius. The result of this scan is summarised in Figure 4.21. The upper plot shows, for both studied jet inputs, the selection efficiency versus cone opening angle. Already this quality marker indicates an optimal value around 0.5 for the calorimeter tower input. For lower values it becomes more difficult to collect enough energy in the cone to exceed the applied jet E_T-cut, while mainly the higher probability for overlapping or merged jets is responsible for the decrease in selection efficiency at higher values. This drop is however less pronounced when generated particles are taken as jet input. Because in this latter case the jet constituents are not deflected by the magnetic field, more narrow jets are reconstructed, resulting in a smaller overlap and merging probability. This difference in broadness between generated and calorimeter jets explains as well the shift to smaller optimal opening angles visible in the $Frac\,(\alpha^{max} + \beta^{max})$ versus cone opening angle plot. However, when multiplying both quality markers, the optimal cone opening angle turns out very similar for both types of jet input. In the case of generated particle jets the highest fraction of selected and well reconstructed events (~11.0%) is found for a jet cone opening angle of 0.4, where for the calorimetric jet input this optimal value shifted to 0.5 corresponding to a fraction of ~9.0% of high quality events. Although the clustering performance between 0.4 and 0.5 for the jet opening angle is almost invariant.

k$_T$ algorithm

For the k$_T$ algorithm the influence of the R-parameter on the reconstruction efficiency is investigated. The results are shown in Figure 4.22, and appear very similar to the case of the iterative cone algorithm. The same arguments to describe the jet selection and clustering performance graphs observed for iterative cone jets as a function of the jet cone radius apply to the k_T-jets. In addition, a small increase is observed of the fraction of selected and well reconstructed events observed for both types of jet input. For generated jets this fraction is now found to be 11.2% (+0.2%) for an optimal R-parameter of 0.5, while for calorimeter jets a maximum of 9.3% (+0.3%) is obtained for a R-parameter equal to 0.6.

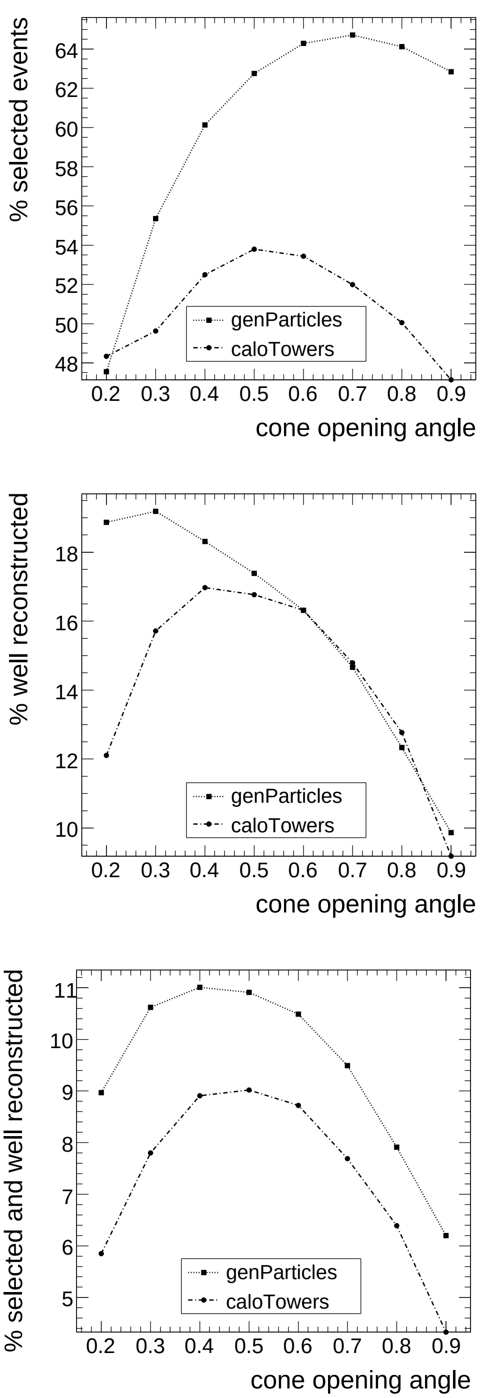


Figure 4.21: Scan of the iterative cone algorithm performance versus the cone opening angle.

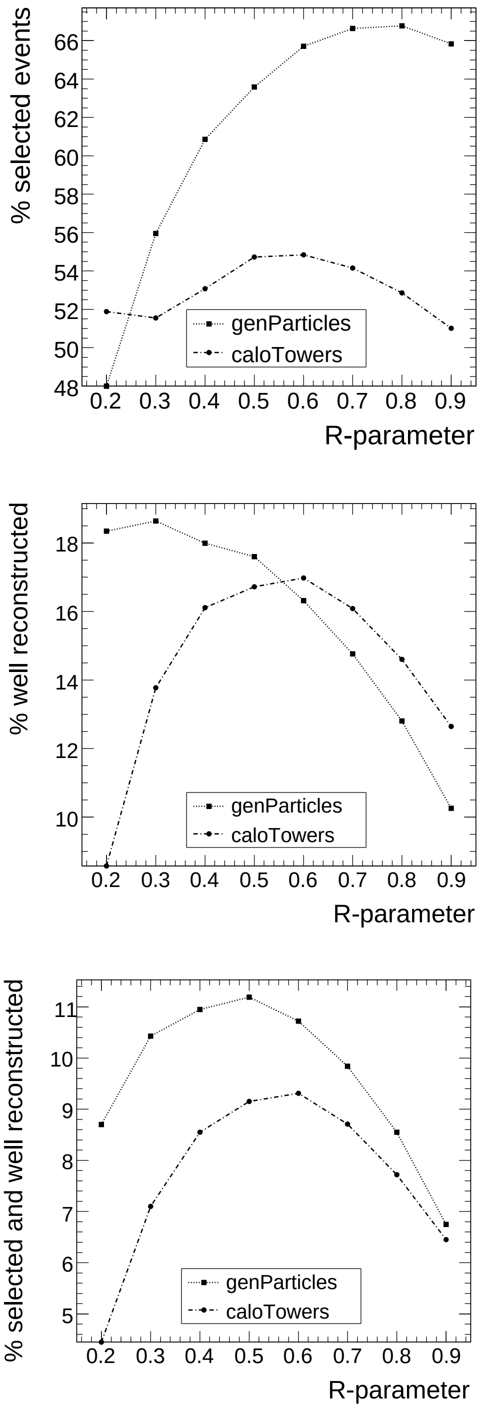


Figure 4.22: Scan of the k_T algorithm performance versus the R-parameter.

	Iterative Cone *jet radius*	k_T $R-parameter$	Midpoint Cone *jet radius / overlap threshold*
generated particles	0.4 (11.0%)	0.5 (11.2%)	0.3/50% (10.9%)
calorimeter towers	0.5 (9.0%)	0.6 (9.3%)	0.4/50% (9.0%)

Table 4.2: Overview of the optimal parameter values with their respective estimate of the fraction of selected and well reconstructed events found for the different scanned jet algorithm parameters.

Midpoint cone algorithm

For the midpoint cone algorithm, two parameters will be mainly responsible for the reconstruction efficiency: the jet cone opening angle and the cone overlap threshold, defined as the fraction of energy that two overlapping cones should share for both jets to be merged. A two dimensional scan, plotted for both types of jet input in Figure 4.23, is performed to identify those regions in the parameter space with the highest selection efficiency, the best reconstruction performance and the highest fraction of selected and well reconstructed events. When we consider the upper plots showing the results for the selection efficiency, for both the generated particles input and the calorimeter towers a jet cone radius around 0.4-0.5 is preferred. The selection efficiency tends to be higher for rather small values for the jet overlap threshold, although this behaviour is less pronounced for the calorimeter input. In the middle plots, where the fraction of well reconstructed events is shown versus both midpoint cone parameters, a clear correlation is observed between the cone opening angle and the overlap threshold. The broader the jet cone is taken, the larger the overlap between two jets should be in order to merge both jets. In the case of generated particles jets the highest $Frac\,(\alpha^{max} + \beta^{max})$ values are obtained for very small cone radii. For these small jet cone opening angles the probability to have overlapping jets is so small that the clustering performance is independent of the chosen value of the jet overlap threshold. This is no longer true when calorimeter towers are selected as jet input. When convoluting the selection efficiency with the fraction of well reconstructed events the bottom plots are obtained. Where for the generated input a comparable fraction of $\sim$10% of selected and well reconstructed events is found in the area of relatively small cone radii, a clear prefered region with $\sim$9% of well clustered events can be identified for the calorimeter jets, pointing to an optimal opening angle close to 0.4 with a jet overlap threshold around 50%.

Eventually, an overview of the main results from this jet study is given in Table 4.2. Because of the very small differences in the inclusive performance of all three algorithms, the iterative cone algorithm is chosen to be used in all analyses presented in this thesis. The inclusion of pile-up collisions is expected to keep the optimal jet clustering parameters quasi invariant, but would reduce the lower overall reconstruction efficiencies.

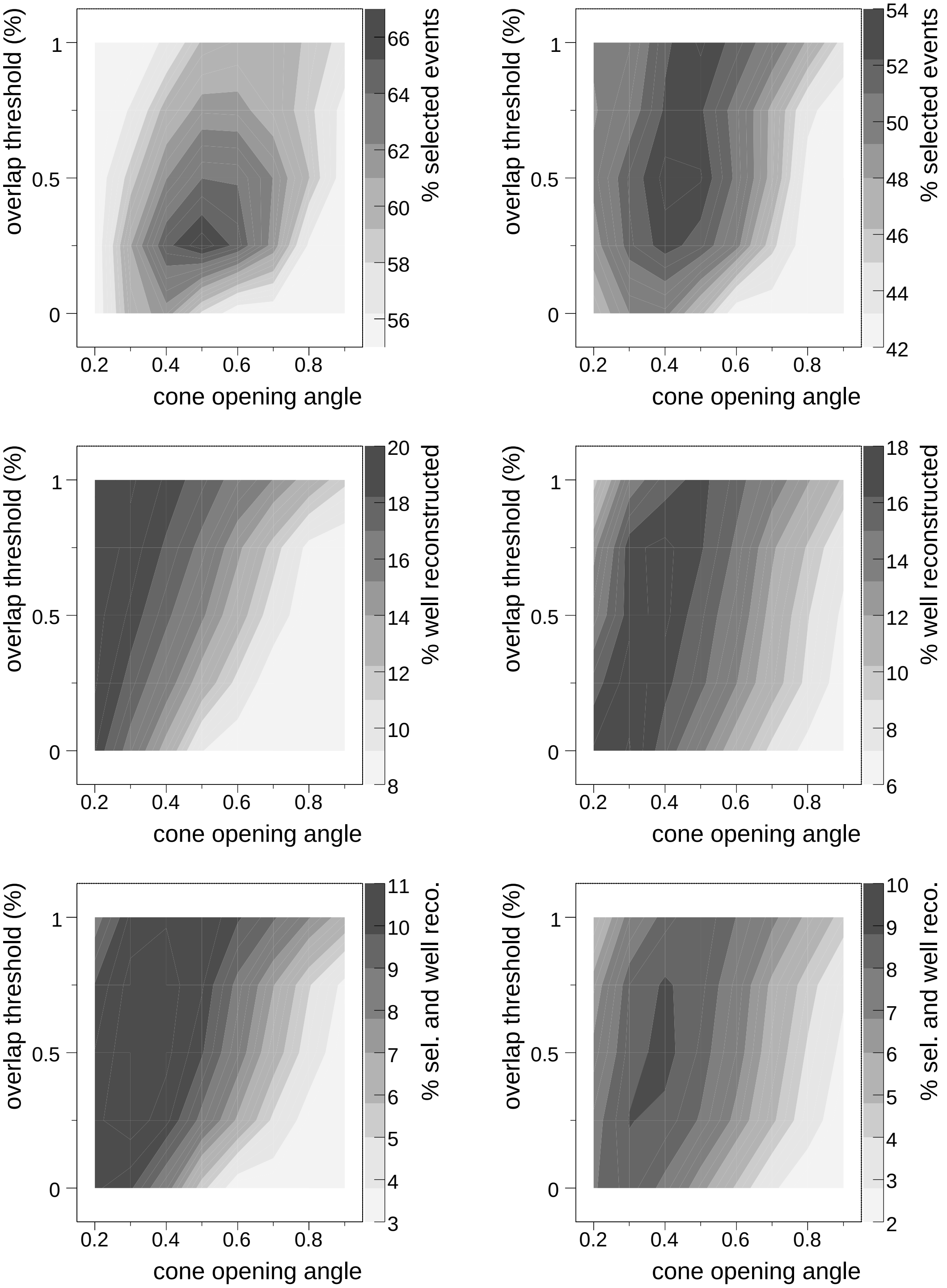


Figure 4.23: Two-dimensional scan of the jet clustering performance for the midpoint cone algorithm versus the jet opening angle and the cone overlap threshold. The plots on the left show the result for generator particles as jet input, while for the right plots calorimeter towers are used.

4.3 Missing Transverse Energy Reconstruction

Beginning with UA1 [119], all major detectors at hadron colliders have been designed to cover as much solid angle as practically possible with calorimetry. The primary motivation was to provide a method to identify the presence of one or more energetic neutrinos or other weakly-interacting stable particles in the final state through apparent missing energy. Because energetic particles produced in the direction of the beam pipe spoil the possibility to measure this missing energy longitudinal to the beam direction, only the transverse energy balance can be determined. This missing transverse energy measurement, however, is found to be a powerful tool to help the detection of physics signatures involving one or more non-interacting particles. To illustrate, the W-boson is discovered and its mass determined to 3% with just 6 events due to the ability of UA1 detector to infer the presence of 40 GeV neutrinos with a resolution of a few GeV [120]. Ever since, the measurement of missing transverse energy has always been a major tool in the search for new phenomena at hadron colliders [121].

Also the CMS detector is designed with an excellent cell segmentation, hermeticity, and good forward coverage to assist in the measurement of E_T^{miss} [122]. Nevertheless, various detector factors are known to degrade the precision achievable, such as the non-linearity of the calorimeter response, the magnetic field and its effect on low p_T charged particles, electronic noise, pile-up collisions, and the underlying event. In spite of all these detector subtleties, the E_T^{miss} resolution in CMS is expected to be dominated by the calorimeter energy resolution. To illustrate, an E_T^{miss} resolution of 6.1 GeV is found in minimum bias events without pile-up collisions and with no intrinsic missing energy produced in the final state [123]. The missing transverse energy in these events is calculated as the vectorial sum of all energy deposits in the calorimeter towers.

Several techniques are however known to improve this detector resolution. A straightforward correction is the inclusion of the reconstructed muon momenta in the E_T^{miss} calculation. Although the effect of this correction will be small in QCD events, an important improvement on the reconstructed E_T^{miss} spectrum is expected in events with high p_T muons produced in the final state. A second technique to improve the E_T^{miss} reconstruction, is to take into account the jet calibration corrections in the reconstruction of the transverse missing momentum by subtracting the shifts in momentum on the reconstructed jets resulting from the Monte-Carlo jet calibration:

$$\left\{ \begin{array}{l} E_{x,corr}^{miss} = E_x^{miss} - \sum_{j=1}^{n}(p_{x,cali}^j - p_x^j) \\ \\ E_{y,corr}^{miss} = E_y^{miss} - \sum_{j=1}^{n}(p_{y,cali}^j - p_y^j), \end{array} \right. \tag{4.3}$$

where the sum includes all jets with a $E_T^{REC} > 20$ GeV. Ultimately, it is believed that the employment of energy flow objects, such as the replacement of charged track energy deposits by the measured track momentum, will significantly improve the E_T^{miss} reconstruction.

In Figure 4.24 and 4.25 respectively the relative bias and resolution on the E_T^{miss} measurement is plotted as a function of the true neutrino transverse energy, showing the influence of the mentioned E_T^{miss} corrections. For both plots semi-muonic decaying $t\bar{t}$-events are used. Before any correction the E_T^{miss} reconstruction is observed to be strongly underestimated, as is expected from the different sources of energy loss in the

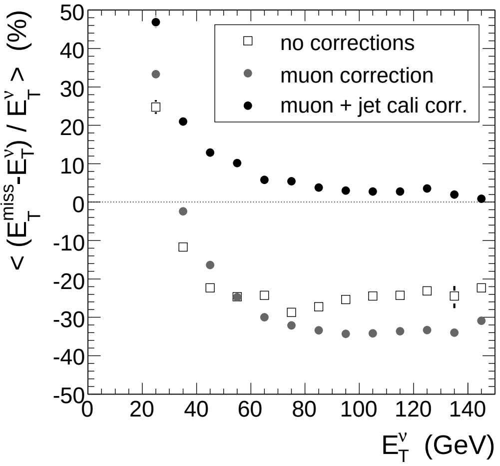


Figure 4.24: The bias on the E_T^{miss} measurement as a function of the true neutrino transverse energy.

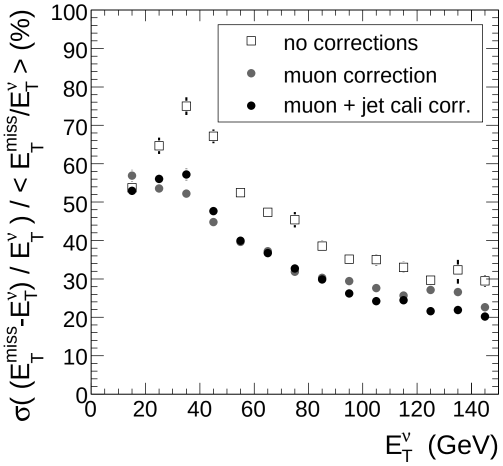


Figure 4.25: The relative resolution on the E_T^{miss} measurement as a function of the true neutrino transverse energy.

calorimeter response. Also the relative resolution, that is rescaled with respect to the average E_T^{miss}/E_T^{ν} response, indicates a very coarse E_T^{miss} reconstruction. It is shown that the addition of the reconstructed muon four-vectors in the calculation of E_T^{miss} mainly results in an improved resolution. The bias, in turn, is reduced impressively after applying the Monte-Carlo jet corrections. Eventually, for an average E_T^{ν}-value of 60 GeV produced in semi-leptonic $t\bar{t}$-events, the E_T^{miss} is on average overcalibrated with ∼6 GeV with a spread of 26 GeV. The reconstruction however improves for higher E_T neutrinos. For a neutrino with E_T^{ν} =120 GeV, an average bias of 2 GeV and a resolution of 30 GeV is found. It therefore can be assumed that the E_T^{miss} measurement in CMS is a highly valuable tool in the detection of high energetic sources of missing transverse energy like predicted in many theories beyond the Standard Model, but that the high uncertainties at moderate energies demand a cautious use of this information in the event reconstruction of for example top quark events.

4.4 Track & Vertex Reconstruction, and Impact Parameter Based *b*-tagging

The high luminosity proton-proton collisions expected at the LHC collider will have thousands of tracks in the acceptance of the tracking system. In order to cope with these conditions, track and vertex reconstruction rely on a system of silicon pixel and microstrip sensors, embedded in a solenoidal magnetic field of 4 Tesla (cfr. Section 3.2.1). The large granularity of the sensors and the high single point resolution translates into an excellent momentum resolution and a precise extrapolation of charged particle trajectories to the interaction region. This allows the reconstruction of primary and secondary vertices, based on the identification of the decays of long-lived particles.

In a first section we will describe the track reconstruction strategy implemented in the CMS software and illustrate the performance of these methods, while techniques used to fit the position of the primary vertex will be the subject of a second section. Finally in a third section, the possibility to use the track impact parameter information for the tagging of *b*-quark jets will be explored.

4.4.1 Track Reconstruction and Performance

Track reconstruction in a dense environment needs an efficient search for hits during the pattern recognition stage and a fast propagation of trajectory candidates. In the CMS tracker, these tasks benefit from the arrangement of the sensitive modules in practically hermetic layers as well as from the almost uniform 4 T magnetic field provided by the CMS solenoid. After the tracker hits have been reconstructed, track reconstruction proceeds through the three main steps: the seed finding, the pattern recognition and the final track fitting.

Seed generation

Seed generation provides the initial trajectory candidates for the full track reconstruction [124]. The seeds should already contain an estimate of all five track parameters [7] that is sufficiently close to the true values to allow the use of linear fitting algorithms. Also the uncertainties on these parameters need to be small enough to narrow the search region for hits. The standard trajectory seeds in the CMS tracker are constructed from pairs of hits, either in the pixel or in the silicon strip detector, compatible with the origin region [8]. Due to the low occupancy and the unambiguous 2-dimensional position information, the pixel layers provide the best seeding. However, mixing the information of both pixel and silicon strip detector as seeding input further improves the global seed generation efficiency in the endcap region due to a better coverage. The overall seed finding efficiency exceeds 99%.

Pattern recognition

The second step in the reconstruction of tracks, the pattern recognition [126], uses the first coarse estimate of the track seed parameters as an input to collect the full set of hits for a charged particle track. The algorithm is based on a combinatorial Kalman filter approach [127]: starting from the seed layer the trajectory is extrapolated outwards to the next layer and compatible hits are selected based on the χ^2 between the predicted and measured positions. With each included layer, the track parameters are updated and better constrained. In the extrapolation of the trajectory from one layer to another, the effects of energy loss and multiple scattering are accounted for. Trajectory candidates are added for each compatible hit, as well as for an unmeasured or 'lost hit' in order to cope with efficiencies in the hit reconstruction. The best trajectory candidates are grown in parallel up to the outermost layers.

---

[7] Usually these 5 track parameters are related to the inverse track momentum, the polar and azimuthal angles, and the transverse and longitudinal point of closest approach to the beam spot [125].

[8] The origin region is defined as a cylinder of radius 2 mm and half length 15 cm, positioned along the beam line at the centre of the detector.

Often ambiguities in track finding arise because a given track may be reconstructed starting from different seeds, or because a given seed may result in more than one trajectory candidate. To avoid double counting of tracks, a trajectory cleaning method is applied based on the fraction of hits that are shared between two trajectories.

Final track fitting

For each trajectory, the pattern recognition stage results in a collection of hits and an estimate of the track parameters. However, the full information is only available at the last hit of the trajectory, and the estimate may be biased by constraints such as the origin region applied during the seeding stage. Therefore the trajectory is refitted and smoothed using a least squares approach. This refitting procedure yields optimal estimates of the parameters at the surface associated with each hit and, especially, at the first and the last hit of the trajectory. Estimates on other surfaces, e.g. at the calorimeter surface, are then derived by extrapolation from the closest hit.

Performance

The global track reconstruction reaches an efficiency for muons higher than 98% over most of the tracker exceptance, and between 75 and 95% for hadrons, depending on the track momentum and the pseudorapidity [128]. The efficiency for hadrons is lower than for muons because hadrons interact with the material present in the tracker. In Figure 4.26 the average number of tracks reconstructed in semi-leptonic decaying $\mathrm{t\bar{t}}$-events is plotted as a function of the track transverse momentum. On average $\sim$100 tracks are reconstructed in these $\mathrm{t\bar{t}}$-events. A clear drop is observed around 0.9 GeV/c, reflecting the inefficient reconstruction of tracks that do not reach the outermost tracker layers due to the strong deflection by the magnetic field, but instead spiral around the beam axis towards the forward regions of the detector. As a consequence, a reduced average number of reconstructed hits is found for these low momentum tracks. This is illustrated in Figure 4.27, where clearly at the cutoff value of 0.9 GeV/c the average number of hits decreases significantly from 9 to 5. For all tracks, the final fitting procedure delivers a χ^2/ndf-value quantifying how well the reconstructed track parameters fit a set of reconstructed hits. The distribution of this normalized χ^2 is shown in Figure 4.28, both for tracks with and without a lost hit among the measured hits. Although the peak around unity for well reconstructed tracks is dominant, the distribution shows an important tail to high values. It is shown that the fraction of badly fitted tracks increases when allowing a lost hit. Finally, in Figure 4.29 the transverse track impact parameter, d_0, is plotted[9], showing a clear seperation between tracks coming from the beam spot area and tracks originating from sources showing a much larger displacement, such as V^0-decays, photon conversions, nuclear interactions in the beampipe or in the first layers of the pixel detector, ... Typically, most of these tracks are rejected when only considering tracks with $d_0 < 2\,\mathrm{mm}$.

---

[9] This transverse track impact parameter d_0 is defined as the closest distance of a track to the beam axis.

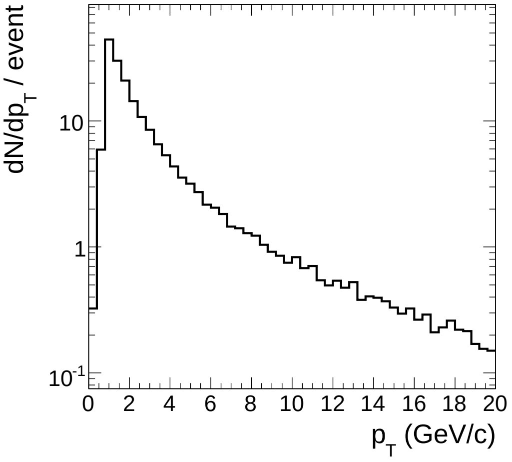


Figure 4.26: Average number of tracks reconstructed in a $t\bar{t}$-event as a function of the track transverse momentum.

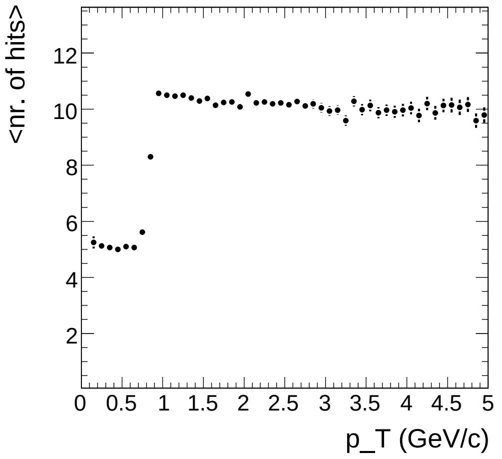


Figure 4.27: The average number of hits composing a track versus the track transverse momentum.

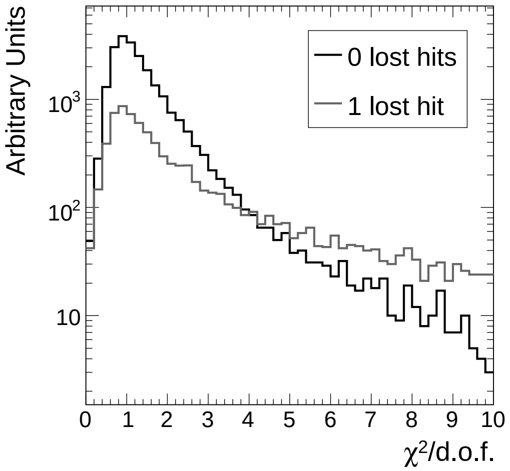


Figure 4.28: The normalised track fit χ^2-value, plotted for track with and without a unmeasured or lost hit.

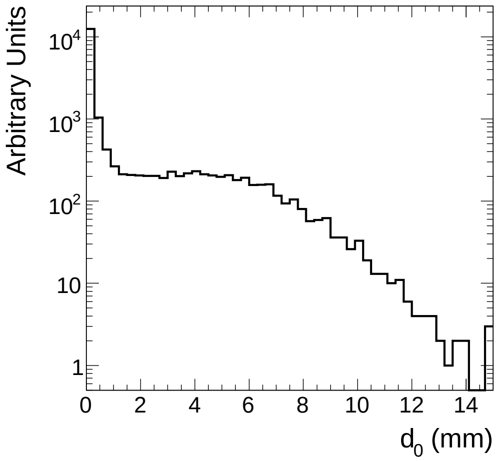


Figure 4.29: The track transverse impact parameter.

4.4.2 Primary Vertex Reconstruction

Once tracks are reconstructed, dedicated algorithms are applied to estimate the primary vertex position and its associated tracks. This vertex reconstruction typically involves two steps: vertex finding, where clusters of tracks originating from the same vertex are grouped together as vertex candidates, and vertex fitting, where, from a set of tracks, the most compatible vertex position is computed and used to constrain track parameters at the vertex.

Vertex finding

For fast primary vertex finding purposes, such as the online trigger system, pixel tracks from pixel hit triplets can be used to efficiently find the vertex positions along the beam line. In offline analyses however, where timing is not an issue, the full information of reconstructed tracks and their corresponding covariance matrices will be used. This offline method [124] starts with a preselection of tracks with a transverse impact parameter significance $d_0/\sigma(d_0) < 5$. This cut will reject secondary vertices and fake tracks, and reduces the computation time. In a second step, the selected tracks are extrapolated to the beam line, and grouped according to their separation in z in order to form primary vertex candidates. A maximum separation of 1 mm between two successive tracks belonging to the same primary vertex candidate is allowed. Each primary vertex candidate is then fitted, and all candidates with a χ^2-probability below 1% or a distance from the beam line larger than 200 μm are rejected. Eventually, the reconstructed primary vertices are sorted in decreasing order of 'hardness', defined by the scalar sum of all squared track transverse momenta, $\sum_{i=1}^{N} (p_T)_i^2$. Tracks with less than 5% compatibility to the leading vertex candidate or less than 1% to the other primary vertex candidates are rejected as tracks originating from a primary vertex, and are used as input to the secundary vertex finding algorithms. This primary vertex finding step is highly efficient in the case of semi-leptonic $\mathrm{t\bar{t}}$-events without pile-up: in more than 99% of the events, only one primary vertex candidate is found.

Vertex fitting

From a set of tracks associated to a given primary vertex advanced vertex fitting algorithms [129] will be used to compute the best estimate of the vertex parameters such as its position and covariance matrix, as well as indicators of the goodness of the fit like the χ^2/ndf-value or track weights. Vertex fitting usually consists in a least-square fit formulated as a Kalman filter. However, this method is only optimal when the uncertainties are Gaussian and the vertex candidate is not contaminated by mismeasured or mis-associated tracks (outliers). Since none of these conditions will hold for the real data at the LHC, robust statistical methods have been envisaged for vertex fitting in CMS, as the Adaptive Vertex Fitter [130] applied in this thesis. This algorithm is based on a re-weighted least-square fit, where outlying tracks are down-weighted during the iteration, enhancing the robustness of the method.

For $\mathrm{t\bar{t}}$-events, this fitting procedure results in an average primary vertex position uncertainty of 11 μm in both the transverse and longitudinal direction. The actual vertex position is smeared with a width of around 10 μm in the transverse plane, and 5 cm along the beam axis.

4.4.3 b-Jet identification using track impact parameters

In many physics processes such as events containing top quarks, b-quarks are produced in the final state. Especially at hadron colliders, these processes however suffer from copious background processes, in which light quark jets or jets resulting from hard gluon radiation in the initial or final state are responsible for a same number of high momentum jets as produced in the signal events. Therefore, it is very important to fully exploit the differences in the decay of b-quarks with respect to the light quarks or gluons, and to define techniques able to tag those jets origination from a b-quark decay.

These tagging algorithms mainly rely upon the large proper lifetime of the b-hadrons ($\tau_B \simeq 1.5\,ps\,,\ c\,\tau_B \simeq 450\mu m$) and their decay into final states with a high charged track multiplicity. On average five charged tracks are produced in the b-hadron decays, and in almost 20% of the cases, a b-hadron decays in an electron or a muon. Two main strategies exist to exploit this lifetime information: searching tracks with a large impact parameter [131], or looking for displaced secondary vertices [132]. We will concentrate on the first method, as this technique is applied in the further analyses. In this subsection we will describe and evaluate the use of two impact parameter based algorithms implemented in CMS: the track counting and track probability b-tagging method.

Both methods start with the association of a collection of tracks to each jet reconstructed in the event. For all jet-track pairs the ΔR-value is calculated between their corresponding directions evaluated at the vertex, and if this value is lower than a certain threshold value, the track is added to the list of matched tracks associated to the jet. When the iterative cone algorithm is used to cluster the jets, obviously this threshold value is taken equal to the cone radius. For cluster based algorithms or cone algorithms with a splitting/merging option implemented more advanced tracks association techniques need to be used.

In this jet-track association step, only tracks are considered that satisfy the following track quality cuts:

- transverse momentum $> 1\,\mathrm{GeV/c}$
- at least 8 reconstructed hits in total
- at least 2 reconstructed hits in the pixel detectors
- χ^2/ndf of the track fit < 5
- transverse impact parameter $< 2\mathrm{mm}$

In Figure 4.26-4.29 the distributions of all these track observables are plotted and discussed. Where the first three cuts are defined to exclude soft tracks curling around the beam axis, the fourth demand eliminates poorly fitted tracks. Finally, the last cut is applied to reject tracks from V^0-decays, photon conversions and nuclear interactions. For all selected tracks, a signed track impact parameter is calculated. This impact parameter (i.p.), illustrated in Figure 4.30, is defined as the distance of closest approach of a track to the primary vertex. Its sign is taken positive if the decay originating the track occurs downstream the jet, i.e. if the scalar product of the jet direction with the impact parameter direction is positive. Negative impact parameters arise most frequently as a result of a bad reconstructed primary vertex, or a finite detector resolution leading to mismeasured track parameters. To take into account the experimental resolution, in both the track counting and track probability b-tagging method the ratio between the impact parameter and its uncertainty, the i.p. significance, is used. This significance

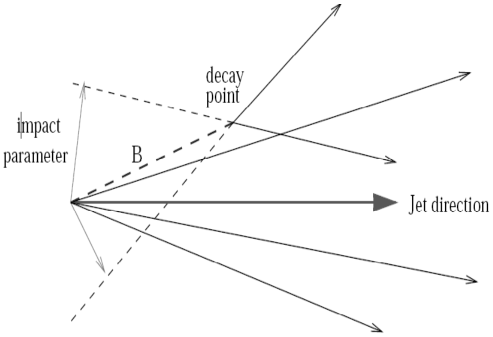


Figure 4.30: The decay of a b quark jet results in a displaced secondary vertex characterised by a set of charged tracks with a relatively large impact parameter with respect to the primary vertex.

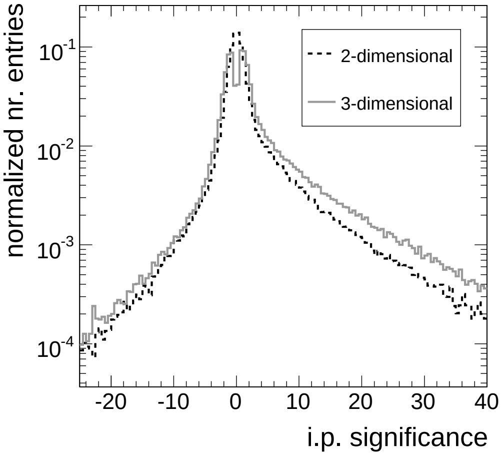


Figure 4.31: A higher fraction of high impact parameter significance tracks is observed when including the z-axis information of the impact parameter measurement. The dip for the 3D case at zero is due to phase-space.

can be computed both in the transverse plane or in three dimensions. Due to the small uncertainty on the primary vertex position along the beam axis ($\sigma_z \simeq 11\mu m$), the use of this extra information leads to more high significance tracks and hence an improved b-tagging performance. Consequently, the 3-dimensional impact parameter significance definition will be further used. This increased fraction of high significance tracks for the 3-D i.p. significance definition is clearly visible in Figure 4.31, where the 2-D and 3-D track i.p. significance is plotted for all selected tracks associated to jets in $t\bar{t}$-events that match a parton closer than 0.2 in the (η, ϕ)-space. The dip around zero in the 3-D i.p. significance distribution is due to phase-space.

Track counting b-tagging method

Originally this method is based on the simple requirement of a minimum number of selected tracks with a impact parameter significance exceeding a certain threshold. Currently, this implementation is replaced by a more flexible algorithm, that is able to deliver a continuous instead of a binary discriminator value, defined by the impact parameter significance of the n-th leading significance track. This discriminator is illustrated in Figure 4.32, where the impact parameter significance is plotted for both the first and the 4th leading track significance, using only $t\bar{t}$-jets with a well matching parton ($\Delta R < 0.2$) reconstructed in the barrel and with a transverse energy exceeding 10 GeV. In the case of light quark jets this significance distribution is observed to be almost symmetric, i.e. consistent with the hypothesis that all tracks originate from the primary vertex. The non-Gaussian tails in the distribution are due to particles in the jet with a rather long lifetime (e.g. K_s) or multiple scattering not correctly taken into account in the impact parameter uncertainty estimate. Tracks from the c-quark and

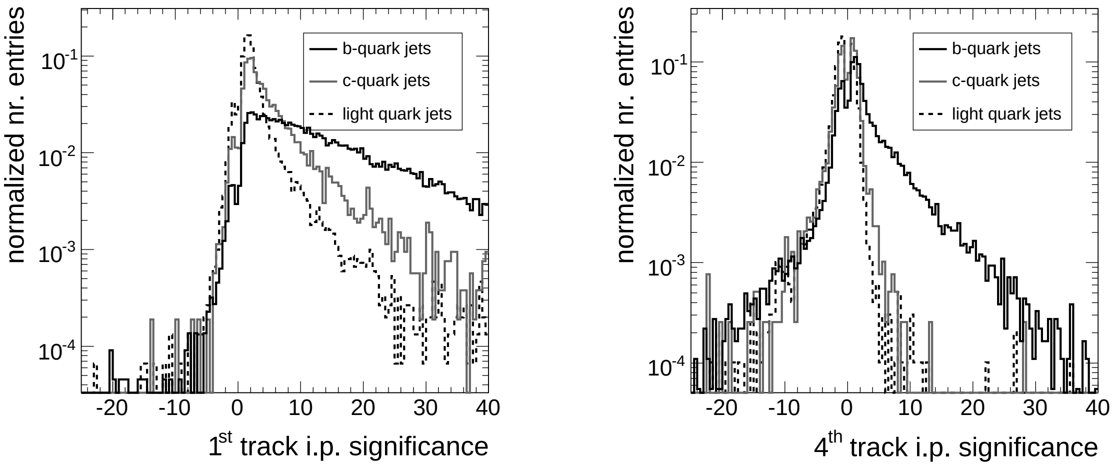


Figure 4.32: Comparison between the leading (left) and the 4th highest (right) track impact parameter significance for well matched $t\bar{t}$-jets.

especially the b-quark jets give an asymmetric distribution at positive values, reflecting the long lifetimes of B and charm hadrons. A clear gain in purity is observed when using the 4th track significance as discriminator. However, this choice would result in an important drop in the b-tagging efficiency, as only ~81% of the well-matched b-quark jets have more than three selected tracks associated. Typical recommended values for n are 2 or 3.

Track probability b-tagging method

In this method the compatibility of a set of selected tracks to all originate from the primary vertex is computed. If this probability is low, the jet is likely to be a b-quark jet. The algorithm works in several steps, starting with the construction of the resolution function $R(x)$, needed to extract the probability $P(S)$ for a track with parameter significance S to come from the primary vertex (eq. 4.4).

This resolution function is defined as a normalised i.p. significance distribution for primary vertex tracks. For this purpose, the negative part of the impact parameter distribution is used because it is mainly made up of primary vertex tracks, and is easy to obtain from the data.

$$P(S) = \mathrm{sgn}(S) \int_{|S|}^{\infty} R(x)\,\mathrm{d}x. \tag{4.4}$$

In order to enhance the discrimination power of the method, for several categories of tracks a dedicated resolution function is produced, depending on the p_T of the track, its pseudorapidity, number of hits, ...

In a second step, all track probabilities $P^i(S)$ are combined. For each jet the probability to have all N associated tracks produced at the primary vertex is defined as the confidence level that any set of N tracks without lifetime would produce the observed value of track probability or any other value that is more unlikely, given by the equation:

$$P_b^{jet} = \Pi \sum_{j=0}^{N-1} \frac{(-\ln \Pi)^j}{j!}, \qquad \text{with} \qquad \Pi = \prod_{i=1}^{N} \hat{P}^i(S), \tag{4.5}$$

where $\hat{P}^i(S) = P^i(S)/2$ in case the impact parameter significance S is positive, and $\hat{P}^i(S) = 1 + P^i(S)/2$ when S is negative, creating a track probability that is always positive. In Figure 4.33 the probability for a jet to have all tracks originating from the primary vertex is plotted for different jet flavours. By definition, this probability is high for light quark jets, while in the case of b-quark jets the presence of high quality tracks produced at a secondary vertex leads to less confidence in the hypothesis. However, the peak visible around $P_b^{jet} = 1$ in the b-quark jets spectrum indicates that these high i.p. significant tracks are not always found in the jet cone.

Performance of both algorithms in $\mathrm{t\bar{t}}$-events

To compare the performance of both methods, the b-tagging purity and the b-quark jet selection efficiency in semi-leptonic decaying $\mathrm{t\bar{t}}$-events are calculated for a given discriminant value c using the equations[10]:

$$\mathcal{P}(c) = \frac{f_b \int_c^\infty B(x)\,\mathrm{d}x}{f_b \int_c^\infty B(x)\,\mathrm{d}x \;+\; f_c \int_c^\infty C(x)\,\mathrm{d}x \;+\; f_l \int_c^\infty L(x)\,\mathrm{d}x} \tag{4.6}$$

$$\mathcal{E}(c) = \epsilon_{method} \int_c^\infty B(x)\,\mathrm{d}x, \tag{4.7}$$

with $B(x)$, $C(x)$ and $L(x)$ the b-tagging discriminant p.d.f.'s for respectively b-, c- and light quark jets. The $f_{b,c,l}$ factors represent the fraction of each jet flavour in semi-leptonic decaying $\mathrm{t\bar{t}}$-events, being respectively 50, 12 and 38%. Finally, ϵ_{method} accounts for the fraction of jets with enough tracks associated to apply the b-tagging method. This efficiency descreased for the track counting method from 93 to 89% when changing the b-tag discriminant definition from the second highest track i.p. significance to the third highest. The track probability method in turn is found applicable in 98% of the selected b-quark jets, as for the other jets no tracks are associated.

Figure 4.34 shows a comparison of the track counting and the track probability method. For the first algorithm, both the second and third highest impact parameter significance configuration are considered. A clear gain in purity is observed when using the second instead of the third track. For a selection efficiency of 70% a b-purity of 80% is achieved. Furthermore, the track probability method is shown to be slightly better than the track counting, and will be used in the further analyses.

4.5 Muon Reconstruction and Selection

A last type of physics object produced in the final state of our selected signal events are muons. Because of the relatively small energy loss of these particles in material,

---

[10] For the track probability b-tagging method the integral should be calculated between 0 and c due to the decreasing b-quark probability for higher discriminant values.

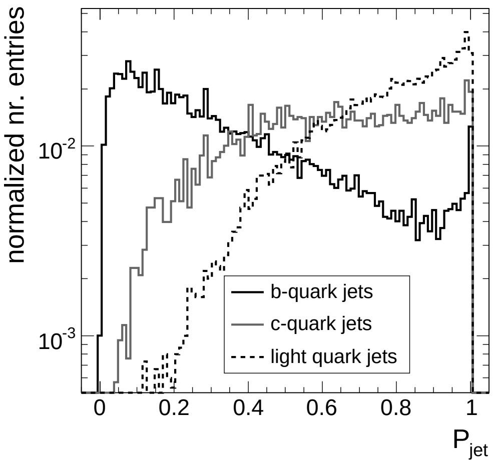


Figure 4.33: The probability that all tracks associated to a jet originate from the primary vertex, resulting from the track probability *b*-tagging method.

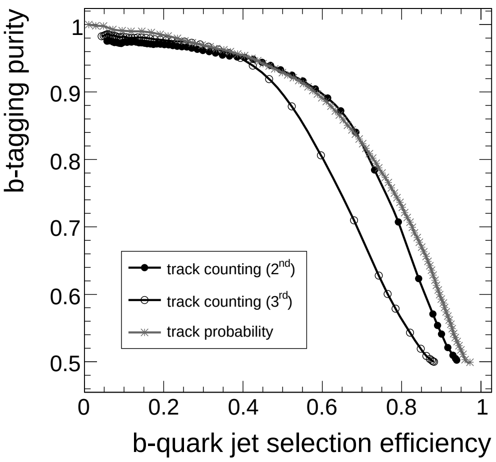


Figure 4.34: Comparison of two Track Counting configurations with the Track Probability *b*-tagging method.

extra outer sampling layers are constructed around the interaction point to detect the traversing muons and to help in the reconstruction of the their trajectories. In the measurement of the muon track parameters, the information of the muon spectrometer is combined with the reconstructed hits in the silicon tracker. The reconstruction of muons is performed in three stages: local reconstruction, standalone reconstruction and global reconstruction. Starting from a seed, the chambers compatible with the seed are identified and local reconstruction is performed only in these chambers (cfr. Section 3.2.4). The standalone reconstruction combines all information from the muon system, while global muon reconstruction also uses silicon tracker hits. In each step in the trajectory building, the Kalman filter technique is employed as described for the track reconstruction in Section 4.4.1, taking into account effects like the energy loss in material, multiple scattering, and the non-constant magnetic field in the muon system. Finally, it should be emphasised that the same software is applied in the online and offline muon reconstruction, except for the muon seed generation. Where in the online case the seeds are provided by the Level-1 Trigger system discussed in Section 3.2.6, segments reconstructed in the CSC and/or DT chambers are started from in the offline reconstruction.

4.5.1 Standalone Muon reconstruction

Mainly the track segments obtained after the local reconstruction in DT and CSC tracking detectors will be used as input to the standalone muon reconstruction. However, especially in the barrel-endcap overlap region where geometric coverage is problematic, RPC detector information is complementary. Starting from the track parameters estimated from segments reconstructed in the innermost muon chambers, an iterative search for matching measurements at outward surfaces is started. In the barrel chambers, complete reconstructed track segments are used as measurements in this Kalman filter procedure; in the endcap chambers, where the magnetic field is inhomogeneous, the individual reconstructed hits belonging to the track segments are used. At each step, the track parameters and the corresponding uncertainties are updated. This procedure continues until the outermost measurement surface of the muon system is reached. A constrained fit to the track parameters is then performed working from outside in, and the track parameters are defined at the innermost muon station. Finally, the track is extrapolated to the interaction region and a vertex constrained fit is applied.

4.5.2 Global Muon Reconstruction

Starting from the standalone reconstructed muon the muon trajectory is extrapolated from the innermost muon station to the outer tracker surface. Silicon layers compatible with the muon trajectory are then determined, and a region of interest within them is defined to perform regional track reconstruction. The determination of the region of interest is based on the uncertainties on the extrapolated standalone muon trajectory. Inside the region of interest, initial candidates for the muon trajectory (regional seeds) are built from pairs of reconstructed hits, and starting from these seeds a regional track is reconstructed following the procedure described in Subsection 4.4.1. The resulting trajectories are then refitted using only silicon tracker hits and the hits in the innermost muon station. The χ^2-probability of this fit is compared with the probability for the tracker only trajectory, in order to detect important muon Bremsstrahlung or any other kind of significant energy loss of the muon before the first muon station. This procedure is especially important for an accurate momentum measurement of high (TeV) p_T-muons.

4.5.3 Performance in semi-muonic $\mathrm{t\bar{t}}$ events

For isolated muons from W-boson decays in semi-leptonic $\mathrm{t\bar{t}}$-events that match the generated muon direction with a $\Delta R < 0.2$ at the vertex, the muon reconstruction efficiency is plotted versus the generated transverse momentum (Figure 4.35) and the generated pseudorapidity (Figure 4.36) for both the standalone and global muons. An increasing muon reconstruction efficiency is observed for higher p_T-values, which is explained by the decreasing fraction of muons suffering multiple scattering. As a function the generated muon pseudorapidity, an almost constant reconstruction efficiency of $\sim$95% is achieved in the $|\eta| < 2.4$ region. The drop of efficiency around $|\eta| = 0.25$ and $|\eta| = 1.2$ are respectively due to a transition between two DT wheels and between the DT and CSC systems. Furthermore, a very small gain in reconstruction efficiency is achieved, particularly visible in the forward η-regions. This increase in reconstruction efficiency for global muons is explained by beam constraint imposed only to the standalone muons. It is observed that in particular in forward regions a relative high fraction of the standalone muons fail this extra requirement, but are recuperated in the global muon reconstruction. Additionally, it should be remarked that the rather similar efficiency obtained for standalone and global muons is very important for the online selection strategy, as both the Level-1 and Level-2 Trigger decisions need to be made in time intervals too short to include tracker information in the muon reconstruction.

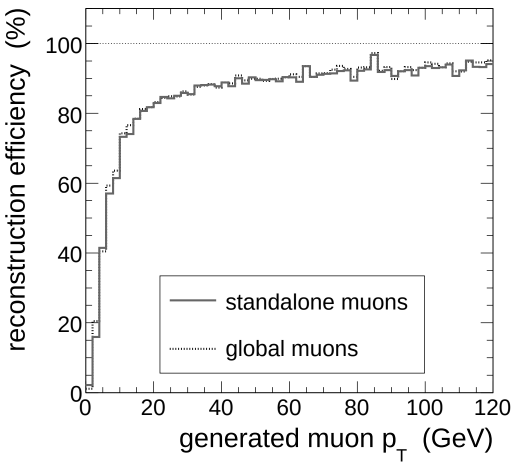


Figure 4.35: Reconstruction efficiency for the isolated muon from the W-boson decay in semi-leptonic $t\bar{t}$-events versus the generated muon transverse momentum.

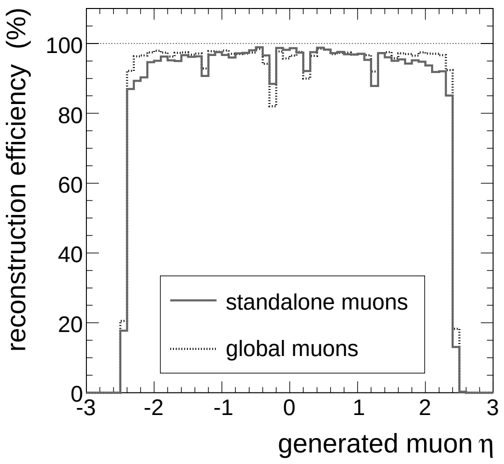


Figure 4.36: Reconstruction efficiency for the isolated muon from the W-boson decay in semi-leptonic $t\bar{t}$-events versus the generated muon pseudorapidity.

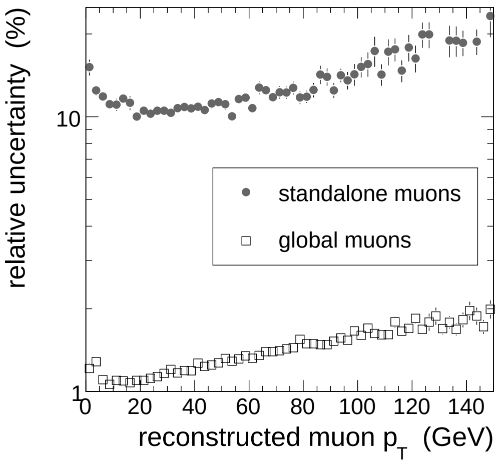


Figure 4.37: Relative uncertainty on the reconstructed muon transverse momentum versus the p_T of the muon.

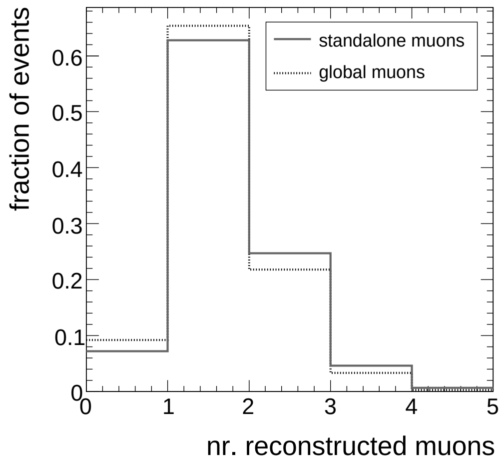


Figure 4.38: The number of muons reconstructed in semi-muonic decaying $t\bar{t}$-events.

The apparent difference in average muon reconstruction efficiency between Figure 4.35 and 4.36 is a result of very forward generated muons.

Figure 4.37 shows the relative uncertainty on the transverse muon momentum measurement versus the reconstructed muon p_T for both stages in the muon reconstruction. A significant improvement in precision (a factor of 10) is observed after inclusion of the tracker information. The standalone muon momentum precision is essentially determined by the measurement in the transverse plane of the muon bending angle at the exit of the solenoid, taking the interaction region as the origin of the muon. For muons with a $p_T < 200\,\mathrm{GeV/c}$ this measurement is however dominated by multiple scattering in the material before the first muon station, and the best momentum resolution is obtained from the track segment reconstructed in the silicon tracker. Additionally, the decrease in trajectory curvature for increasing p_T-values results in higher uncertainties in the momentum measurement.

Finally, in Figure 4.38 the number of muons reconstructed in semi-muonic decaying $\mathrm{t\bar{t}}$-events is illustrated. Although in more than 60% of the events exactly one muon is reconstructed, around 30% of the events contain multiple muons detected in the final state. These extra muons mainly arise from heavy quark decays and are expected to be part of the jet cone. Only isolated muons should therefore be accepted as a correct leptonic W-boson decay product.

4.6 Selection of the Reconstructed Top Quark Decay Objects

The generated final state topology of the semi-muonic $\mathrm{t\bar{t}}$ decay channel consists of four coloured partons of which two are heavy, a muon and a neutrino. The detected final state can therefore be characterised by four hadronic jets of which two originate from a b-quark, an isolated muon and transverse missing energy. In the previous sections we described the reconstruction aspects for all three types of physics objects detected by the CMS detector.

In this section the selection cuts will be introduced that are applied to the reconstructed physics objects. These cuts not only envisage a possible reconstruction of the total event decay topology, but should also enhance the fraction of signal events in the selected event sample. All analyses described in the next chapters will use only events in which the reconstructed objects fulfil the following requirements:

- **The event is accepted by the isolated single-muon HLT.**
  Before considering any other reconstructed object selection cut, the events are required to pass the the single-muon Level-1 trigger threshold and the corresponding HLT requirements as discussed in Section 3.2.6.

- **The triggered muon should have a $p_T > 20\,\mathrm{GeV/c}$ and should be reconstructed in the muon system acceptance ($|\eta| < 2.5$)**
  For all events passing the HLT, the reconstructed global muon candidate matching best the direction of the triggered muon is selected. For both signal events and several background contributions the transverse momentum spectrum of these selected muons is plotted in Figure 4.39. The probability for the highest p_T muon to have a reconstructed p_T close to the trigger threshold value of $19\,\mathrm{GeV/c}$ is shown to be very high for all event types. Therefore, the transverse momentum selection cut is chosen slightly higher than the HLT cut-off, in order to stay above the HLT turn-on curve, and still select most signal events. The few events with a transverse momentum below $19\,\mathrm{GeV/c}$ are due to events where no global muon is found corresponding to the HLT inducing muon, and a different global muon candidate is selected. The transverse momentum cut at $20\,\mathrm{GeV/c}$ will anyhow ensure that these rare events are rejected. Furthermore, the choice to consider the global muon that matches best the HLT muon as the decay product of the

leptonic W-boson decay, $W \rightarrow \mu\nu_\mu$ is correct in almost 99.7% of the triggered signal events. In [133] a method is developed to further enhance this purity based on a Likelihood Ratio Discriminant approach combining the discimination power of several observables. It is shown that in the case of muons, the purity of selecting the HLT muon as the correct lepton from the W-boson decay is similar to the purity reachable with the Likelihood Ratio method. In the study of the semi-electron $\mathrm{t\bar{t}}$-decay channel however, the use of the LR-method will result in an important gain in purity.

- **Minimum four jets should be reconstructed in the event containing at least 3 constituents and with a raw and calibrated jet E_T exceeding respectively 10 and 30 GeV. To allow b-tagging the jets are asked to be directed to the tracker region, and have minimum one associated track.**
  All jets in the presented analyses are clustered using the Iterative Cone algorithm with a jet cone radius of $R = 0.5$ and a minimum seed E_T equal to 1 GeV, as is motivated in Section 4.2. In order to allow for b-tagging, jets are asked to be reconstructed in the tracker acceptance and have minimum one associated track. This last cut, together with the requirement for a minimal E_T^{REC} of 10 GeV and the cut on the number of constituents allow an important rejection of many fake jets in the event. These can arise from the underlying event, pile-up collisions, electrons, ... In Figure 4.40 and Figure 4.41 respectively the reconstructed E_T and the number of calorimeter towers clustered in a jet is plotted. Only events from the exclusive `AlpGen` $\mathrm{t\bar{t}} + 0j$-sample, with all four semi-leptonic $\mathrm{t\bar{t}}$-jets matched to an initial quark closer than 0.2 in (θ, ϕ)-space, are selected. This to ensure that no jets originating from radiative processes are added to the fake jets distributions. It is shown that both the cuts on E_T^{REC} and the number of constituents in a jet are highly effective in the reduction of the number of fake jets in the event. The selected iterative cone jets are calibrated using the Monte-Carlo calibration technique described in Section 4.1.3, and a minimal jet E_T^{CAL} of 30 GeV is required. Lower E_T^{CAL} thresholds would result in a higher fraction of poorly reconstructed jets, as can be concluded from the jet resolution plots shown in Figure 4.16. An increase of this E_T^{CAL} cut-off value on the other hand would be effective to enhance the signal to background ratio, which is illustrated in Figure 4.42. However, such an altering of the E_T^{CAL} threshold would also result in an important reduction of the fraction of $\mathrm{t\bar{t}}0j$ signal events, demonstrated in Figure 4.43. Analogue, the probability to still have all four hadronic $\mathrm{t\bar{t}}$-decay products among the selected jets will decrease for a higher E_T^{CAL} cut-off value, what would spoil the possibility for a later event reconstruction. Hence, a rather low calibrated jet E_T threshold is advisable, the more because other techniques exist to reject the same background events, as for example b-tagging.

  After these jet selection cuts many events will still have more than four selected jets. In Section 5.1 the strategy is described to choose the four hadronic semi-muonic $\mathrm{t\bar{t}}$ decay products among these remaining jets.

- **The event should contain minimum two jets among the selected jets**

with a $P_b^{jet} < 50\%$.
Minimum two jets should be found among the selected jets in the event with a probability of having all associated tracks originating from the primary vertex less than 50%. The result of this cut is illustrated in Figure 4.44, where the distribution of the second lowest P_b^{jet}-value is shown for the different event types. The requirement to have two jets with a $P_b^{jet} < 50\%$ rejects many background events, and in particular the $W+jets$ events. For signal jets with a well matching parton, the threshold value of $P_b^{jet} = 50\%$ corresponds to a b-quark jet selection efficiency of ∼75% and a b-tagging purity of ∼80% (cfr. Figure 4.33).

The relative efficiencies of these cuts and the remaining number of events after the complete object selection are summarised in Table 4.3, for all considered signal and background samples and rescaled for an integrated luminosity of $1\,\mathrm{fb}^{-1}$. For all processes the LO cross sections were employed. Details on the simulation of these different Monte-Carlo samples are given in Section 3.2.9. The semi-muonic decaying $\mathrm{t\bar{t}}$ signal events are observed to pass the isolated single muon HLT stream in around 62% of the cases. For the other $\mathrm{t\bar{t}}$-decays ∼6% survived the trigger cuts, being mainly semi-leptonic tau decays and fully leptonic muon decays. For the W+jet events an efficiency around 18% is observed for the samples containing no heavy quark jets. This high fraction of selected events compared to the 6% obtained in the Wbb+jet samples is because in the first set of samples the W-boson is only allowed to decay leptonically. After all muon cuts, the signal-to-background (S/B) ratio increased with a factor of 6.

The event selection efficiency after the requirement to find minimum four energetic jets is, as could be expected, very dependent on the jet multiplicity in the generated `AlpGen` subsample. Where around 70% of the signal $\mathrm{t\bar{t}}0j$ events that pass the muon criteria are rejected after the jet cut, this fraction is only 4% in the case of the signal $\mathrm{t\bar{t}}4j$ events. Hence, many signal events will be selected for further analysis due to the presence of high energetic initial or final state radiation jets. As a consequence, the fraction of selected signal events with the two correct b-quark jets among the selected jets will be smaller for the high jet multiplicity bins. This is clearly visible from the efficiencies obtained after the cut on the P_b^{jet}-value of the jets. The requirement to have minimum two jets with a high probability to originate from the decay of a b-quark is more stringent for the high multiplicity subsamples, indicating that in these bins more often jets different from the correct b-quark jets are selected after the previous jet E_T-cut. Furthermore, the b-tagging requirement is observed to be very efficient in the rejection of $W + jets$ events, as is also demonstrated by Figure 4.44.

After all four requirements more than twice as many signal events are selected with respect to background events. This corresponds with an improvement of the signal-to-background ratio with a factor of ∼23. When other $\mathrm{t\bar{t}}$-decays are considered top quark signal, a S/B-ratio of 11.2 is obtained. In the next chapter, methods to exploit the different kinematics and topology between the signal and background events will be introduced, resulting in a further enhancement of this signal-to-background ratio.

The effect of the described reconstructed object selection cuts on the composition of respectively the signal and background event sample is visualised in Figure 4.45 and Figure 4.46. Both before and after the object selection, the relative fraction of the

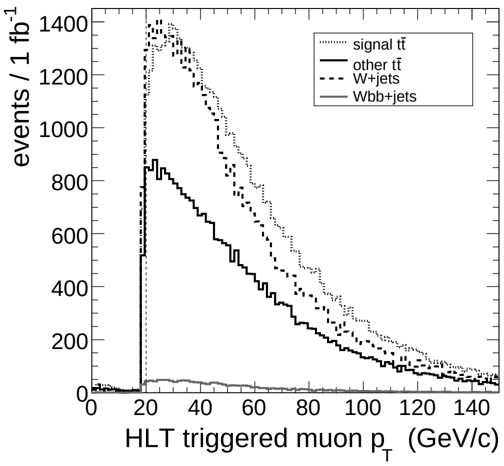


Figure 4.39: The reconstructed transverse momentum for the global muons matching best the HLT muons. The applied cut-value of 20 GeV/c is indicated.

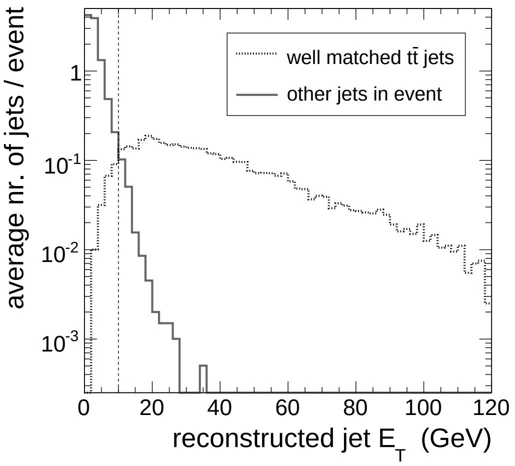


Figure 4.40: The reconstructed transverse energy for correctly matched $t\bar{t}$-jets and fake jets, selected from the `AlpGen` $t\bar{t}+0j$-event sample.

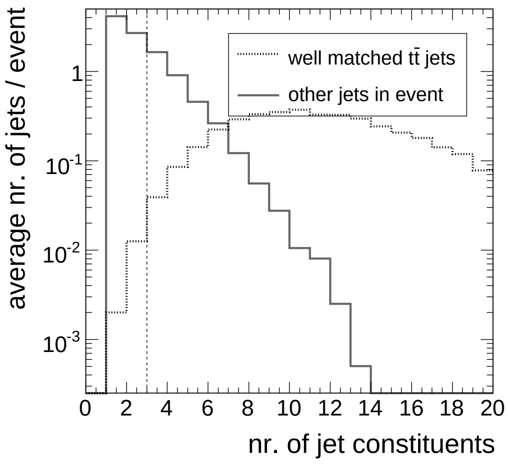


Figure 4.41: The number of constituents in correctly matched $t\bar{t}$-jets and in fake jets, selected from the `AlpGen` $t\bar{t} + 0j$-event sample.

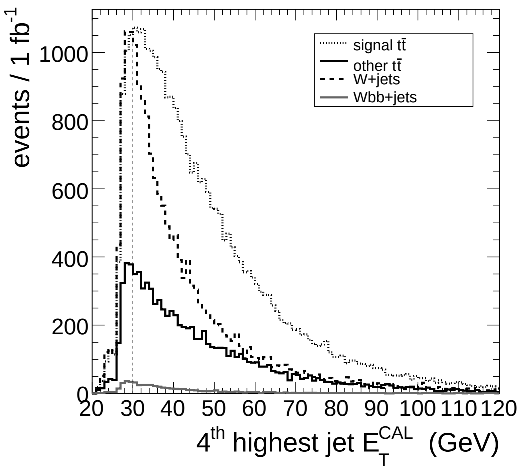


Figure 4.42: The reconstructed transverse energy for the 4^{th} highest E_T jet for all events passing the muon cuts. The applied cut-value of 30 GeV is indicated.

different signal and background event contributions are plotted. For the signal events, a clear shift towards high jet multiplicity `AlpGen` subsamples is observed due to the higher efficiency to pass the calibrated jet E_T requirement. Looking at the background composition, it is shown that the reconstructed object selection cuts are very efficient to reduce the fully hadronic and the semi-electron $t\bar{t}$-decays. On the contrary, the relative fraction of fully leptonic $t\bar{t}$-decays increases significantly after the selection cuts. Where in the initially generated samples, these events only contributed for 8.3% of the events, fully leptonic $t\bar{t}$-decays become the leading fraction (42.3%) in the selected channel background event sample. Furthermore, the reduction of the fully hadronic and

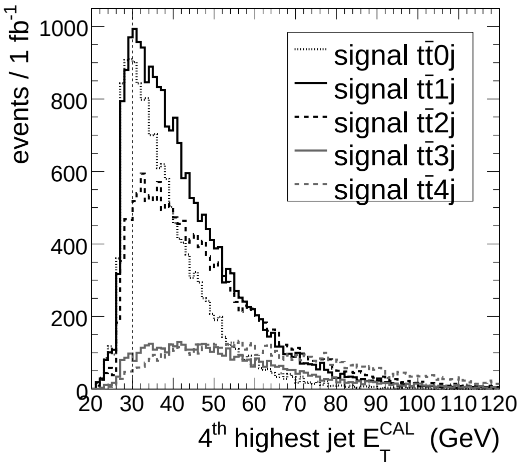


Figure 4.43: The reconstructed transverse energy for the 4^{th} highest E_T jet for all signal events passing all muon cuts. The applied cut-value of 30 GeV is indicated.

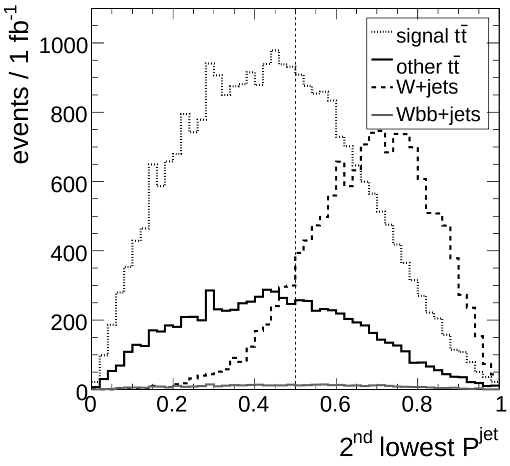


Figure 4.44: The second lowest P_b^{jet}-value among the selected jets, with P_b^{jet} defined in Eq. 4.5. The applied cut-value of 50% is indicated.

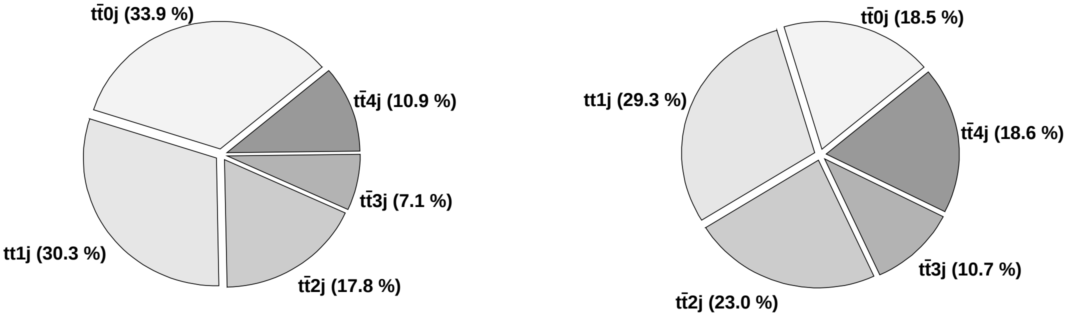


Figure 4.45: The composition of the semi-muonic signal event sample before (left) and after (right) the physics object selection.

semi-electron tt̄-decays is balanced by an increased fraction of W+jets events (39.7%) and semi-tau tt̄-decays (14.3%).

Apart from the W+jets, Wbb+jets and other tt̄-decays channel backgrounds, also the efficiency of this reconstructed physics object selection should be investigated with respect to QCD events. However, given the huge QCD production cross-section and the relatively small available Monte-Carlo simulated event samples[11] a precise estimate of this efficiency is not possible, as no event is left after all physics object selection cuts. Nevertheless, if we assume the muon related selection cuts are completely uncorrelated to the requirements posed on the jets, a rough estimate of the QCD selection efficiencies

---

[11] For a list of all analysed QCD subsamples with their respective cross section and number of events we refer to Table 3.2.

	#evts.	L1+HLT	muon p_T	$\geq$ 4 jets	2 b-jets	sel. #evts
signal						
$\mathrm{t\bar{t}}0j$ (excl)	28.1 k	62.4 %	60.2 %	17.3 %	11.3 %	(3.17±0.03) k
$\mathrm{t\bar{t}}1j$ (excl)	25.2 k	63.5 %	61.1 %	32.1 %	19.3 %	(4.86±0.04) k
$\mathrm{t\bar{t}}2j$ (excl)	14.8 k	62.6 %	60.6 %	43.5 %	25.2 %	(3.73±0.03) k
$\mathrm{t\bar{t}}3j$ (excl)	5.9 k	62.3 %	60.1 %	50.2 %	28.7 %	(1.70±0.03) k
$\mathrm{t\bar{t}}4j$ (incl)	9.0 k	60.0 %	58.3 %	55.5 %	32.5 %	(2.98±0.04) k
incl. semi-μ	83 k	62.5 %	60.3 %	33.0 %	19.8 %	(16.43±0.08) k
background						
$\mathrm{t\bar{t}}0j$ (excl)	162 k	5.98 %	5.74 %	0.28 %	0.18 %	(0.29±0.01) k
$\mathrm{t\bar{t}}1j$ (excl)	145 k	6.17 %	5.92 %	1.03 %	0.63 %	(0.92±0.02) k
$\mathrm{t\bar{t}}2j$ (excl)	85.0 k	6.06 %	5.83 %	2.21 %	1.34 %	(1.14±0.02) k
$\mathrm{t\bar{t}}3j$ (excl)	34.1 k	6.20 %	5.95 %	3.45 %	2.02 %	(0.69±0.02) k
$\mathrm{t\bar{t}}4j$ (incl)	52.0 k	6.03 %	5.80 %	5.11 %	2.95 %	(1.53±0.03) k
incl. other-$\mathrm{t\bar{t}}$	478 k	6.07 %	5.83 %	1.60 %	0.96 %	(4.57±0.05) k
$W4j$ (excl)	174 k	18.7 %	17.9 %	3.72 %	0.34 %	(0.60±0.04) k
$W5j$ (excl)	45.0 k	18.3 %	17.6 %	8.09 %	0.90 %	(0.41±0.02) k
$W6j$ (incl)	31.0 k	17.3 %	16.6 %	12.6 %	2.10 %	(0.65±0.03) k
$Wbb2j$ (excl)	17.0 k	6.35 %	6.08 %	1.31 %	0.75 %	(0.120±0.004) k
$Wbb3j$ (excl)	7.0 k	6.16 %	5.87 %	2.57 %	1.22 %	(0.085±0.005) k
incl. $W(bb)$	274 k	17.4 %	16.7 %	5.27 %	0.68 %	(1.86±0.05) k
S/B	0.11	0.68	0.68	1.24	2.56	2.56±0.03

Table 4.3: Overview of efficiencies obtained after the different object selection cuts applied to the signal and background event samples. The indicated number of events before and after the cuts are rescaled to an integrated luminosity of $\mathcal{L}^{int}$ =1 fb^{-1} using LO cross-sections. The notation 'signal $\mathrm{t\bar{t}}nj$' refers to semi-muonic decaying $\mathrm{t\bar{t}}$-events, while all the other $\mathrm{t\bar{t}}$-decays are considered as 'background $\mathrm{t\bar{t}}nj$'.

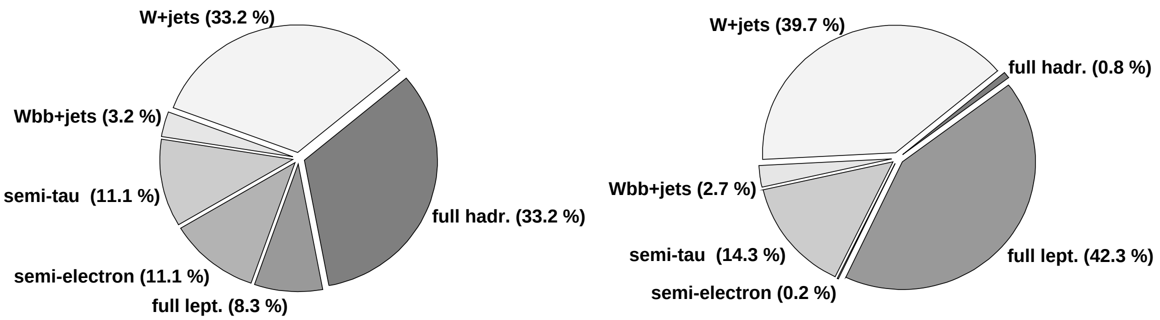


Figure 4.46: The relative fractions of the background event contributions before (left) and after (right) the physics object selection. In the simulation of the W+jet events the W-boson was only allowed to decay leptonically.

can be calculated. In this case the total efficiency ϵ_{tot} is estimated as:

$$\epsilon_{tot} = \epsilon_{muon} \times \epsilon_{(b)-jets}, \tag{4.8}$$

with ϵ_{muon} the selection efficiency to pass both the HLT and the addition muon requirements and $\epsilon_{(b)-jets}$ the event selection efficiency after the jets and b-tagging requirements.

The resulting efficiencies for the different QCD samples are listed in Table 4.4. It is observed that the muon requirements are by far the most efficient cut in the rejection of QCD events. For the low and very high $\hat{p_T}$-bins not even one event passed the muon cuts. For these bins, the analysed statistics only allows to state that less than 1 event is selected, which results in the mentioned lower bound on the selection efficiencies. The muon candidates in the few selected QCD events are in general real muons produced in the fragmentation of a heavy flavour quark. The fraction of QCD events with minimum four high-E_T jets reconstructed in the final state plus two jets fulfilling the b-tagging requirements increases significantly for higher $\hat{p_T}$-bins. This is expected due to the increasing transverse energy of both QCD jets, and the higher probability for hard initial and final state gluon radiation to occur in the event. This increasing number of jets reconstructed in the event will also enhance the probability to find two jets in an event that pass the b-tagging requirement. After combining the muon and (b-quark) jets selection efficiencies using Eq. 4.8, an estimate for a total selection efficiency for each of the QCD subsamples is listed. Compared to the results obtained for the $t\bar{t}$ and $W(bb)$+jets events in Table 4.3, very small fractions of the QCD events will survive all requirements. In the last column finally, an estimation of the number of selected QCD events for $1\,\text{fb}^{-1}$ of data is given. The increasing selection efficiencies for the higher $\hat{p_T}$-bins are now compensated with a decreasing production cross section. Due to this effect, only in $\hat{p_T}$-bins between 30 and 230 GeV more than 100 events per bin are expected for $1\,\text{fb}^{-1}$. Summing all contributions results in a total of 5.4k QCD events. Comparing this number to the results in Table 4.3 learns a factor three more QCD events are expected to pass all cuts compared to the $W(bb)$ event rate.

$\hat{p_T}$-bin	$\epsilon_{muon}(\%)$	$\epsilon_{jet+b-tag}(\%)$	$\epsilon_{tot}(\%)$	sel. evts/fb^{-1}
15-20 GeV/c	$<(2\pm2)\,10^{-3}$	$<0.002\pm0.002$	$<(3\pm5)\,10^{-8}$	$<<(5\pm7)\,10^{-2}$
20-30 GeV/c	$<(1\pm1)\,10^{-3}$	0.001 ± 0.001	$<(1\pm1)\,10^{-8}$	$<(7\pm10)\,10^{1}$
30-50 GeV/c	$(5\pm5)\,10^{-4}$	0.024 ± 0.004	$(1\pm1)\,10^{-7}$	$(2\pm2)\,10^{2}$
50-80 GeV/c	$(3\pm1)\,10^{-3}$	0.21 ± 0.01	$(5\pm3)\,10^{-6}$	$(11\pm7)\,10^{2}$
80-120 GeV/c	$(10\pm5)\,10^{-3}$	0.93 ± 0.05	$(9\pm5)\,10^{-5}$	$(3\pm1)\,10^{3}$
120-170 GeV/c	$(7\pm5)\,10^{-3}$	2.00 ± 0.08	$(2\pm1)\,10^{-4}$	$(7\pm5)\,10^{2}$
170-230 GeV/c	$(4\pm4)\,10^{-3}$	3.4 ± 0.1	$(1\pm1)\,10^{-4}$	$(1\pm1)\,10^{2}$
230-300 GeV/c	$(2\pm2)\,10^{-3}$	4.47 ± 0.09	$(8\pm8)\,10^{-5}$	$(8\pm8)\,10^{1}$
300-380 GeV/c	$(2\pm2)\,10^{-3}$	5.0 ± 0.1	$(1\pm1)\,10^{-4}$	$(1\pm1)\,10^{2}$
380-470 GeV/c	$(3\pm3)\,10^{-3}$	5.8 ± 0.1	$(2\pm2)\,10^{-4}$	$(4\pm4)\,10^{0}$
470-600 GeV/c	$(1\pm1)\,10^{-3}$	6.12 ± 0.08	$(7\pm7)\,10^{-5}$	$(5\pm5)\,10^{-1}$
600-800 GeV/c	$<(3\pm3)\,10^{-2}$	6.3 ± 0.4	$<(2\pm2)\,10^{-3}$	$<(4\pm4)\,10^{0}$
800-1000 GeV/c	$<(3\pm3)\,10^{-3}$	6.7 ± 0.1	$<(2\pm2)\,10^{-4}$	$<(8\pm8)\,10^{-2}$
			$\sum$	5.4k

Table 4.4: For all analysed QCD event samples, the selection efficiencies for the different cuts defined above. The HLT requirement and the additional muon cuts are merged because of their important mutual correlation.

Due to the limited Monte-Carlo statistics and the extremely low selection efficiencies obtained for all scanned QCD $\hat{p_T}$-bins, QCD channel background events will not be directly considered in the setup and definition of the analyses presented in the next chapters.

Chapter 5

Data-driven Light-Quark Jet Energy Scale Calibration

In the complex environment of proton-proton collisions at the Large Hadron Collider it will be challenging to calibrate the absolute energy scale of hadronic jets in the final state. However, in the search for new physics or when aiming for precision measurements, the knowledge of the absolute energy scale of the reconstructed jets originating from the quarks in the high-Q^2 collision is crucial. Although accurate estimates of the JES calibration factors can be obtained from adequate Monte-Carlo simulated event studies, in general these studies suffer from important systematic uncertainties arising from the uncertainties in the ISR/FSR rate, the fragmentation and underlying event model, pile-up collisions, detector inefficiencies,... Therefore, if possible, data-driven strategies are preferred to directly measure the JES correction factors, or as a closure test of the already available calibration factors. In order to extract accurate jet correction factors from the data, the complex issue of jet calibration is factorised in separate levels. The resulting calibration chain was discussed in Section 4.1.3, and examples were given of physics processes to be employed for the measurement of the different calibration factors.

The huge $\mathrm{t\bar{t}}$ statistics expected at the LHC not only allows for an improvement in the precision of several top quark properties, but will also create a possibility to use the top quark events for commissioning and calibration tasks. Already in both the CDF and DØ experiment, semi-leptonically decaying $\mathrm{t\bar{t}}$-events were selected to simultaneously measure the top quark mass and the calorimeter jet energy scale for light-quark jets [134, 135]. However, these analyses emphasise on a more optimal employment of the available statistical and kinematic information to reduce the total systematic uncertainty on the top mass measurement, instead of a more general extraction of calibration constants.

In this chapter, a study will be presented that allows the estimation of an inclusive light-quark jet energy scale (JES) correction factor similar to [116]. As an input, semi-muonic decaying $\mathrm{t\bar{t}}$-events are used because of the large S/B ratio achievable after the event selection. Instead of using the reconstructed four-momenta for a precision measurement, one can invert the process and determine via the precisely known W-boson mass the absolute energy scale miscalibration of the reconstructed jets in the observed $W \rightarrow q\bar{q}$ decay. Consequently, the reconstruction of the hadronic W-boson mass will be the starting point in this analysis. This will be subject of Section 5.1. The method developed to estimate the inclusive JES calibration factor will be described in Section 5.2. Section 5.3 finally will summarise the robustness of the JES correction estimate with respect to the main systematic uncertainties.

In terms of the jet calibration levels introduced in Section 4.1.3, the correction factor obtained via this analysis can be seen as a combination of Level-4 and Level-6: a jet flavour correction up to the parton level. Hence, a factor to be applied after the offset, relative η and absolute p_T correction. However, with enough statistics accumulated, the current estimation of the inclusive correction factor might be extended to a measurement depending on the jet's p_T or pseudorapidity.

5.1 Reconstructing the hadronic W-boson mass

In the last section of the previous chapter, the cuts on the reconstructed physics objects were described that defined an event to be accepted or rejected for further analysis. In order to be able to fully reconstruct the semi-muonic $\mathrm{t\bar{t}}$-decay, several requirements on the kinematics and characteristics of both muon and jets had to be fulfilled. This reconstructed objects selection resulted in an important increase of the signal to channel background ratio with a factor of 24. In order to reconstruct the four-momentum of the hadronically decaying W-boson, the two resulting light quark jets need to be identified among the reconstructed jets in the events. After the jet selection cuts, many events still have more than four jets reconstructed and selected in the final state, where at tree level only four hadronic decay products are expected for the studied single-muon signal event topology. Therefore, the following two step approach was applied to select the signal jets [1]:

1. The two jets with the minimal P_b^{jet}-value are selected as hadronic and leptonic b-quark candidate.
2. Among the remaining jets the two jets with leading E_T are searched and selected as the light quark candidates resulting from the hadronic W-boson decay.

For all jets assumed to be the light quark jets originating from the W-boson decay, the invariant W-mass spectrum is plotted in Figure 5.1 for an integrated luminosity rescaled to $1\,\mathrm{fb}^{-1}$. All events passing the reconstructed objects selection cuts summarised in Table 4.3 were included. The plot shows a differentiation between events with respectively correct or wrong reconstructed light quark jets, and channel background events. A hadronic W-mass contribution was defined correctly reconstructed if for both selected light quark jet candidates a well matching light quark was found originating from the W-boson decay. This was done using an angular criterion of $\alpha_{jp} < 0.2$ rad, where α_{jp} reflects the angle between the jet momentum and the parton momentum vectors in the (θ, ϕ) metric. All semi-leptonic $\mathrm{t\bar{t}}$-decays were considered as signal events, as they all contain a hadronic W-boson decay. As a consequence, the events considered as channel background contributions in Figure 5.1 are the result of either

---

[1] This procedure to select the hadronic semi-muonic $\mathrm{t\bar{t}}$-decay products is applied throughout the further presented work.

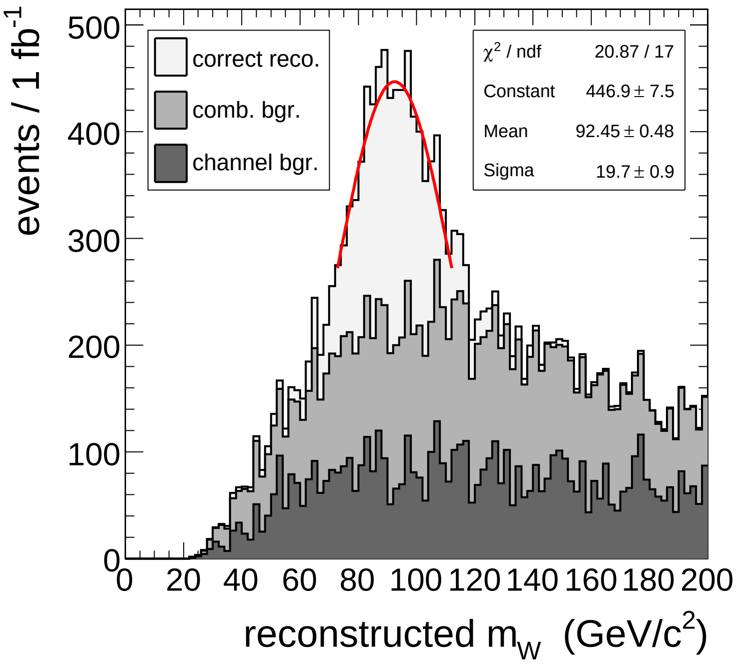


Figure 5.1: The reconstructed W-boson mass spectrum. A differentiation between correct reconstructions, combinatorial and channel background has been made. The histogram is rescaled for an integrated luminosity of $1\,\mathrm{fb}^{-1}$.

fully leptonic or fully-hadronic $t\bar{t}$-decays or W(bb)+jet events. In order to account for the number of selected QCD events, the contribution of the $W(bb)$ events is multiplied with a scaling factor of four. This factor is motivated by the expected QCD event rate left after the physics object selection requirements described in Section 4.6. For $1\,\mathrm{fb}^{-1}$ of accumulated data, three times more QCD events are expected to survive the selection cuts compared to $W(bb)$-events. Also, in both the W-boson and QCD events the two light jet candidates are produced in radiative processes and most QCD events are expected for $\hat{p_T}$-values between 50 and 120 GeV/c and hence at scales comparable to the W-boson production. Because of these analogies, the same reconstructed m_W shape is assumed for both processes. After this rescale, a channel background event fraction of 24% was obtained in a reconstructed W-boson mass window between 70 and 120 GeV/c^2. In this m_W-interval most of the correctly reconstructed m_W-contributions are found. Also, it should be remarked that the remaining channel background distribution is almost flat in the considered mass region. Therefore, it it expected that the systematic uncertainty on the gaussian fitted W-boson mass spectrum due to uncertainties on the channel background level will be negligible. In Section 5.3 we will come back to this. Additionally, in more than 52% of the signal events the correct jets where selected and well reconstructed in the same W-boson mass window. These correct m_W-reconstructions are clearly visible in the inclusive W-mass spectrum by the appearance of a nice gaussian peak in the region of 90 GeV/c^2. Around the peak of the invariant W-mass spectrum, a simple Gaussian function $G(m_W)$ is fitted and the mean value m_W is taken as an estimate of the reconstructed W-boson mass. This value is found to be $m_W = 92.45 \pm 0.48\,\mathrm{GeV/c^2}$, which is much higher than the world average, $M_W = 80.403 \pm 0.029\,\mathrm{GeV/c^2}$ [3]. This discrepancy was expected because of the overestimated light quark jet energy scale after the Monte-Carlo jet calibration. This behaviour was discussed in Subsection 4.1.3 and illustrated in Figure 4.9. The overestimated JES will be by far the dominant reason for the large bias between the reconstructed and true W-boson mass. This can be concluded from the much worse relative resolution on the jet energy measurement (Figure 4.13) compared to the uncertainties on the jet direction (Figure 4.16). Furthermore, it should be noted that the fitted m_W-value can be obtained with a statistical precision better than 1% with only $1\,\mathrm{fb}^{-1}$ of accumulated data. Also, no advanced signal event selection cuts where applied to improve the signal to background ratio.

5.2 Estimator for the Absolute Light Quark JES

It is possible to actually correct for the light-quark JES overcalibration via the determination of the relative light quark jet energy correction factor that results in a m_W-measurement in agreement with the world average. More concrete, the W-boson mass spectrum will be constructed from jets which obtain an extra relative calibration shift ΔC (in %) on their Monte-Carlo calibrated jet energy scale E^{CAL}:

$$E^{\Delta C} = E^{CAL}\,(1 + \Delta C). \tag{5.1}$$

The direction of the jet is kept invariant during this rescaling, while the magnitude of its momentum is rescaled in order to keep the $E/|\vec{p}|$ ratio of the jet invariant. For ΔC values varying between -40 to 40%, the reconstructed W-boson mass spectrum is illustrated in Figure 5.2. Each of the inclusive histograms was fitted with a gaussian function $G_i(m_W^{\Delta C_i})$. The fit range is chosen that value between 1σ and 2σ which results in a minimal $(\chi^2/ndf - 1)$-value.

This procedure results in 21 $m_W^{\Delta C_i}$ values corresponding to the 21 ΔC corrections. The linear relation between both is demonstrated in Figure 5.3. The value of ΔC which results in a fitted W-boson mass, m_W, in agreement with the precise world average M_W is the best estimate of the correction needed on the absolute light quark jet energy scale. Using this method on the initially Monte-Carlo calibrated jets results in a best value of ΔC = -13.60 ± 0.46%. The uncertainty on ΔC is estimated from the uncertainty on the fitted W-boson mass and the uncertainty on the slope of the curve shown in Figure 5.3.

An essential test of this ΔC-estimator is to check its bias with respect to the Monte-Carlo true best relative correction factor. From the selected signal events one can identify those light quark jets which match well to the Monte Carlo simulated parton ($\alpha_{jp} < 0.2\,\mathrm{rad}$). For these jets the Monte-Carlo calibrated energy, E^{CAL}, is compared to the generated parton energy, E^{GEN}. Figure 5.4 shows the distribution of E^{GEN}/E^{CAL} for the uncorrected Monte-Carlo calibrated and well matched light quark jets. The Gaussian fitted mean value of the (E^{GEN}/E^{CAL} - 1) distribution is taken as the true energy scale correction factor. Its value is found to be ΔC^{true}= -13.55 ± 0.22%, and has to be compared with the estimated correction of ΔC = -13.60 ± 0.46%, resulting

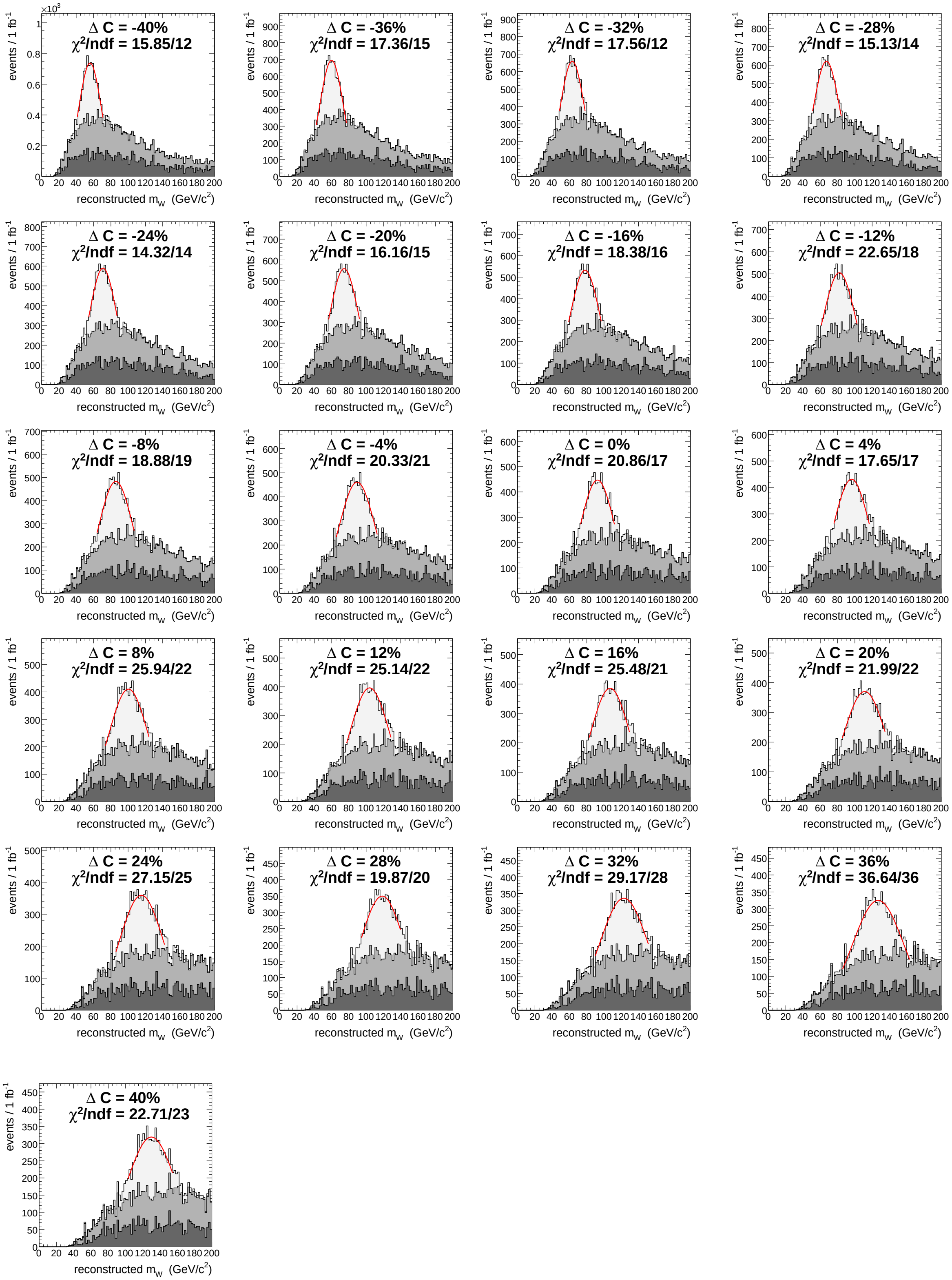


Figure 5.2: Reconstructed W-boson mass spectra for relative light-quark jet energy scale shifts ΔC varying between -40 and 40%.

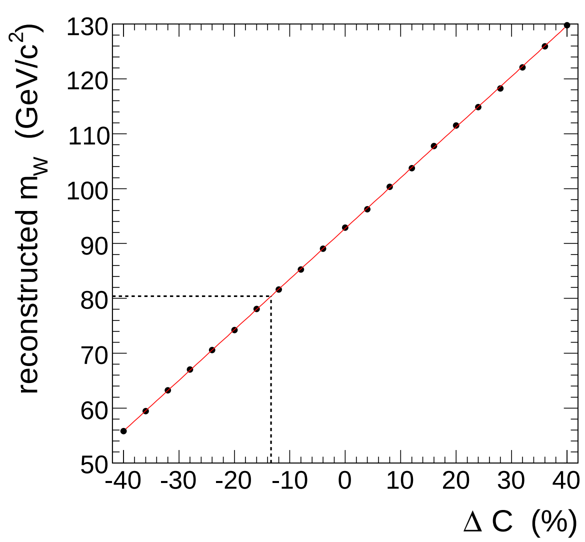


Figure 5.3: The dependency of the W boson mass estimator m_W and the applied extra jet energy scale correction ΔC.

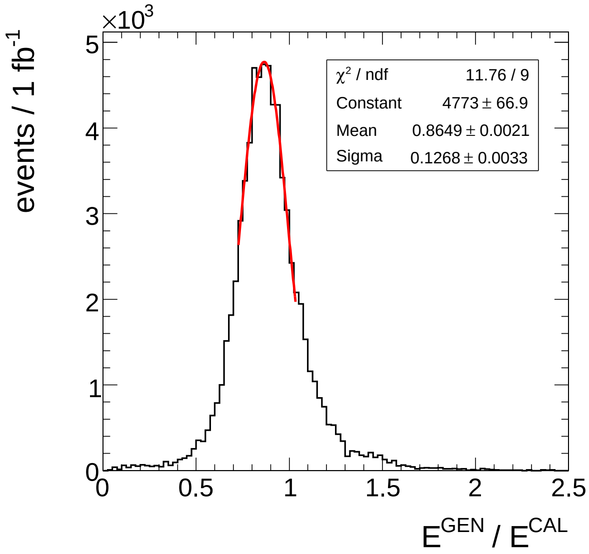


Figure 5.4: Distribution of E^{GEN}/E^{CAL} used to determine ΔC^{true}. Only well matching light-quark jets (α_{jp} <0.2) were used.

in a bias ΔC^{bias} defined by:

$$\Delta C^{bias} = |\Delta C - \Delta C^{true}| = 0.05 \pm 0.51\%. \tag{5.2}$$

The robustness of the bias ΔC^{bias} versus the criterion used for the jet-to-parton matching is illustrated in Figure 5.5. It is shown that the Gaussian fitted mean of the E^{GEN}/E^{CAL} distribution is invariant to the level of 0.4% versus the allowed angle α_{jp} between the vectors of the jet and primary parton. This angle was changed from 0.1 rad to 0.3 rad in the (θ, ϕ) metric. The perfect agreement between the experimentally measured and theoretically predicted ΔC value confirms the assumption that the mismeasured light quark JES is by far the most important cause for the significant shift observed in the initial m_W-distribution in Figure 5.1. This result illustrates the possibility to use semi-leptonic top quark decays for the data-driven determination of jet energy calibration factors, even with limited accumulated events. However, the robustness of this bias with respect to the main systematic uncertainties still has to be proven.

5.3 Study of the Systematic Uncertainties

As the proposed measurement of the absolute jet energy scale will not be limited by its statistical accuracy, an appropriate study of the systematical uncertainties on the ΔC-estimator is crucial. The influence of systematic effects on the estimation of m_W via the Gaussian fit on the spectrum is translated into a systematic uncertainty on the estimation of ΔC, $\Delta\, C^{syst}$. The proposed measurement of the inclusive light quark

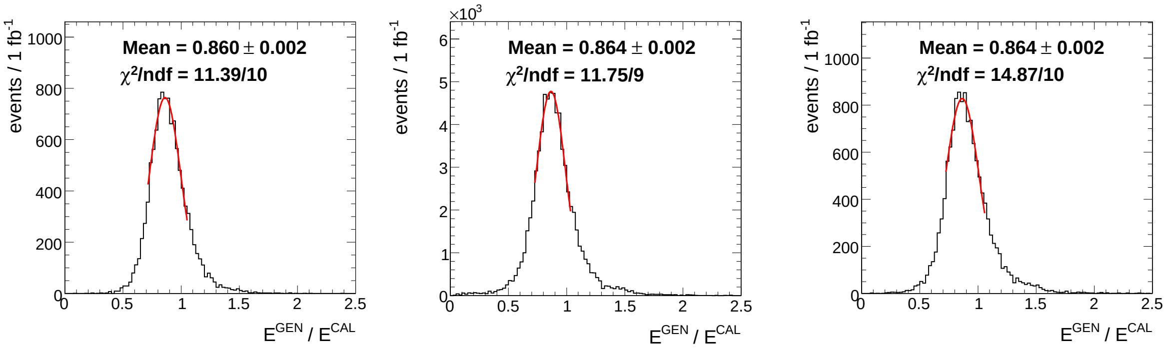


Figure 5.5: Distributions of E^{GEN}/E^{CAL} using the (θ, ϕ) metric for all jets matching better to a parton than 0.1 (left), 0.2 (middle) and 0.3 (right).

JES correction factor is fully data-driven; the method does not rely on any knowledge obtained from the study of Monte-Carlo simulated events, as the estimator is shown to be unbiased with respect to the true JES correction. Consequently, only the systematic effect arising from the uncertainty on the employed process production cross-sections, experimental techniques such as b-tagging and the effect of pile-up collisions have to be accounted for. The below described strategies to estimate the systematic bias on the ΔC-measurement are based on the results presented in [116].

5.3.1 Pile-up collisions

Fluctuations in the rate and impact of minimum bias collisions that are piled-up to the hard-Q^2 event will result in a systematic uncertainty on the measured light quark JES correction factor ΔC. In order to study this effect, a Monte-Carlo simulated event sample with and without the inclusion of pile-up collisions was generated using the `PYTHIA` program. The number of minimum bias events added to the hard event is randomly selected from a Poisson distribution with an expectation value λ set to 3.5 minimum bias events. This corresponds to the expected number of additional collisions during the same bunch crossing and for a LHC luminosity of $2.10^{33}\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The systematic uncertainty on ΔC due to the inclusion of pile-up (PU) collisions is taken as the difference between the result obtained for the sample with and without minimum bias events added,

$$\Delta\, C_{PU}^{syst} = (\Delta\, C)_{PU} - (\Delta\, C)_{noPU} = -0.87 \pm 0.21\% \tag{5.3}$$

The negative sign of this bias is a result of the extra energy deposits in the selected light quark jets originating from the pile-up collisions. Therefore, the Monte-Carlo calibrated light quark JES becomes even more overcalibrated, effecting into a more negative $\Delta\, C$ correction factor.

The systematic effect of pile-up collisions is hence found non-negligible, the more because in this check only in-time pile-up events were accounted for. On the other hand, the choice to estimate the systematic bias from a sample with and without pile-up events is rather extreme. Once data is available, appropriate tools will envisage

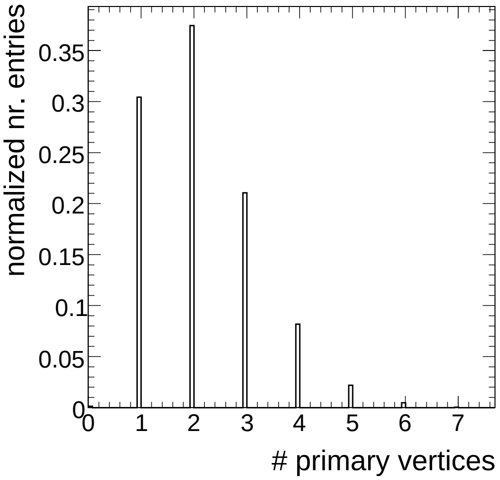


Figure 5.6: The number of reconstructed vertices in the event sample containing pile-up collisions.

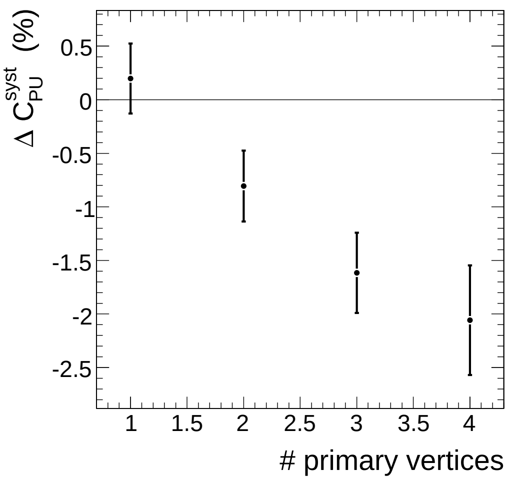


Figure 5.7: The systematic effect on the ΔC measurement due to the inclusion of pile-up collisions evaluated for semi-muonic signal event subsamples with a different number of reconstructed primary vertices.

to measure the pile-up rate. The uncertainty on this number will then offer a more adequate way to estimate the systematic bias on a measurement due to pile-up uncertainties. Furthermore, the pile-up systematic effect can be significantly reduced with a cut on the number of reconstructed primary vertices in the event[2], as is illustrated in Figure 5.7. When only events with one primary vertex are selected, the whole systematic effect on ΔC due to pile-up collisions appears to be cancelled. From Figure 5.6 it is observed that still $\simeq$30% of the events would survive such an extra requirement.

5.3.2 Process background contribution

Apart from the pile-up effect, the uncertainty in the ratio of $\mathrm{t\bar{t}}$ to channel background events will result in a systematic uncertainty on the ΔC correction factor. In this thesis LO cross-sections for all studied processes were employed, mainly because for the Wbb-production NLO calculations are not yet available. Hence, the uncertainty on the $\sigma_{\mathrm{t\bar{t}}}/\sigma_{proc.\,bgr}$-ratio will be large and difficult to estimate. In order to get an idea of the light quark JES measurement's sensitivity on a shift of the channel background level, the nominal level is increased by another factor of two. We remind that in the presented measurement the inclusive $W(bb)$ contribution was already raised by a factor of four to account for the expected QCD event rate after the physics object selection.

The extra rescaling factor of two results in a $\Delta C^{syst}_{proc\,bgr}$ equal to $0.31 \pm 0.99\,\%$. Hence, an increase of the W-boson or QCD production cross-sections is expected to have a limited systematic effect on the ΔC measurement, although this is difficult to

---

[2] For a description of the primary vertex reconstruction, we refer to Section 4.4.2 of the previous chapter.

state considering the insignificance of this uncertainty estimate. However, an uncertainty in the channel background level of a factor of two is very optimistic. Unfortunately, the limited amount of simulated statistics in the $W(bb)$-samples did not allow to test the measurement's bias with respect to higher rescaling factors. Apart from the cross-sections also the m_W shape from the QCD events has to be studied directly from the data.

5.3.3 *b*-tagging efficiency

The sensitivity of the ΔC-estimator is checked on a variation of the b-tagging efficiency within its expected uncertainties. In [53] a method is described to measure this efficiency from real data. The potential relative uncertainty on this measurement is found to be about 6% (4%) in the barrel and 10% (5%) in the endcap regions for $1\,\mathrm{fb}^{-1}$ ($10\,\mathrm{fb}^{-1}$) of accumulated data[3]. The method is however dominated by systematic uncertainties, and these potential uncertainties can be considered as conservative.

Any uncertainty on the b-tag efficiency will change the mistag rate and hence the fraction of combinatorial background events in the reconstructed m_W-distribution, analogue to the effect of a variation of the P^b_{jet}-cut value. In Figure 5.8 the relation between the b-tag efficiency and the selected track probability b-tag definition is illustrated. The b-tag efficiency is obtained from a sample of $\mathrm{t\bar{t}}$ jets that match well the direction of an initial b-quark. The b-tag requirement $P^b_{jet} < 0.5$ results in a b-tag efficiency of $\simeq$70%. We will asume an overall, rather conservative, b-tag efficiency uncertainty of 8% for $1\,\mathrm{fb}^{-1}$ of data. In order to select ±8% more or less b-quarks in a jet sample, the P_b^{jet} value should be shifted between 0.4 and 0.6. The systematic effect on ΔC due to the uncertainty on the b-tagging efficiency, $\Delta C^{syst}_{b-tagging}$, is defined as

$$\Delta C^{syst}_{b-tagging} = 1/2\left[(\Delta\, C)_{P_b^{jet}<0.4} - (\Delta\, C)_{P_b^{jet}<0.6}\right]\,, \tag{5.4}$$

and found to be 0.15±0.18%. Consequently, the proposed ΔC measurement is observed to be stable with respect to a change in the b-tagging working point.

The systematic uncertainties obtained for all three effects are summarised in Table 5.1. For both the process background and b-tagging systematic, the uncertainty on the bias is listed, as the bias itself is not significant. Hence, the total uncertainty of $\simeq$1.5% is conservative. The more because the most important effect, the impact of pile-up collisions, can easily be reduced with a cut on the number of reconstructed primary vertices in the event. On the contrary, the high uncertainty in the channel backgound event spectrum and level requires caution. Although a first test indicated the robustness of the presented JES correction determination versus a change in the channel background level by a factor of two, dedicated QCD data studies will allow to gain further confidence in this observation.

---

[3] Although these uncertainties are acquired for a given cut on the combined secondary vertex b-tag discriminator [132], similar uncertainties might be expected for the track probability algorithm.

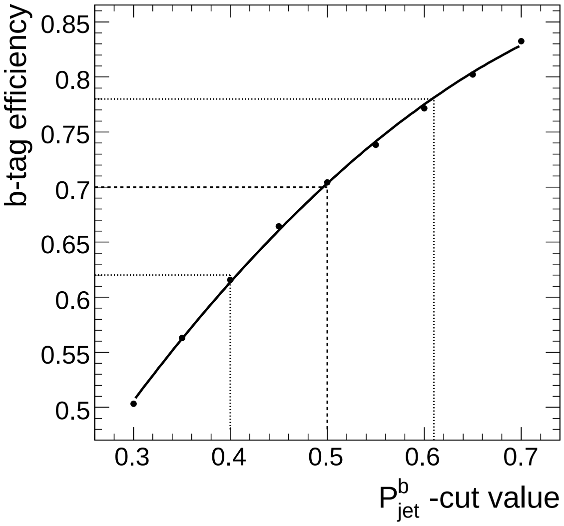


Figure 5.8: The b-tagging efficiency versus the track probability P^b_{jet} cut value.

Effect	Uncertainty Δ^{syst}
Pile-up collisions	0.87%
Channel background level	0.99%
b-tagging	0.15%
Total systematic uncertainty	1.32%
Statistical uncertainty ($1\,\mathrm{fb}^{-1}$)	0.46%
Total	1.40%

Table 5.1: Overview of the investigated systematic influences on the ΔC estimator, compared to the statistical uncertainty achievable for $1\,\mathrm{fb}^{-1}$ of accumulated data.

5.4 Conclusions

A first study is presented to demonstrate the importance of using the abundantly produced top quark events for the calibration of the jet energy scale. After a Monte-Carlo jet calibration as discussed in Section 4.1.3, the residual inclusive light quark JES miscalibration factor was determined via a fit to the reconstructed W-boson mass spectrum. The presented method should allow to measure this inclusive calibration factor ΔC with a precision of the order of 1%. Already for $1\,\mathrm{fb}^{-1}$ of data systematic effects dominate the statistical precision on this estimation. The influence of pile-up collisions is shown to introduce the dominant systematic uncertainty on the ΔC measurement. However, this effect will be reduced impressively with a cut on the number of reconstructed primary vertices. Furthermore, the systematic uncertainties due to a shift in W-boson or QCD production cross-section and due to the applied b-tagging algorithm are found negligible. Nevertheless, the method does rely on an efficient and well understood b-tagging algorithm, which is probably not present at the start-up of the LHC machine. Once data is available, also the effect of the process background events should be studied carefully (e.g. the shape of the m_W-distribution for QCD events).

As soon as this method is well understood and stable in data, and more statistics is accumulated, the estimated value of ΔC can be differentiated in the jet observable space, for example as a function of the pseudo-rapidity or the transverse momentum of the jet. Also, a similar method exploiting both the W-boson mass and the top quark mass constraints would allow to estimate simultaneously the absolute jet energy scale of jets originating from light and b-quarks. On an event-by-event basis one can force the two light quark jets from the W-decay to obtain the world average measured W-boson mass. The energy scale of the b-quark can then be adapted to fit the world average measured top quark mass.

In both the $\mathrm{t\bar{t}}$-production cross-section and the top quark mass measurement presented in the next chapters, the inclusive light quark JES correction derived in this study is applied. Concrete, the fourmomenta of both light quark jet candidates are multiplied with a fixed correction factor of $1 - 0.136 = 0.84$, in order to account for the light quark JES overcalibration after the Monte-Carlo jet calibration. The two b-quark jet candidates are left unchanged.

Chapter 6

Advanced Event Selection & the $\sigma_{t\bar{t}}$ Measurement

In Section 4.6 the threshold cuts on the reconstructed physics objects in the CMS detector were introduced. Apart from the necessary requirements for an event to be triggered or for the reconstructed objects to be accepted and reasonably well measured, these cuts ensured the possibility to be able to reconstruct the full semi-muonic $t\bar{t}$ event topology. Also, this event selection allowed a significant reduction of the $W(bb)+jets$, QCD and other-$t\bar{t}$ channel background events, summarised in Table 4.3.

However, important kinematical and topological information is still unexploited (e.g. $m_{jj} = M_W$), or limited employed (e.g. *b*-tagging information). The combination of all this extra information will be the topic of Section 6.2: by means of a Likelihood Ratio method, the discriminating power of several event observables will be merged. This allows not only to further enhance the signal-to-background ratio, but also to define a probability an event is signal and well reconstructed. In a first section however, the notion of kinematic fitting is introduced, in order to be able to exploit the topological constraints in the most optimal way.

Once the event selection is fully determined, Section 6.3 will elaborate on the definition of a $t\bar{t}$ production cross-section estimator. Because of the high signal-to-background ratio in the remaining event sample, this measurement can be treated as a simple counting experiment. It will be shown that already for an integrated luminosity of $1\,\text{fb}^{-1}$ this $\sigma(t\bar{t})$-estimation is dominated by systematic effects.

6.1 Fitting of Event Topologies with External Constraints

Many physical processes lead to final state topologies with measurable and unmeasurable quantities. While the first category typically represents four-vector estimators of certain reconstructed objects (e.g. jets, tracks, etc), the unmeasured quantities are supposed to mark the undetected particles of the underlying primary event structure (e.g. neutrinos). Additionally, one often can define certain hypotheses the kinematics of the events should fulfil, such as a known invariant mass or the conservation of energy and momentum. These constraints can be used to slightly change the measured values within their uncertainties in order to fulfil the kinematic requirements. This procedure is known as kinematic fitting.

The application of a kinematic fit has already proven itself as a powerful tool in the event reconstruction. In particular at the LEP-accelerator the accurately measured incoming electron and positron momenta allowed to put stringent constraints on the reconstructed objects kinematics, leading to significantly improved estimators of e.g. the W-boson mass [136]. Also in the determination of the CKM matrix elements V_{cb} and V_{ub} in the BABAR experiment the use of a kinematic fit was of great importance [137–139]. Finally, kinematic fitting techniques can be applied to verify a hypothetical kinematic topology, as for example in the search for the Standard Model Higgs boson at LEP-2 [140].

Hence, the outcome of a kinematic fit is twofold: it leads to improved estimators of the underlying kinematics, and it supplies a χ^2-value that can be interpreted as a probability the imposed event hypothesis was true. Both aspects will be employed in what follows. But first, we will briefly sketch the mathematical concept of a non-linear least square fit using Lagrange Multipliers. For more information about the implementation and use of this kinematic fit package we refer to [141].

6.1.1 Using Lagrange Multipliers for a Non-Linear Least Square Fit

Let us consider a physical problem with n measured parameters $\vec{y}$ and p unmeasured parameters $\vec{a}$ that are supposed to fulfil m constraints $\vec{f}$ as defined in (6.1). By definition, all requirements are fulfilled for the true parameters $\bar{a}$ and $\bar{y}$.

$$\begin{aligned} f_1(\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_p, \bar{y}_1, \bar{y}_2, \ldots, \bar{y}_n) &= 0 \\ f_2(\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_p, \bar{y}_1, \bar{y}_2, \ldots, \bar{y}_n) &= 0 \\ &\vdots \\ f_m(\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_p, \bar{y}_1, \bar{y}_2, \ldots, \bar{y}_n) &= 0. \end{aligned} \tag{6.1}$$

However, the resolutions on the measured objects will in general ensure that these equations are not solved by the measured values $\vec{y}$. Corrections $\Delta\vec{y}$ have to be calculated, requiring the sum $\vec{y}\,' = \vec{y} + \Delta\vec{y}$ to fulfil all constraints. For these corrections, the weighted sum

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

should be minimal, with $\mathcal{V}$ being the covariance matrix of the measured parameters. A general method to determine local extrema of non-linear functions of many variables is the use of Lagrange Multipliers. The likelihood is then defined as follows:

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

with $\vec{\lambda}$ the Lagrange Multipliers. The necessary condition for a local minimum of this function is then equivalent to the condition for a minimum of $S(\vec{y})$ under the constraints

$f_k(\vec{y},\vec{a}) = 0$. With linear constraints the solution can be calculated analytically, otherwise the constraint functions are linearised, and a iterative procedure is started to solve the equations and find the corresponding optimal corrections $\Delta\vec{y}$ within the required numerical precision.

6.1.2 Fitting the semi-muonic $t\bar{t}$ decay channel

At hadron colliders like the LHC less stringent constraints on energy and momentum conservation between initial and final state kinematics can be posted compared to e.g. at LEP, because of the unknown boost of the final particles along the Z-axis. Nevertheless, the use of a kinematic fit will still be a very important tool in the study of many physics channels, and in particular in the analyses based on the semi-muonic decaying $t\bar{t}$-events.

This decay channel results in a final state topology consisting of four jets, an isolated muon and missing transverse energy. Both for muon and jets the reconstructed four-vectors can be considered as a direct measurement of the generated quarks and muon four-momenta. Also the invisible neutrino can be treated as a measured particle, where the transverse missing energy measurement is used to deliver a coarse estimate of the particle's p_x and p_y component.

For this event topology many theoretical kinematic constraints can be examined: the hadronic and leptonic W-boson and top quark invariant masses, the back to back production of the top and anti-top quark in the transverse plane and the conservation of the transverse momentum. However, the non-negligible transverse momentum contribution of the underlying event, and the inclusion of the unknown longitudinal momentum of the neutrino had a bad influence on the overall fit performance. Therefore we restrict ourselves to only impose constraints on the hadronic, fully-determined part of the $t\bar{t}$-decay.

The demand for a light quark jets invariant mass equal to the precisely known W-boson mass, $m_{jj} = m_W$ will not only result in improved jet resolutions, but will also be shown useful in the enhancement of the S/B-ratio (Section 6.2). Additionally, the hadronic top quark mass constraint, $m_{\mu\nu_\mu b} = m_{top}$, will be used in Chapter 7 to scan the probability that a given set of three jets originates from the decay of a top quark with mass m_{top}.

Apart from the constraints, all involved jet four-momenta and their corresponding covariance matrices should be given as an input to the kinematic fit package. These covariance matrices are needed in order to define the freedom of the kinematic fit. Only their diagonal terms are determined via the evaluation of the resolution functions that were plotted for various jet observables in Figure 4.16.

Furthermore, several four-vector parameterisations are implemented in the kinematic fit package:

- **Momentum deviating four-vector parameterisation**
  This four-vector parameterisation describes the deviations from the measured momentum in a local coordinate system:

$$\vec{u_1} = \frac{\vec{p}_{meas}}{|\vec{p}_{meas}|}, \quad \vec{u_2} = \frac{\vec{u_3} \times \vec{u_1}}{|\vec{u_3} \times \vec{u_1}|}, \quad \vec{u_3} = \frac{\vec{u_z} \times \vec{u_1}}{|\vec{u_z} \times \vec{u_1}|},$$

	`EMomDev`	`EtThetaPhi`	`EtEtaPhi`
convergence rate (%)	98.8	100	99.96

Table 6.1: Comparison of the convergence rate of a kinematic fit imposing both the hadronic W-boson and top quark mass constraint on well matching semi-muonic decaying $t\bar{t}$-events for different jet parameterisations and corresponding

where $\vec{u_z}$ is the unit vector of the detector's z-axis. The coordinates a, b and c of the fitted momentum are defined in this frame as

$$\vec{p}_{fit} = a\,|\vec{p}_{meas}|\,\vec{u_1} + b\,\vec{u_2} + c\,\vec{u_3},$$

and are by definition equal to (1,0,0) for the measured momentum. For the jet energy it is common to choose a free floating fitted energy, and consequently also mass.
(`TFitParticleEMomDev`):

$$E_{fit} = d\,E_{meas}.$$

- **Two commonly used parameterisations at hadron colliders**
  Especially in the study of jets at hadron colliders, often the use of the Lorentz-boost invariant E_T measurement is preferred to the jet's energy. Two 3D implementations will be considered: the `TFitParticleEtThetaPhi` parameterisation, in which the object's E_T, θ and ϕ measurements will be fitted, and the `TFitParticleEtEtaPhi` parameterisation where the object's θ angle is replaced by its pseudorapidity. In this implementation the jet mass is tied to the fitted values of the other three parameters.

Choosing the optimal jet parameterisations

As only constraints on the hadronic part of the semi-muonic $t\bar{t}$-decay will be imposed, finding the optimal object parameterisations is limited to a search for the best jet parameterisation. Choosing the most optimal four-vector and covariance matrix parameterisation is of major importance, as both the convergence rate of the kinematic fit and the resolutions on the fitted jet kinematics will depend on it. As a consequence, also the performance and precision of the later top quark analyses will be affected.

In order to compare different jet parameterisations, the outcome of a kinematic fit setup with both the hadronic W-mass and top quark mass constraints was studied. Only events were considered in which for all three hadronic top quark decay products, a well matched jet was found with $\alpha_{jp} < 0.2$. Hence, only those jet combinations were selected where the kinematic fit is supposed to converge. The obtained convergence rates for the different jet parameterisations are summarised in Table 6.1, and are shown to be close to 100% in all three cases.

An objective criterion to find the most optimal jet parameterisation, is to look at the resolutions on the fitted direction and energy of the jet, respectively illustrated in Figure 6.1 and Figure 6.2. From these plots it is clear that due to the much larger uncertainties on the jet energy measurement compared to its direction, only the jet

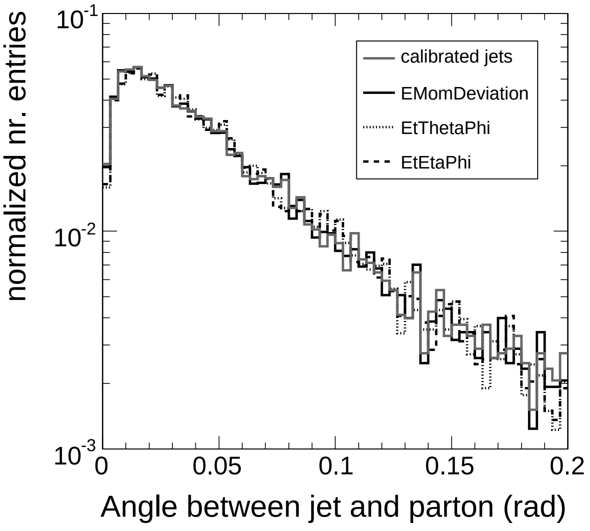


Figure 6.1: The difference in calibrated and fitted and generated jet direction with different choices of the jet parameterisation. Both the hadronic W-boson and top quark mass constraint were applied on $t\bar{t}$-events with a well matching parton for each of the three hadronic jets.

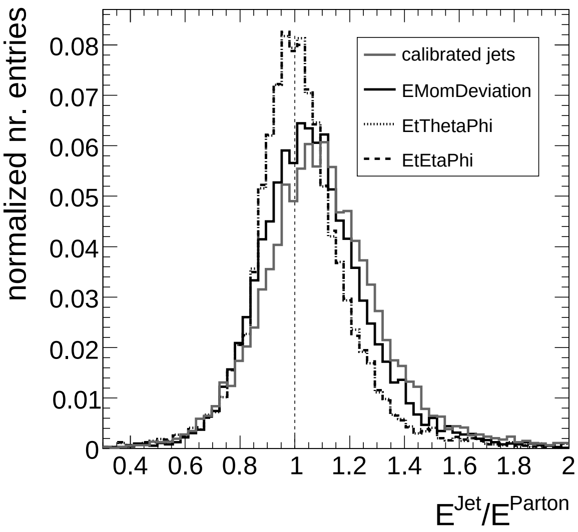


Figure 6.2: The relative calibrated and fitted jet energy resolution with different choices of the jet parameterisation. Both the hadronic W-boson and top quark mass constraint were applied on $t\bar{t}$-events with a well matching parton for each of the three hadronic jets.

energies are affected by the kinematic fit. Although already a reduced bias and spread on the fitted E^{FIT}/E^{PARTON}-distribution is visible with the `EMomDev` parameterisation, the best results are obtained with the `EtThetaPhi` and `EtEtaPhi` parameterisation. Throughout the remainder of this work, the `EtThetaPhi` parameterisation is selected and employed as the optimal jet parameterisation.

6.2 Enhancing the Signal to Channel Background Ratio

In Table 4.3 in the previous chapter the selection efficiencies obtained after some selection cuts on the various reconstructed objects were summarised. The trigger, muon and jet related requirements resulted in a final signal-to-background ratio of ∼2.56, an improvement with a factor of around 23 with respect to the S/B value of the initially expected number of events. This fraction of signal events can however be enhanced by exploiting the characteristics of the signal event topology. In this section several observables will be defined with a discrimination power to extract signal from background. The information of all these observables will then be combined by means of a Likelihood Ratio (LR) method. Finally, a minimal LR-discriminant value is defined for an event to be selected, as well as a probability that a given reconstructed event corresponds to a well reconstructed signal event. Both aspects will be shown useful in further analyses.

6.2.1 Sensitive observables

Several observables were determined to separate the semi-muonic signal $t\bar{t}$-events from the channel background events resulting from other $t\bar{t}$-decays and multi-jet W-boson events. All are calculated from the kinematics and properties of the six reconstructed and selected semi-muonic decay products:

- the missing transverse energy, 'E_T^{miss}'
- the logarithm of the χ^2-probability of a kinematic fit imposing the hadronic W-mass constraint, '$\log(P_{\chi^2}^{hadr\,W})$'
- the sum of the P_b^{jet}-values for the two selected b-quark candidates, '$P_{b1}^{jet} + P_{b2}^{jet}$'
- the aplanarity, as defined in [142], which measures the transverse momentum component out of the event plane.

To further enhance the discrimination power for the last three above defined observables, the signal probability density function (pdf), $P_i^S(x_i)$, was constructed from events in which the relevant jets matched well to a parton. More precisely, an event's $P_{\chi^2}^{hadr.W}$-value was considered a signal contribution only if for both selected light quark jet candidates a well matching parton was found ($\alpha_{jp} < 0.2$), independent of the $t\bar{t}$-decay. Analogue, two well reconstructed b-quark jets were required to fill the b-tag related observable value to the signal histogram, and all jets were asked to be well matched for the aplanarity observable. For all four sensitive observables, the resulting signal and background probability density functions $P_i^S(x_i)$ and $P_i^B(x_i)$ are illustrated in the left plots in Figure 6.3. Only events passing the reconstructed objects selection cuts were included in the plots, while both for the inclusive signal and background histograms the cross-sections of the different subsamples were taken into account.

From these probability density functions $P_i^S(x_i)$ and $P_i^B(x_i)$ determined for each of the above defined observables x_i with $i \in \{1, ..., 4\}$, a distribution $R_i(x_i)$ is defined as

$$R_i(x_i) = \frac{P_i^S(x_i)}{P_i^S(x_i) + P_i^B(x_i)}. \tag{6.4}$$

By definition, all $R_i(x_i)$ values are constrained in the interval [0,1], what facilitates the search for accurate fit functions $F_i^R(x_i)$. For each observable i, the $R_i(x_i)$ function with its respective fit function $F_i^R(x_i)$ is shown on the right plots in Figure 6.3.

Using these signal and background definitions, the highest discrimination power is observed for the probability of the χ^2 and the b-tagging observable. For this first observable much higher values are preferred for signal compared to background contributions, reflecting the higher probability that both selected light quark jet candidates did indeed result from the decay of a W-boson. The on average lower values in the background pdf are mainly due to the missing hadronic W-boson decay in both W+jets events and fully leptonic $t\bar{t}$-decays, which were the two leading contributions among

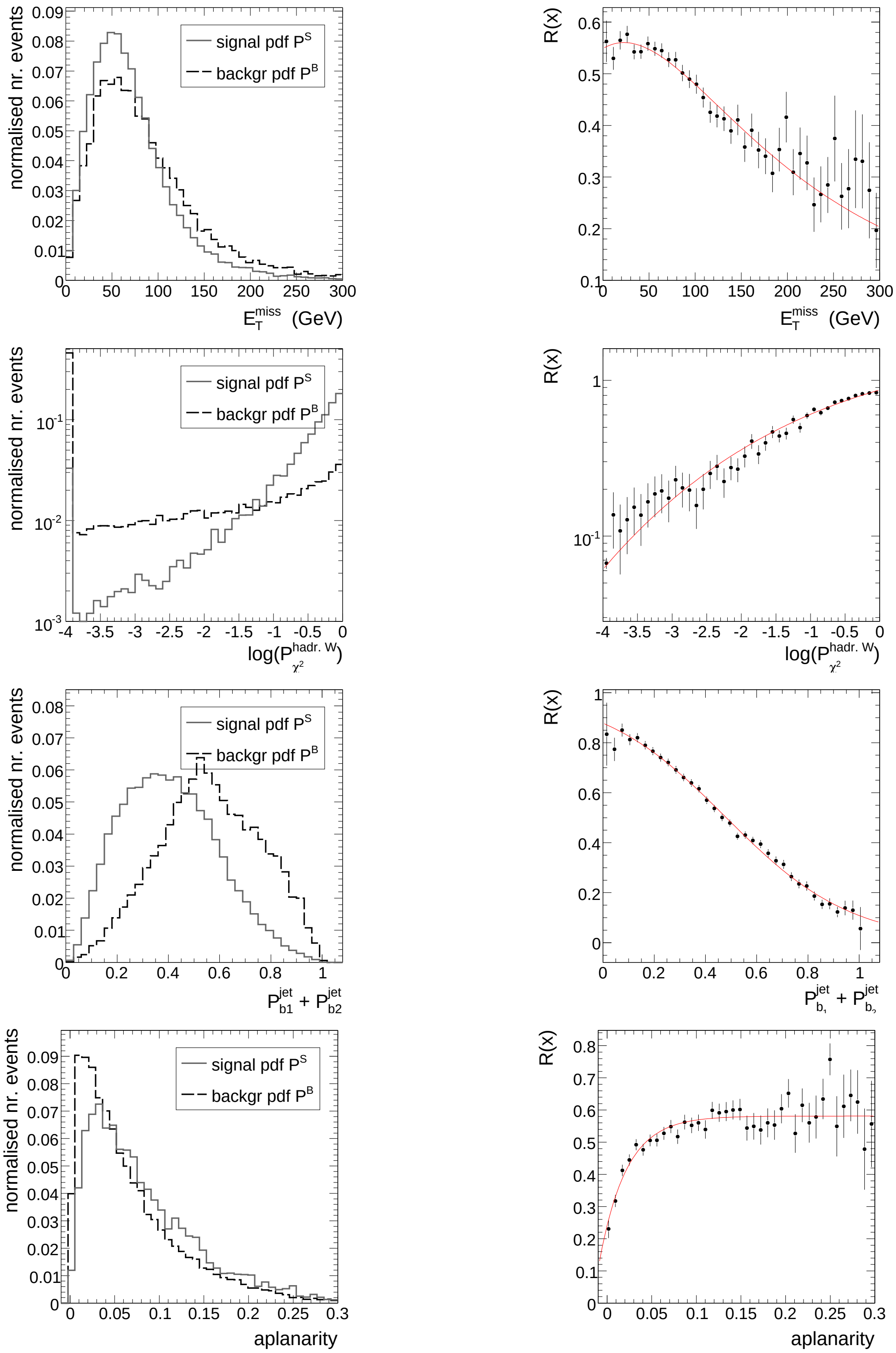


Figure 6.3: Distributions of the E_T^{miss}, $P_{b1}^{jet} + P_{b2}^{jet}$, $P_{\chi^2}^{hadr.W}$ and the aplanarity observables, along with their respective fitted $R_i = P_i^S/(P_i^S + P_i^B)$ distribution.

the selected channel background events (cfr. Figure 4.46). The peak at the lower edge comprises all events with a $\log(P_{\chi^2}^{hadr\,W})$ value lower than -4.

The sensitivity of the b-tag related observable is also expected, because of the lack of b-quarks in the final state of the W+jets events and the fact that signal events with badly reconstructed or unselected b-quark jets were added to the background pdf. Because of the previously introduced requirement to have minimum two jets with their P_b^{jet}-value lower than 0.5, the value of the $P_{b1}^{jet} + P_{b2}^{jet}$-observable is limited till 1.0 . From the negligible population of signal contributions at this upper bound it can be concluded that the signal events rejected after the P_b^{jet}-cut simply had very poorly reconstructed b-quark jets or b-quark jets that were already rejected by the jet E_T cuts.

Furthermore, the missing transverse energy and the aplanarity are both proven to contain a certain discriminating power, but both are probably sensitive to systematics. On average higher E_T^{miss}-values are observed for background events, explained by the high fraction of fully leptonic decays among the channel background events. In these events two instead of one neutrino is produced in the final state, what results in a relatively higher missing transverse energy. The last observable finally reflects a larger aplanarity for well reconstructed signal events with respect to background contributions. The decay products of signal events are hence on average more isotropically distributed in space.

6.2.2 Combining the information in a Likelihood Ratio discriminant

For each of the four selected observables, the probability density functions for both signal and background contributions were estimated using Monte-Carlo simulated event samples. This knowledge allows us to employ the Neyman-Pearson theorem [143]. Following this lemma, it is possible to define for each of the observables a probability or Likelihood Ratio function:

$$LR_i(x_i) = \frac{P_i^S(x_i)}{P_i^B(x_i)}, \tag{6.5}$$

with the most optimal discrimination power between signal and background contributions. Differently quoted, if we would define a test for the hypothesis that an event is signal given its observable value, the rejection region corresponding to the minimal α and β-values[1] corresponds to a one-dimensional cut on the $LR_i(x_i)$ distribution: $LR_i(x_i) < c$.

When using the relation in (6.4), Eq. 6.5 can be written as:

$$LR_i(x_i) = \frac{P_i^S(x_i)}{P_i^B(x_i)} = \frac{1}{\left(F_i^R(x_i)\right)^{-1} - 1}. \tag{6.6}$$

---

[1] The α-value denotes the probability a contribution is labelled signal while the opposite is true. The β-value on the other hand corresponds to the probability an event is not found to be signal when it actually is a signal contribution.

c_{ij}	E_T^{miss}	$log(P_{\chi^2}^{hadr.W})$	$P_{b1}^{jet} + P_{b2}^{jet}$	$Aplanarity$
E_T^{miss}	1	-0.09	0.04	-0.09
$log(P_{\chi^2}^{hadr.W})$	-	1	-0.04	0.05
$P_{b1}^{jet} + P_{b2}^{jet}$	-	-	1	0.06
$Aplanarity$	-	-	-	1

Table 6.2: Correlations between the signal event topology sensitive observables, estimated from the signal events, and asumed linear.

Further, the statistical information of the four observables might be combined by multiplying the likelihood ratios $LR_i(x_i)$ of the individual observables[2]:

$$LR^{comb} = \prod_{i=1}^{n} LR_i(x_i), \tag{6.7}$$

However, theoretically, this is only true if all observables are uncorrelated, which will not be the case in most physics problems. Also the observables defined for this signal event enhancement method are correlated. This is demonstrated in Table 6.2, where the estimated correlation factors c_{ij} between each pair of observables i and j is calculated for signal events. Although only observables were selected with a small correlation, equation (6.7) was replaced with the function LR' [133]:

$$LR'^{comb} = \prod_{i=1}^{n} LR_i(x_i)^{\frac{1}{\sum_{j=1}^{n} |c_{ij}|}}. \tag{6.8}$$

The ad-hoc correction in the exponent of the $LR_i(x_i)$ factors causes a reduction in power of each observable according to the mutual correlations. Finally, the obtained LR'^{comb} values are projected in the interval [0,1]:

$$LR_{comb}^{signSel} = \frac{LR'^{comb}}{1 + LR'^{comb}}. \tag{6.9}$$

In Figure 6.4 the combined likelihood ratio discriminant distribution for the inclusive signal and channel background samples is plotted. By definition, the region close to unity is highly dominated by well reconstructed signal events. However, for an important fraction of the selected semi-muonic signal events a low, background-like, combined LR value is obtained. This is a result of the particular definition of the observable's signal and background pdf's. An observable's value was only considered signal if the relevant jets to calculate the given observable value were well matched to a parton. As a consequence, all signal events with poorly reconstructed or wrongly selected jets receive a low combined likelihood ratio value. This conclusion is further demonstrated in Figure 6.5. In this plot, the angular jet reconstruction quality marker

---

[2] This equation is strictly only valid if exactly the same signal and background definition is used in all pdf's and if all observables are uncorrelated. The constructed LR^{comb} discriminator is thus an approximation of the exact Neyman-Pearson observable.

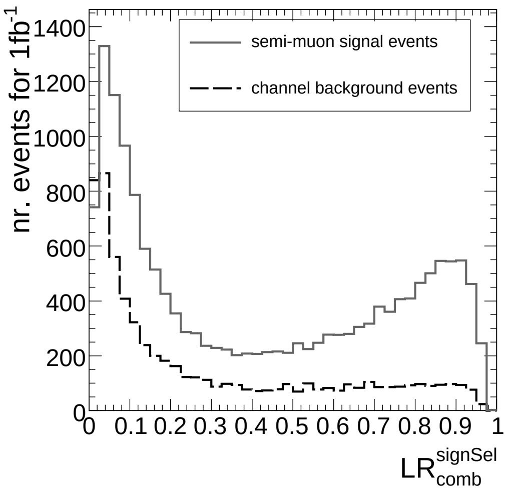


Figure 6.4: The combined LR discriminant value for the inclusive signal and background samples using only events passing the object selection.

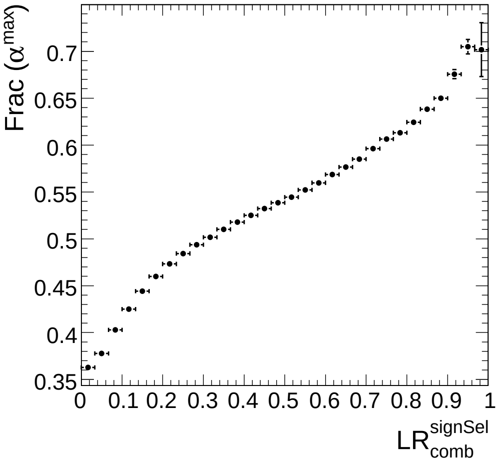


Figure 6.5: The Frac(α^{max}) jet reconstruction quality marker as defined in Section 4.2 after an increased cut on the combined Likelihood Ratio discriminant value.

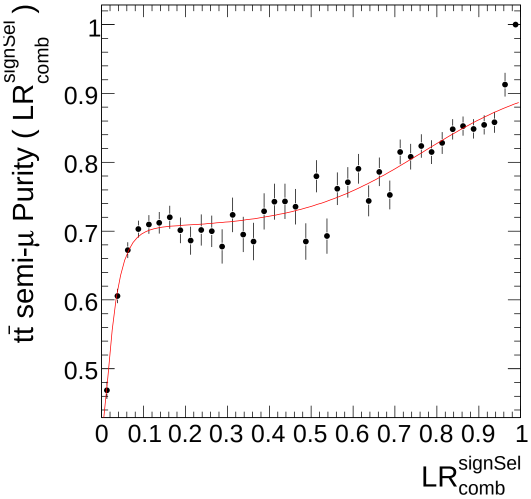


Figure 6.6: The bin-by-bin signal event purity given a certain combined likelihood ratio value.

$Frac(\alpha^{max})$[3] is calculated for an increasing cut-value on the combined likelihood ratio, $LR_{comb}^{signSel}$. Only semi-muonic decaying $t\bar{t}$ signal events were used as input to this plot. Clearly, the fraction of events with all four jets well reconstructed ($\alpha_{jp} < 0.2$) increases impressively with an increasing $LR_{comb}^{signSel}$ cut-value. Hence, a cut on the combined LR-discriminant value will not only reduce the fraction of channel background events

---

[3] This quality marker was introduced in the framework of the jet study presented in Section 4.2.

in the selected event sample, but will also deliver a way to exclude badly reconstructed signal events from further analysis. This, due to the particular signal and background definition employed to determine the observable's pdf's.

Furthermore, the inclusive Likelihood Ratio discriminant distributions showed in Figure 6.4 can be employed to calculate bin-by-bin the fraction of signal events for a given combined likelihood value:

$$P[signal]\,(LR_{comb}^{signSel}) = \frac{N_S(LR_{comb}^{signSel})}{N_S(LR_{comb}^{signSel}) + N_B(LR_{comb}^{signSel})}. \tag{6.10}$$

This fraction can be interpreted as an estimate for the probability a certain selected event is a semi-muonic event, given its $LR_{comb}^{signSel}$-value. The resulting curve and respective fit function is illustrated in Figure 6.6. An evaluation of this function will allow the weighting of events in the later top quark mass measurement (cfr. Chapter 7).

6.2.3 Choosing the optimal cut value

The optimal $LR_{comb}^{signSel}$ cut-value is depending on the definition of the background used in the final analysis. A procedure is presented to improve the significance of the signal to the background in a certain analysis. This background definition will be different in both upcoming analyses: the estimate of the top quark production cross-section $\sigma_{t\bar{t}}$ by means of a counting experiment, and the measurement of the top quark mass.

In the first study, $W(bb) + jets$ and QCD events are considered as background in the cross-section estimation. Uncertainties in the production cross-section of these processes would indeed lead to important systematic uncertainties on the final precision of the $\sigma(t\bar{t})$-measurement. Hence, for this study the aim is to search the $LR_{comb}^{signSel}$-value which maximises the significance $S_{\sigma(t\bar{t})}$ of $t\bar{t}$ events in the selected event sample:

$$S_{\sigma(t\bar{t})} = \frac{N_{t\bar{t}}}{\sqrt{N_{W(bb)+jets}}}. \tag{6.11}$$

For an increasing cut on the combined LR-discriminant, this $t\bar{t}$-events significance was calculated in each of the remaining event samples. The result of this scan is plotted in Figure 6.7. Already without any cut, it is possible to observe the $t\bar{t}$-events above the $W(bb)$+jet channel background with a significance of around 500. However, it is observed that a cut on the $LR_{comb}^{signSel}$-values allows to push this significance to even higher values. The maximum $S_{\sigma(t\bar{t})}$ is obtained when only events are selected with a LR-value above 0.7-0.8 .

The same strategy is now employed for the case of the top quark mass measurement. Only the definition of what is considered as a signal and background contributions is changed. As the top quark measurement presented in the next chapter will be based only on the hadronic part of the semi-muonic $t\bar{t}$-decay, the possibility to reconstruct the four-momentum of the hadronic top quark is fully determined by how well the three hadronic decay products are reconstructed. Consequently, signal events can be defined as semi-muonic events with well reconstructed and selected jets in the final state ($\alpha_{jp} < 0.2$), while all other events can be considered background contributions. The

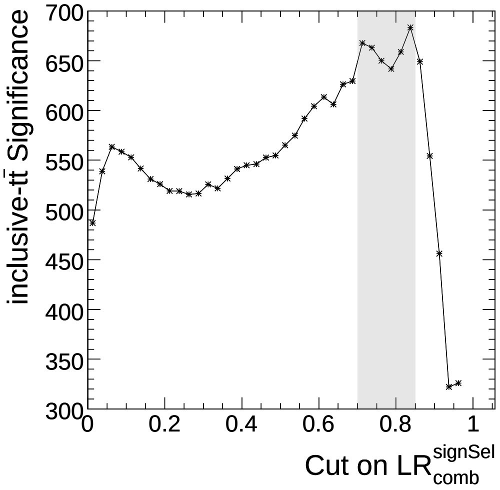


Figure 6.7: The significance $S_{\sigma(t\bar{t})}$ to select $t\bar{t}$-events from $W(bb)$+jet channel background, with $S_{\sigma(t\bar{t})}$ defined in Eq. 6.11 ($1\,\mathrm{fb}^{-1}$).

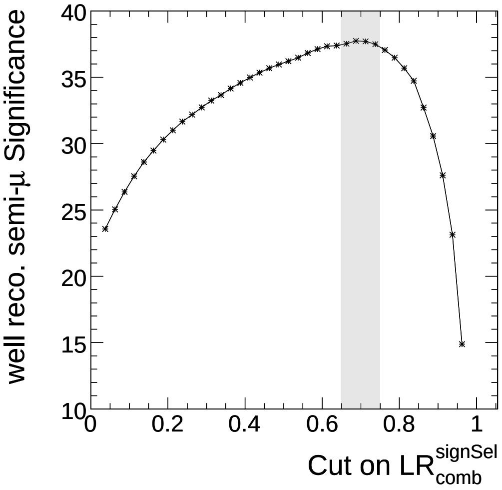


Figure 6.8: The significance $S_{M_{top}}$ to select well reconstructed semi-muonic signal events from all other events, with $S_{M_{top}}$ defined in Eq. 6.12 ($1\,\mathrm{fb}^{-1}$).

$LR^{signSel}_{comb}$ cut maximising the significance of the useful signal events to the background,

$$S_{M_{top}} = \frac{N_{well\ reco.\ signal}}{\sqrt{N_{badly\ reco.\ signal} + N_{channel\ background}}}, \tag{6.12}$$

is now to be determined. From Figure 6.8 an optimal LR-cut region around ∼0.7 was found, resulting in a maximal significance of ∼37. An altering of the jet-parton matching tolerance to $\alpha_{jp} < 0.3$ did leave this optimal value invariant. Applying an additional $LR^{signSel}_{comb}$-cut at 0.7 results in an increase of the fraction of signal events with all jets well reconstructed from angular point of view from 36% to 60% (cfr. Figure 6.5).

Hence, for both definitions of signal events a rather high optimal LR-cut is preferred. Because of this coincidence, it is decided to apply the same additional event selection cut in both the $\sigma(t\bar{t})$ and M_{top} measurement. Concrete, as an input to both analyses, only events will be accepted with a $LR^{signSel}_{comb}$ value higher than 0.7. Furthermore, it should be remarked that a high LR-cut value is also favoured to reject the QCD events passing the object selection cuts defined in Section 4.6.

The event selection efficiencies after this cut at $LR^{signSel}_{comb} = 0.7$ are summarised in Table 6.3 for each analysed signal and background `AlpGen` subsample. All mentioned statistics and their uncertainties are calculated for an integrated luminosity of $1\,\mathrm{fb}^{-1}$. In a third column, the relative reduction for each subsample is given when applying the Likelihood Ratio cut. For the different semi-muonic signal contributions, clearly the fraction of rejected events increases with the subsample's jet multiplicity. Where still 64% of the $t\bar{t}0j$-events survive the $LR^{signSel}_{comb}$-cut, only 9% of the $t\bar{t}4j$-events pass. This is another manifestation of the correlation between obtained LR-values and the angular reconstruction quality of the selected jets as was illustrated in Figure 6.5. Because the probability to pick up a wrong jet increases in events polluted with hard radiation jets, more events are rejected. After the full event selection, 5.85% of the initialy produced semi-muonic signal events survive, corresponding to (4857±44) events for $1\,\mathrm{fb}^{-1}$. In the calculation of this uncertainty only the limited size of the Monte-Carlo produced event samples was considered.

Also for events resulting from other $\mathrm{t\bar{t}}$-decays, the $LR^{signSel}_{comb}$-cut is observed to be more stringent for the high jet multiplicity `AlpGen` subsamples. Another time, this is due to the drop in probability to have correct (b-quark) jets selected and well reconstructed. Compared to the reduction factors obtained for the signal events, a relatively small increase is observed. This was expected as most of the remaining other $\mathrm{t\bar{t}}$ decays were the result of either a fully-leptonic or semi-tau $\mathrm{t\bar{t}}$ decay. In all these events two b-quark jets were produced with the same kinematics as in our signal events. Additionally, in the case of the semi-tau events, also correct light quark jets might be selected among the reconstructed jets. Finally, after all event selection cuts, a total of (854±20) other-$\mathrm{t\bar{t}}$ decay events remain for $1\,\mathrm{fb}^{-1}$.

In contrast to the observations for the other-$\mathrm{t\bar{t}}$ channel background events, the application of the extra LR-cut is very efficient in the reduction of the $W(bb)$+jet events. This was posed as an important necessity in order to allow a precise determination of the $\mathrm{t\bar{t}}$ production cross-section. Especially the fraction of W-boson events produced in association of only light quark jets decreases impressively: only a few percent of these events passing the reconstructed object selection cuts remain after the $LR^{signSel}_{comb}$-cut. For Wbb+jet events the last cut is somewhat more tolerant, as in these events the presence of b-quark jets will result in relatively higher values for the b-tagging related LR-observable. Finally, only (97±7) $W(bb)$+jet events are expected to survive the event selection when using $1\,\mathrm{fb}^{-1}$ of accumulated data. This corresponds to a global $W(bb)$+jet event selection efficiency of $\sim 3.5\,10^{-4}$. For QCD events one would expect the same order of reduction as for the W+jet events, because also in these events very low $LR^{signSel}_{comb}$ values are expected due to the missing b-quark jets and hadronic W-boson decay.

These main conclusions, a relative enhancement of low jet multiplicity signal events together with a very efficient reduction of $W(bb)$+jet events, are visualised in Figure 6.9. These pie charts give the relative event contribution after all reconstructed object selection cuts and the additional Likelihood Ratio cut, respectively for the different semi-muonic signal subsamples (left) and the main sources of channel background events (right). These plots can be compared with the ones obtained before the $LR^{signSel}_{comb}$-cut in Figure 4.45 and Figure 4.46.

Ultimately, this signal enhancing Likelihood Ratio method enables an increase of the signal-to-background ratio to a value of 5.2, which is more than twice the value obtained after the reconstructed objects selection cuts.

6.3 Estimation of the $\mathrm{t\bar{t}}$ Production cross-section

Measuring the cross-section of the $pp \rightarrow \mathrm{t\bar{t}} \rightarrow bq\bar{q}\mu\bar{\nu_{mu}}$ process at the LHC provides an essential test of the Standard Model. Radiative corrections predicted either in the Standard Model or in models beyond can be constrained via this measurement. Currently, such a comparison of experiment and theoretical prediction is possible at

sample	selected #evts.	# after LR-cut	reduction	overall sel.eff. (%)
signal				
$\mathrm{t\bar{t}}0j$ (excl)	(3.17±0.03)k	(1716±25)	-46%	6.09±0.09
$\mathrm{t\bar{t}}1j$ (excl)	(4.86±0.04)k	(1657±25)	-66%	6.58±0.1
$\mathrm{t\bar{t}}2j$ (excl)	(3.73±0.03)k	(921±19)	-75%	6.22±0.1
$\mathrm{t\bar{t}}3j$ (excl)	(1.70±0.03)k	(290±13)	-83%	4.90±0.2
$\mathrm{t\bar{t}}4j$ (incl)	(2.98±0.04)k	(274±13)	-91%	3.03±0.1
incl. semi-μ	(16.43±0.08)k	(4858±44)	-70%	5.85±0.11
background				
$\mathrm{t\bar{t}}0j$ (excl)	(0.29±0.01)k	(133±7)	-54%	0.08±0.005
$\mathrm{t\bar{t}}1j$ (excl)	(0.92±0.02)k	(236±10)	-74%	0.16±0.007
$\mathrm{t\bar{t}}2j$ (excl)	(1.14±0.02)k	(232±10)	-80%	0.27±0.01
$\mathrm{t\bar{t}}3j$ (excl)	(0.69±0.02)k	(111±8)	-84%	0.33±0.02
$\mathrm{t\bar{t}}4j$ (incl)	(1.53±0.03)k	(142±9)	-91%	0.27±0.02
incl. other-$\mathrm{t\bar{t}}$	(4.57±0.05)k	(854±20)	-81%	0.18±0.01
$W4j$ (excl)	(0.60±0.04)k	(7±4)	-99%	0.004±0.002
$W5j$ (excl)	(0.41±0.02)k	(20±4)	-95%	0.04±0.009
$W6j$ (incl)	(0.65±0.03)k	(13±4)	-98%	0.04±0.01
$Wbb2j$ (excl)	(0.120±0.004)k	(23±2)	-81%	0.14±0.01
$Wbb3j$ (excl)	(0.085±0.005)k	(11±2)	-87%	0.15±0.03
incl. $W(bb)$	(1.86±0.05)k	(73±8)	-96%	0.027±0.006
S/B	(2.56±0.03)	(5.24±0.13)		

Table 6.3: Overview of the number of events before and after applying the extra $LR^{signSel}_{comb}$ cut for all studied `AlpGen` samples. The initial number of events corresponds to the number of selected event for 1 fb^{-1} and after the full reconstructed objects selection cuts summarised in Table 4.3. The third column gives the relative reduction factors for each sample, while in the last column the overall event selection efficiency is listed after all cuts applied.

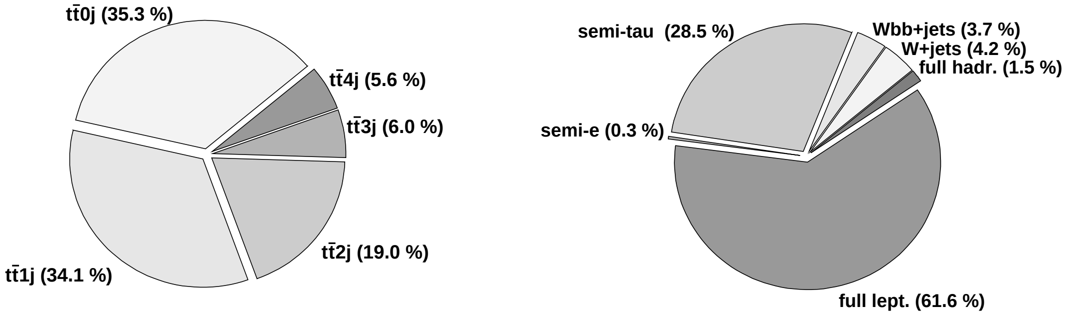


Figure 6.9: The composition of the inclusive semi-muonic signal event sample (left) and of the channel background sample (right) after the $LR_{comb}^{signSel}$ cut.

the Tevatron collider. The analysis of 760 pb^{-1} of Run-II data by both the CDF and DO collaboration resulted in a combined $t\bar{t}$ production cross-section of $\sigma_{t\bar{t}} = 7.3 \pm 0.5(stat.) \pm 0.6(syst.) \pm 0.4(lum.)$ pb [144]. This measurement is in good agreement with the theoretical SM value of 6.7 ± 0.8 pb, calculated for a top quark mass equal to 175 GeV/c^2. At the LHC, the up-scaled centre of mass energy requires a new consistency check of the Standard Model. The production of resonances decaying into $t\bar{t}$-pairs is for example predicted in various scenarios to become visible via the measurement of the differential $t\bar{t}$-production cross-section spectrum $d\,\sigma(t\bar{t})/d\,m_{t\bar{t}}$ [145, 146]. Furthermore, the much higher luminosity foreseen at the LHC will allow a $t\bar{t}$-event selection with a much higher purity compared to the purities obtained at the Tevatron collider. Especially the theoretical uncertainties on the background scale might be reduced significantly at the LHC due to the possibility to almost completely suppress the $W + jets$ and QCD channel background events in the selected event sample.

In this section a cross-section estimator is constructed. Where the actual estimator definition and its statistical precision is described in a first subsection, Section 6.4 will elaborate on the study of the most relevant systematic uncertainties on this cross-section measurement for $1\,\text{fb}^{-1}$.

6.3.1 Definition of the cross-section estimator

In a first strategy described in [147], the possibility was examined to estimate the cross-section based on the study of topological shape observables. Analogue to a measurement published by the DØ collaboration [148], six observables were defined and their statistical power to separate $t\bar{t}$ from W+jets events was investigated. However, already after the sequential reconstructed objects selection cuts described in Section 4.6, the few remaining $W + jets$ events provide observable distributions similar to the signal events, which spoils the sensitivity of a method based on topological observables.

Moreover, the higher statistics obtained at the LHC, and consequently the possibility for a highly efficient $t\bar{t}$-event selection, allows to simplify the $t\bar{t}$ production cross-section measurement to a simple counting experiment.

Starting from the number of data events counted after the full event selection (N_{sel}), the production cross-section for semi-muonic decaying $t\bar{t}$-pairs can be written as:

$$\sigma_{semi-\mu} \quad = \quad N_{sel} \;.\; P^{t\bar{t}}_{sim} \;.\; F^{semi-\mu}_{sim} \;.\; \frac{1}{\epsilon^{semi-\mu}_{sim}} \;.\; \frac{1}{\mathcal{L}}, \tag{6.13}$$

with $\mathcal{L}$ the integrated luminosity of the collected data sample, and:

- $P^{t\bar{t}}_{sim}$: the purity of $t\bar{t}$ events in the selected event sample
- $F^{semi-\mu}_{sim}$: the fraction of semi-muonic signal events among the selected $t\bar{t}$ events
- $\epsilon^{semi-\mu}_{sim}$: the selection efficiency of semi-muonic $t\bar{t}$-events after the full event selection

Each of these numbers need to be estimated from simulated Monte-Carlo event samples.

For all simulated and analysed `AlpGen` signal and background event subsamples, the event selection efficiencies and their respective uncertainties were shown in Table 6.3. From these data, the above defined quantities can be calculated as follows[4]:

$$\begin{aligned}
P^{t\bar{t}}_{sim} &= \frac{\sum N^{t\bar{t}}_{sel}}{N^{all}_{sel}} = (95.1 \pm 0.8)\% \\
F^{semi-\mu}_{sim} &= \frac{\sum N^{semi-\mu}_{sel}}{\sum N^{t\bar{t}}_{sel}} = (85.1 \pm 0.8)\% \\
\epsilon^{semi-\mu}_{sim} &= \frac{\sum N^{semi-\mu}_{sel}}{\sum N^{semi-\mu}_{produced}} = (5.85 \pm 0.05)\%.
\end{aligned}$$

An error propagation for all three above uncertainties in Eq. 6.13 results in a relative uncertainty on the cross-section measurement of 1.6%. This uncertainty is due to the limited number of Monte-Carlo simulated events, and can simply be reduced by simulating more events. However, it will be clear after the next paragraph where we describe the systematic effects that this is not the dominant source of uncertainty in the cross-section measurement.

Additionally, the statistical uncertainty on the number of selected data events should be accounted for. For a cross-section measurement based on $1\,\text{fb}^{-1}$ of accumulated data, a total of ~5800 events is expected to survive the full event selection. Assuming that N^{sel} is Poisson distributed, the uncertainty on this number becomes $\sqrt{N_{sel}}$, which on its turn translates into a relative uncertainty on the cross-section measurement of 1.3%. The expected precision on the $\sigma_{semi-\mu}$ estimation due to both statistics related sources of uncertainty are summarized in Table 6.4 for $1\,\text{fb}^{-1}$ of accumulated data.

---

[4] In this calculation of $P^{t\bar{t}}_{sim}$ the contribution of $W(bb)$ events is rescaled with a factor of four to incorporate the expected rate of QCD events remaining after the Likelihood ratio cut. The same approach was used in the determination of the light quark JES correction factor presented in the previous chapter.

	relative uncertainty on $\sigma_{semi-\mu}$ ($1\,\mathrm{fb}^{-1}$)
limited MC-statistics	1.6%
stat. uncertainty on N^{sel}	1.3%
total	2.1%

Table 6.4: Influence of the statistical uncertainties on the precision achievable in the semi-muonic decaying $\mathrm{t\bar{t}}$ production cross-section measurement.

6.4 Study of the Systematic Uncertainties on $\sigma_{\mathrm{t\bar{t}}}$

Analogue as in the measurement of the light quark JES scale calibration factor in the previous chapter, already for $1\,\mathrm{fb}^{-1}$ of data the systematic effects on the $\sigma_{semi-\mu}$ estimation will by far dominate the achieved statistical uncertainty. In this section, the influence on the measurement of $\sigma_{semi-\mu}$ is quantified, arising from the uncertainties in our theoretical, phenomenological and experimental knowledge. In contrast to the JES measurement, the estimation of the semi-muonic $\mathrm{t\bar{t}}$ production cross-section fully relies on a correct Monte-Carlo estimation of the signal event selection efficiency and purity. As a consequence, also the uncertainties in this simulation should be translated into a systematic effect on the cross-section measurement. Each of the studied systematics is introduced only on the signal events; the systematic effect in the relatively few selected channel background events is considered second order. For all systematic effects the relative difference in inferred cross-section is shown between the nominal event sample and the event sample generated with the deviation from this default as indicated, or between the two event samples with extreme changes in the parameter settings.

6.4.1 Pile-up Collisions

For the semi-muonic signal a simulated event sample is produced with and without the superposition of low-luminosity pile-up collisions (with an instantaneous luminosity of $2{\times}10^{33}\mathrm{cm}^{-2}\mathrm{s}^{-1}$). Only in-time inelastic pile-up collisions are taken into account, while the out-of-time contribution could be as large as the in-time contribution. The systematic uncertainty is taken as the relative difference between the semi-muonic cross-section obtained from the simulated sample with and without pile-up collisions,

$$\Delta\sigma^{syst,PU}_{semi-\mu} = |\sigma^{PU}_{semi-\mu} - \sigma^{no\,PU}_{semi-\mu}| \;/\; \sigma^{no\,PU}_{semi-\mu}, \tag{6.14}$$

and found to be $5.5 \pm 1.2\,\%$. Again, a cut on the number of reconstructed primary vertices is observed to be highly effective to reduce this systematic sensitivity, as is illustrated in Figure 6.10. When only events with one primary vertex are selected, the relative cross-section shift vanishes completely, while only increasing the statistical uncertainty from 2.1 to 3.8% for $1\,\mathrm{fb}^{-1}$.

6.4.2 The `PYTHIA` event description

The `PYTHIA` program was employed to simulate the parton showering and consequent hadronisation of the partons produced in both the hard interaction and the proton

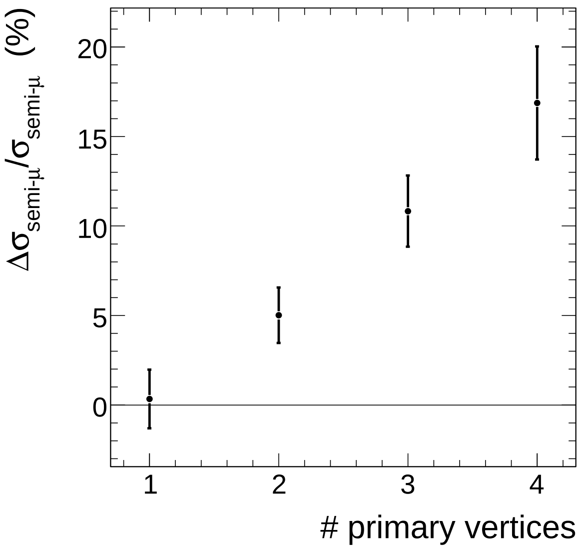


Figure 6.10: The systematic effect on the $\sigma_{semi-\mu}$ measurement due to the inclusion of pile-up collisions evaluated for event subsamples with a different number of reconstructed primary vertices.

remnants. In Chapter 2 all the different steps in the event generation were introduced, and for each of these the main `PYTHIA` parameters were listed with their respective tuned value and uncertainty.

In order to estimate the systematic effect on the $t\bar{t}$ cross-section originating from the uncertainties in the simulated Monte-Carlo event description, the main model parameters are varied within their uncertainties. Concrete, for each parameter deviation a different `PYTHIA` signal event sample is simulated using the fast simulation CMS software. The systematic uncertainty is then defined as half the relative $\sigma_{semi-\mu}$ difference obtained between the positive and negative shifted event sample:

$$\Delta\sigma^{syst}_{semi-\mu} = \frac{1}{2}\,|\sigma^{positive\,deviation}_{semi-\mu} - \sigma^{negative\,deviation}_{semi-\mu}|\;/\;\sigma^{nominal}_{semi-\mu}\,, \qquad (6.15)$$

assuming a linear dependence between the estimator and the shifted parameter value.

This procedure is applied on the main parameters in the description of the underlying event structure, the perturbative QCD radiation and the hadronisation model. The outcome is summarised in Table 6.5. For each effect, the reference to the corresponding section in Chapter 2 is given, where more information on the physical meaning and origin of the listed parameters and their uncertainties can be found.

- **Underlying Event**
  The CMS fast simulation software was used to generate a sample with two extreme values of the colour screening p_T cut-off value, which is the main parameter in the underlying event model implemented in the `PYTHIA` program. Using Eq. 6.15 a 1.2% relative shift on the cross-section estimate is obtained due to the variation of the $p_\perp^{UE}$ value. This estimation is still conservative as the 3σ confidence interval for the color screening $p_\perp^{UE}$ cut-off value is used to obtain a 1σ systematic uncertainty of the cross-section estimator.

Systematic effect	$\Delta\sigma^{syst.}_{semi-\mu}$	Reference
Underlying event 1/2 ($[p_{\perp}^{UE} = 3.4]$ - $[p_{\perp}^{UE} = 2.4]$)	1.2±0.6%	Section 2.8
Perturbative QCD radiation 1/2 ($[\Lambda_{QCD} = 0.35\,GeV]$ - $[\Lambda_{QCD} = 0.15\,GeV]$)	5.5±0.6%	Section 2.4.4
1/2 ($[{Q^2_{max}}^{ISR} = 4\,Q^2_{hard}]$ - $[{Q^2_{max}}^{ISR} = 0.25\,Q^2_{hard}]$) ($[{Q^2_{max}}^{FSR} = 16\,Q^2_{hard}]$ - $[{Q^2_{max}}^{FSR} = Q^2_{hard}]$)	0.5±0.6%	
Hadronisation 1/2 ($[rms\ p_{\perp} = 0.43\,\mathrm{GeV/c}]$ - $[rms\ p_{\perp} = 0.37\,\mathrm{GeV/c}]$) ($[Lund\ b = 0.56\,\mathrm{GeV}^{-2}\mathrm{c}^2]$ - $[Lund\ b = 0.48\,\mathrm{GeV}^{-2}\mathrm{c}^2]$)	0.1±0.6%	Section 2.7
1/2 ($[\epsilon_c = -0.020]$ - $[\epsilon_c = -0.042]$) ($[\epsilon_b = -0.0037]$ - $[\epsilon_b = -0.0045]$)	-0.4±0.6%	

Table 6.5: Overview of the systematic uncertainties on the semi-muonic cross-section measurement due to the uncertainties in the `PYTHIA` event description.

- **Perturbative QCD radiation**
  The main parameters in the description of the hard perturbative QCD radiation according to the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution equations (Eq. 2.13) are the general QCD scale parameter Λ_{QCD} and the virtuality cut-off scale Q^2_{max} which defines the allowed phase-space for initial state radiation (${Q^2_{max}}^{ISR}$) and indicates where the final state radiation takes over (${Q^2_{max}}^{FSR}$). It is found that the cross-section estimator is rather sensitive to the quality of the tuning of these perturbative QCD parameters, especially towards a variation of the Λ_{QCD} parameter. The systematic uncertainty on the cross section estimate due to perturbative QCD effects is defined as the maximum of the relative $\sigma_{semi-\mu}$ shifts resulting from the Λ_{QCD} and Q^2_{max} variation. This choice is motivated by the large correlation between both parameters, and results in a pQCD systematic uncertainty of 5.5%. It is expected that the use of the `AlpGen` generator reduces this systematic uncertainty, as high energetic radiation is no longer described by the parton showering. In a way, the uncertainty on the initial and final state radiation effects will be partly translated into an uncertainty on the exclusive cross-section of the different jet multiplicity `AlpGen` subsamples. However, the relative contribution of each subsample should rather be tuned to the physics observable's spectra observed in data.

- **Hadronisation**
  The `PYTHIA` implementation of the non-perturbative evolution of the parton's decay or so-called hadronisation process is based on the string fragmentation model. For the light quarks, two main parameters can be identified, the others being strongly correlated to these. The first is *Lund b*, which is strongly anti-correlated to *Lund a*, as both arise in the same Lund fragmentation functions. These functions express the probability that a hadron consumes a given fraction of the available longitudinal energy-momentum. For the heavy quark fragmentation the Peterson function is used instead, which is different for c and b quark. The transverse momenta of the hadrons are generated according to a flavour independent Gaussian probability density function of width $rms\ p_{\perp}$, being the second main parameter.
  In a first check the influence of the uncertainty on both the *Lund b* and $rms\ p_{\perp}$ parameter on the estimated cross-section is investigated. A second test concentrates on the heavy quark fragmentation functions. In this latter case the Peterson function's parameter ϵ_c and ϵ_b are varied in their uncertainties[5]. Both sets of deviations result in a non-significant and negligible systematic effect on the cross-section estimate. As a quantification of the size of this systematic effect, the statistical uncertainty on the relative $\sigma_{semi-\mu}$ shifts, 0.6%, is taken.

6.4.3 Parton Density Functions

The parton density functions or PDF's of protons are an essential ingredient in the simulation of the proton-proton scattering process. In Section 2.2 these functions were introduced and a method was described to estimate the systematic uncertainty on a measurement due to the uncertainties on the PDF fits. (Eq. 2.3). Instead of producing 40 different event samples, corresponding to the 20 `CTEQ6M` parameters that might be varied in positive and negative direction, an event weighting technique was preferred [149]. In a first step, the momentum fraction x, the initial flavour of the partons in the hard process and the Q^2 value of an event (which is taken equal to M_{top}^2) is extracted. With these values, 40 different event weights are calculated according to the 40 PDF sets shifted from the central `CTEQ6M` set. A combination of all weights results in 40 shifted cross-section estimates. The actual effect due to the uncertainties arising from the `CTEQ` fits is determined via the following formula :

$$(\Delta^{+}_{PDF}\sigma_{semi-\mu})^2 = \frac{1}{{\sigma^{c}_{semi-\mu}}^2} \sum_{j=1}^{20} \Big(max\left[\sigma^{+,j}_{semi-\mu} - \sigma^{c}_{semi-\mu},\, \sigma^{-,j}_{semi-\mu} - \sigma^{c}_{semi-\mu},\, 0\right] \Big)^2$$

---

[5] These uncertainties are however only estimations of the resolution of these parameters within the Lund and Peterson hadronisation model as implemented in the `PYTHIA` program. The fit of the model parameters is performed on several data distributions, but even the best fit results in a χ^2/NDF which is significantly deviating from unity, what indicates that the phenomenology of the model does not reflect the true physics.

and

$$(\Delta^{-}_{PDF}\sigma_{semi-\mu})^2 = \frac{1}{{\sigma^c_{semi-\mu}}^2}\sum_{j=1}^{20}\Big(max\left[\sigma^c_{semi-\mu} - \sigma^{+,j}_{semi-\mu},\, \sigma^c_{semi-\mu} - \sigma^{-,j}_{semi-\mu},\, 0\right]\Big)^2 \tag{6.16}$$

where $\sigma^c_{semi-\mu}$ is the central cross-section inferred from the `PYTHIA` fast simulated event sample applying the central `CTEQ6M` fit. This procedure takes into account the sign of the correlation between each of the PDF parameters and the observable of interest, being in this case the semi-muonic $\mathrm{t\bar{t}}$ cross-section. The resulting systematic uncertainties are $\Delta^{+}_{PDF}\sigma_{semi-\mu} = 3.5\%$ and $\Delta^{-}_{PDF}\sigma_{semi-\mu} = 3.7\%$. The largest of the two is taken as a systematic uncertainty.

6.4.4 Jet Energy Scale

The top quark events are reconstructed with jets in the final state and the event selection includes cuts on the transverse momentum of the leading jets. The reconstructed angles of the jets have a much better resolution compared to the energy scale of these jets. In this paragraph both the effect of a bias and a smearing of the jet energy scale is investigated.

In order to study the effect of a systematic shift on the inclusive jet energy scale, the four-momenta of the jets are scaled by a factor α using:

$$p^{\nu,jet}_{\pm\alpha} = (\,(1\pm\alpha)\,E,\;(1\pm\alpha)\,p_x,\;(1\pm\alpha)\,p_y,\;(1\pm\alpha)\,p_z\,)\,. \tag{6.17}$$

The shift α is applied on all reconstructed jets before the event selection and before the kinematic fit. Important to notice however is that due to the application of the data-driven light quark JES correction factor to all light quark jet candidates[6], the systematic shift on the initial light quark JES is expected to be impressively reduced. This is indeed what can be concluded from Figure 6.11. For α-values varying between -6 and 6% the mean value of a Gaussian fit on the (E^{CAL}/E^{PARTON}-1) distribution is plotted. In order to calculate this 'true' bias, only well matched jets where used from events that remained after the physics object selection cuts and the likelihood ratio cut. Where for the shifted b-quark JES a slope compatible to unity is found, the use of the light quark JES correction almost completely cancels the α-dependence. This reduction would have been even more pronounced if the kinematic fit to the hadronic W-boson mass was applied before the event selection (as will be illustrated in the next chapter). To conclude, any systematic shift on the initial JES will translate in an uncertainty on the final measured quantity, which is in this case the semi-muonic $\mathrm{t\bar{t}}$ cross-section. This uncertainty will by far be dominated by the systematic shift on the b-quark JES.

For each of the applied shifts, the relative cross-section bias is plotted in Figure 6.12. Paragraph 4.1.3 documents the precision on the JES that is expected to be achievable for 1 fb^{-1} of accumulated data. For the light quark JES 1% should be reachable with

---

[6] For each value of α, the light quark JES correction factor as defined in the previous chapter was recalculated. The resulting ΔC^{α} value is then applied to all light quark jets just before the jet selection.

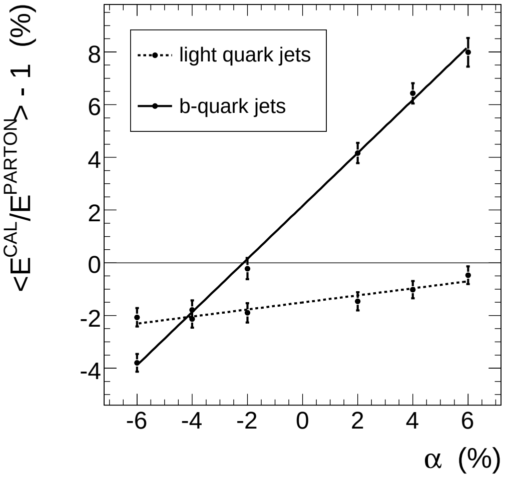


Figure 6.11: The Gaussian fitted mean of the $(E^{CAL}/E^{PARTON} - 1)$-distribution versus the applied inclusive JES shift. Only well matched jets were used from events passing all event selection cuts. A differentiation is made between light and b-quark jets.

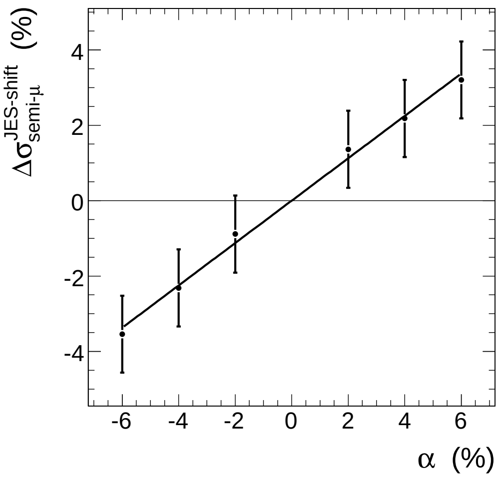


Figure 6.12: The relative cross-section shift for several fixed jet energy scale shifts.

the extra calibration factor derived from the study of semi-leptonic decaying $t\bar{t}$ events as introduced in the previous chapter. However, for the b-quark JES an uncertainty of 5% should be accounted for. Consequently, the systematic uncertainty on the cross-section measurement due to the uncertainty on the (b-quark) jet energy scale becomes $2.8 \pm 0.5\%$, where for the uncertainty the error on the slope in Figure 6.12 is used. For this, the correlations between the points are neglected.

Apart from a check of the measurement's robustness with respect to a fixed shift of the JES, also the influence of an additional jet energy smearing is investigated. Many effects, both detector and jet reconstruction issues, might result in a worse jet resolution in data compared to the one estimated from Monte-Carlo. Therefore, the jet energies are randomly and Gaussian smeared with an additional 20% relative to the measured value, and the resulting relative shift on the cross-section measurement is taken as an estimate of the systematic effect. For this purpose `AlpGen` samples were employed, which were simulated using the CMS fast simulation software. The extra 20% JES smearing results in an relative shift on $\sigma_{semi-\mu}$ of -6.7 ± 2.7%, and hence gives rise to a non-negligible uncertainty on the cross-section measurement. The reason for this shift is mainly the ≃6% drop in selection efficiency after the jet requirements defined in Section 4.6. To illustrate this, Figure 6.13 shows the transverse energy spectrum of the fourth leading E_T jet reconstructed in semi-muonic decaying `AlpGen` $t\bar{t}$ events. It is observed that the distribution is much broader after the extra 20% JES smearing, resulting in less events that will pass the $E_T^{CAL} > 30\,\text{GeV}$ requirement in the nominal event selection.

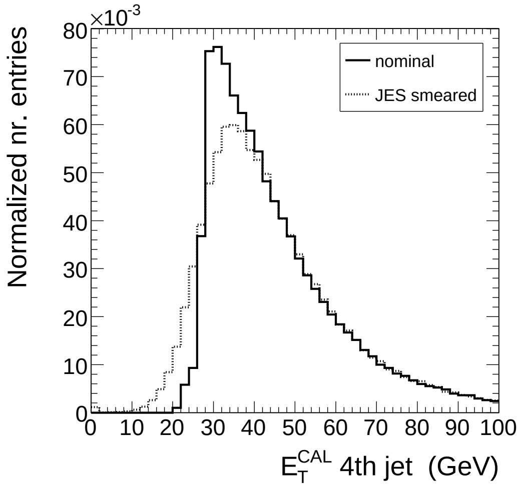


Figure 6.13: The 4th highest Monte-Carlo calibrated jet transverse energy, reconstructed in semi-muonic decaying `AlpGen` $t\bar{t}$ events, both for the nominal and the 20% extra JES smeared event sample.

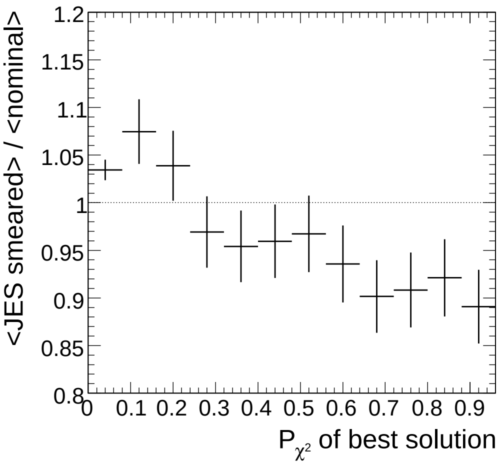


Figure 6.14: For all semi-muonic $t\bar{t}$ events passing the physics object selection cuts, the ratio of the number of events in the JES smearing enhanced sample and nominal sample is plotted as a function of the probability of the kinematic fit for the hadronic W-boson constraint.

Another possible reason for the decrease in selected event rate after the smearing, is the change in probability of the kinematic fit applied to the two light quark jet candidates in the event. The underestimation of the jet parameter resolutions in the JES smeared sample might result in a drop in probability and hence on average lower $LR_{comb}^{SignSel}$-values. This effect is indeed observed in Figure 6.14, where the relative mismatch of the P_{χ^2}-observable distribution is shown between the shifted and nominal sample. However, compared to the difference in E_T-spectrum of the fourth jet, this discrepancy can be seen as a second order, but non negligible, effect on the total jet smearing systematic uncertainty.

Finally it should be remarked that the above cross-section systematic uncertainty only takes into account the shift in the event selection efficiency $\epsilon_{sim}^{semi-\mu}$. However, the effect on $F_{sim}^{semi-\mu}$ was found to be negligible, which is also expected for the effect on $P_{sim}^{t\bar{t}}$.

6.4.5 *b*-tagging efficiency

In order to check the systematic uncertainty on the cross-section estimate due to the limited precision achievable in a data-driven method to measure the b-tagging efficiency, the b-tag working point is varied. In Section 5.3 an uncertainty of 8% on the b-tag efficiency measurement (as expected for $1\,\text{fb}^{-1}$ of data) was shown to correspond to a shift of the P_{jet}^{b} cut value from 0.4 to 0.6. This variation results in a relative cross-section bias equal to:

$$\begin{aligned}\Delta\sigma_{semi-\mu}^{b-tagging} &= 1/2\ |\Delta\sigma_{semi-\mu}(P_b^{jet} < 0.4) - \Delta\sigma_{semi-\mu}(P_b^{jet} < 0.6)|\ /\ \sigma_{semi-\mu}^{nominal} \\ &= 9.0 \pm 0.7\%\ .\end{aligned} \tag{6.18}$$

In the calculation of this uncertainty, only the shift on the event selection efficiency was found relevant. Between both extremes the fraction of $t\bar{t}$ events $P_{sim}^{t\bar{t}}$ changed with less than 2%, while the fraction of semi-muonic events among the selected $t\bar{t}$ events stayed constant. As expected, an accurate knowledge of the b-tagging efficiency will be crucial for a precision measurement of the $t\bar{t}$ cross-section. However, apart from the uncertainty on the b-tag efficiency and mistag rate at a selected working point, any difference between data and Monte-Carlo in the b-tagging discriminant distribution will give rise to a systematic uncertainty on the cross-section estimate. As will be illustrated in the next paragraph, the sensitivity of the measurement towards this extra source of uncertainty is strongly correlated to the systematic uncertainty resulting from an imperfectly aligned and calibrated detector.

6.4.6 Misalignment / miscalibration

In Section 3.2.5 the procedure is described of how to align and calibrate the different CMS detector compounds. Several scenario's are defined depending on the amount of accumulated data in the experiment. At the time the first 1 fb^{-1} of data is collected in the experiment, it is expected that the misalignment and miscalibration constants are still derived from the detailed study of the first 100 pb^{-1} data sample. Therefore, in this work, the systematic effect due to the imperfect detector calibration and alignment is estimated as the relative shift obtained between an expected knowledge at 100 pb^{-1} and the nominal, perfectly aligned and calibrated `AlpGen` fast simulation sample[7]:

$$\begin{aligned}\Delta\sigma_{semi-\mu}^{MisCal/MisAlign.} &= |\Delta\sigma_{semi-\mu}(100\,\mathrm{pb\ scenario}) - \Delta\sigma_{semi-\mu}(\mathrm{nominal})|\ /\ \sigma_{semi-\mu}^{nominal} \\ &= 9.1 \pm 2.7\ \%.\end{aligned}$$

The systematic uncertainty on the semi-muonic cross-section measurement due to an imperfect knowledge of the calibration and alignment constants is observed to be one of the main systematic uncertainties, and should hence be better understood.

The question can be posed if either the miscalibration or misalignment issue causes this important shift. The impact of the miscalibration is investigated via the comparison of the jet smearing obtained for the miscalibrated/misaligned event sample to the ones from the nominal and JES smeared sample. The result, plotted in Figure 6.15, clearly indicated the miscalibration has a negligible effect on the JES smearing.

On the contrary, Figure 6.16 illustrates the important mismatch between the b-tagging discriminant distribution obtained for the nominal and miscalibrated/misaligned event sample. All jets remaining after the jet selection requirements defined in Section 4.6 are included in the plot. It is clear that if only jets with a P_{jet}^{b} $<$0.5 are

---

[7] The uncertainty on $F_{sim}^{semi-\mu}$ was found negligible compared to the shift observed in the event selection efficiency.

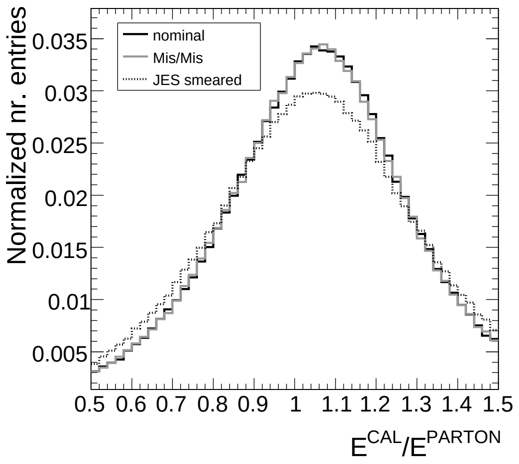


Figure 6.15: The JES smearing obtained from well matching semi-muonic $t\bar{t}$ jets, for the nominal, miscalibrated/misaligned and JES smeared event samples.

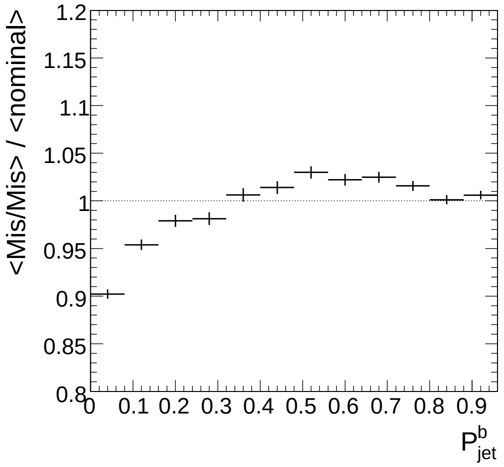


Figure 6.16: The ratio of the P^b_{jet} distribution obtained for the miscalibrated/misaligned and nominal sample. Only jets reconstructed in semi-muonic decaying $t\bar{t}$ events and passing the jet requirements defined in Section 4.6 were used.

accepted as b-quark jet candidates, relatively less events will be selected in the miscalibrated/misaligned event sample. Especially because two b-quark jets are required in the event selection, this will be the main cause for the rather large systematic shift on the cross-section estimate. Decreasing the P^b_{jet} value to increase the b-quark purity is expected to enhance this systematic effect even more.

In contrast to the systematic uncertainty obtained from the b-tagging efficiency uncertainty propagation described in the previous paragraph, this effect originates from a difference in the b-tagging discriminant distribution between data and Monte-Carlo. As a consequence, the systematic contributions due to the uncertainty on the JES smearing, the b-tagging efficiency and the detector calibration/alignment should be considered as rather uncorrelated sources of uncertainty, and consequently be added quadratically.

6.4.7 Process Background

The large uncertainties in the $W(bb)$ and QCD event production cross-section and hard radiation rate expected early in the experiment will give rise to a relatively large systematic uncertainty on the semi-muonic $t\bar{t}$ cross-section. In paragraph 5.3.2 the channel background sensitivity was checked by increasing of the inclusive channel background level by an additional scaling factor of two[8]. This was the maximum scaling factor that still allowed the reconstruction of a meaningful channel background m_W spectrum. However, as the cross-section measurement is just a counting experiment, this

---

[8] Indeed, ‘additional’, as the $W(bb)$ channel background contribution was already rescaled with a factor of four in order to incorporate the expected rate of selected QCD events.

factor can be increased to a more realistic factor of five. Such a rescale of the process background results in a shift of the $t\bar{t}$ event purity $P^{t\bar{t}}_{sim}$ from 95.1 to 79.6%, what corresponds to a systematic uncertainty on the cross-section estimate of -16.1 ± 1.1%. It should be stresses that this systematic uncertainy only includes a global increase of the selected number of $W(bb)$ events, in which the assumption was made that these events have the same kinematic topology than QCD events. The correctness of this hypothesis, as well as more accurate channel backgound level estimation is to be extracted from the data itself.

6.4.8 Integrated Luminosity

Another important systematic effect originates from the uncertainty on the integrated luminosity measurement. Following the predictions presented in Chapter 8 of [88], the integrated luminosity is assumed to be known with a precision of 10% from $1\,\text{fb}^{-1}$ of data. This uncertainty directly translates to an uncertainty on the cross-section estimate.

6.4.9 Combination

In Table 6.6 all determined systematic uncertainties are quoted. For the hadronisation effect the listed value corresponds to the statistical uncertainty on the $\sigma_{semi-\mu}$ bias, because the shift itself was found insignificant.

In the combination of the individual components of the total systematic uncertainty, the correlation between the effects was not accounted for; all contributions are quadratically summed. It is observed that the systematic uncertainty dominates by far the statistical uncertainty, even for a relatively limited amount of data ($1\,\text{fb}^{-1}$). The precise measurement of the semi-muonic $t\bar{t}$ production cross-section fully relies on the ability to reduce the b-tagging, miscalibration/misalignment, JES smearing and in particular the process background systematic.

The question can be posed how these main systematics behave if a different cut-value on the combined likelihood ratio value $LR^{signSel}_{comb}$ had been chosen. These dependencies are summarised in Figure 6.17. It is observed that cutting harder on the LR-discriminant allows an impressive reduction of the b-tagging efficiency and channel background systematic uncertainty, but on the contrary results in an increase of the miscalibration/misalignment systematic, the JES smearing systematic and the statistical uncertainty. All these dependencies are understood. Increasing the likelihood ratio discriminant threshold results in:

- a higher suppression of the number of channel background events, and hence an increase of the sample's purity $P^{t\bar{t}}_{sim}$. Consequently, the uncertainty on the channel background level has less impact on the relative cross-section uncertainty when the LR-cut is increased.

- in on average lower $P^{jet}_{b1} + P^{jet}_{b2}$ observable values in the LR combination. Hence, events in which the b-quark jets have relatively smaller P^{b}_{jet} values are enhanced when increasing the LR-cut. Consequently, the variation of the P^{b}_{jet}-threshold to check the effect of the uncertainty on the b-tagging efficiency results in a reduced systematic uncertainty on $\sigma_{sem-\mu}$, because in a higher fraction of events the b-quark jet candidates are selected anyhow.

- a decrease of the miscalibration/misalignment systematic effect. Cutting harder on $LR^{signSel}_{comb}$ enhances the fraction of events in which the b-quark jets have a low P^{b}_{jet} values (and hence a higher probability to be a true b-quark jet). Figure 6.16 shows that the relative difference in shape between the miscalibrated/misaligned and nominal b-tagging discriminant distribution increases when moving to low P^{b}_{jet} values. As a consequence, also the systematic effect on the cross-section estimate is expect to increase when the cut on the LR-discriminant is raised.

- a decrease of the JES smearing systematic uncertainty. An analogue argumentation is possible based on the result plotted in Figure 6.14. Increasing the LR-cut will enhance the fraction of events with high P_{χ^2}-values. As this region is observed to be relatively more sensitive to the extra JES smearing, a higher systematic effect is obtained.

- less selected events N^{sel}, what increases the statistical uncertainty.

As soon as the level of the channel background is measured from data and the experimental uncertainties on this level indicate the applied rescaling factor of five can be lowered, the total uncertainty on the semi-muonic $t\bar{t}$ production cross-section measurement will drop significantly. In this case other systematic effects like the ones arising from the b-tagging efficiency, detector miscalibration/misalignment and the JES smearing uncertainty become more relevant. A LR-cut at 0.7 can be considered optimal in this scenario. Additionally, new b-tagging algorithms should be tested, that are based purely on the presence of a reconstructed secondary vertex, and consequently are expected to be more robust towards misalignment [150].

The total relative uncertainty on the developed cross-section estimator for $1\,\mathrm{fb}^{-1}$ of accumulated data is found to be $2.1\,(\mathrm{stat.}) \pm 23.5\,(\mathrm{syst}) \pm 10.0\,(\mathrm{lumi})\,\%$. The top quark mass can be inferred from the knowledge of the cross-section of the prosses $pp \rightarrow t\bar{t}$, as discussed in Section 1.2.3. A total uncertainty of 26% on the determination of the cross-section would give an indirect measurement of the top quark mass with an uncertainty of more than $8\,\mathrm{GeV/c}^2$. Hence, in order to actually be able to make a sensitive consistency check of the Standard Model through the comparision of the directly and indirectly measured top quark mass, an important reduction of this total uncertainty is required.

	Uncertainty $\Delta\sigma_{semi-\mu}$
limited MC-statistics	1.6%
Pile-up collisions	5.5%
Underlying event	1.2%
Perturbative QCD radiation	5.5%
Hadronisation	0.6%
Parton Density Functions	3.7%
Jet Energy Scale Shift (5%)	2.8%
Jet Energy Scale Smearing	6.7%
b-tagging	9.0%
Misalignment/Miscalibration	9.1%
Channel background level (×5)	16.1%
Total systematic effects	23.5%
statistical uncertainty for 1 fb^{-1}	1.3%
Integrated Luminosity uncertainty for 1 fb^{-1}	10%
Total uncertainty for 1 fb^{-1}	26.0%

Table 6.6: Overview of the investigated systematic influences on the $\sigma_{semi-\mu}$ estimator evaluated after a cut at $LR^{signSel}_{comb}$=0.7 .

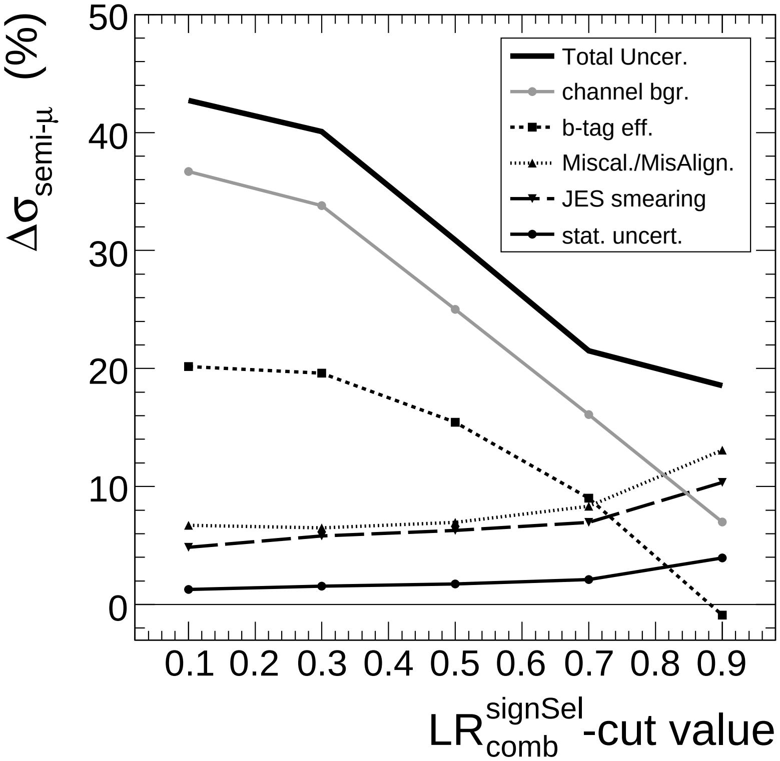


Figure 6.17: The most important systematic uncertainties on the $t\bar{t}$-cross-section estimate evaluated for several $LR^{signSel}_{comb}$ cut values. Also the statistical uncertainty is plotted, as well as the quadratic sum of all five contributions.

Chapter 7

Measurement of the Top Quark Mass

One of the most important properties of the top quark which is at the heart of the Standard Model is its pole mass. The extremely high measured value for this parameter leads to large Yukawa couplings, and consequently to the general assumption that the top quark might play an important role in the mechanism of electroweak symmetry breaking (EWSB).

The top quark mass is currently being measured by the CDF and DØ experiments at the Tevatron Collider in Fermilab. Today, the combination of the best top quark mass measurement of both experiments result in a value of $M_{top} = 170.9 \pm 1.8\,\mathrm{GeV/c^2}$ for $1\,\mathrm{fb}^{-1}$ of accumulated data [35].

Also at the Large Hadron Collider, the precise measurement of the top quark mass is one of the main physics goals. The enormous amount of $pp \rightarrow \mathrm{t\bar{t}}$ events in the semi-leptonic decay mode $\mathrm{t\bar{t}} \rightarrow q\bar{q}b\bar{b}\mu\bar{\nu_{\mu}}$ will allow to measure the top quark mass with an unprecedented accuracy, by far dominated by systematic uncertainties. While the Tevatron is aiming to determine M_{top} with a combined uncertainty of $1.5\,\mathrm{GeV/c^2}$, both the ATLAS and CMS collaboration envisage a precision better than $1\,\mathrm{GeV/c^2}$. This increase in knowledge will lead to a significant improvement in our understanding of the Standard Model and the EWSB mechanism.

This chapter will elaborate on the accurate measurement of the top quark mass based on $1\,\mathrm{fb}^{-1}$ of integrated luminosity, and is based on the results presented in [151]. As an input to this analysis, all events remaining after the cut on the $LR^{signSel}_{comb}$ observable will be used (cfr. Table 6.3).

In a first section, the procedure to handle the ambiguity in the selection of the three hadronic top quark decay products is defined. Another time the information of sensitive observables is merged by means of a Likelihood Ratio technique, and the probability a given jet combination is the correct one is estimated.

With this additional knowledge, all building blocks are introduced to start the description of the actual top quark mass estimators in Section 7.2. For each event, the fitted top mass resolution function or event ideogram is constructed and convoluted with a theoretical template function depending on the true top quark mass M_{top}. Via the Maximum Likelihood method the final top quark mass estimate is obtained. Also, some resampling tests are performed to check the robustness and linearity of the different estimators.

Finally, an extended study of the influence of systematic uncertainties on the measured M_{top}-value is described. It will be motivated that the leading uncertainty on M_{top} arises from the imperfect knowledge of the b-quark jet energy scale.

7.1 Choosing the Correct Jet Combination

In Section 5.1 the strategy was defined to choose the hadronic $\mathrm{t\bar{t}}$-decay products among the jets remaining after the jet and b-tag requirements. The two selected jets with the lowest P_b^{jet}-value and hence the highest probability to originate from the decay of a b-quark, were postulated to be the b-quark jets coming from the top and anti-top decay. Among the remaining selected jets the two jets with the highest transverse energy were on their turn assumed to be the light quark jets produced in the hadronic W-boson decay. The only ambiguity remaining after these choices is which of both b-quark jets belongs to the hadronic part of the semi-muonic $\mathrm{t\bar{t}}$-decay. This can indeed be considered as the only uncertainty in the allocation of the selected jets, as the reconstructed kinematics of both the hadronic W-boson and top-quark are independent of a permutation of both light quark jet candidates.

In this section, a likelihood ratio method is constructed in order to select the correct jet combination among both possibilities. Only semi-muonic events passing the reconstructed objects selection cuts presented in Section 4.6 and the $LR_{comb}^{signSel}$-cut mentioned in Section 6.2 were used. The light quark jet energy scale was corrected for an overall bias in the reconstructed W-boson mass, according to the method described in Chapter 5. Concrete, an inclusive jet energy scale correction factor ΔC =-13.4% was applied only to the light quark jet candidates. The energy scale of the b-quark jet candidates is unchanged. The reconstructed hadronic W-boson mass spectrum after this correction is illustrated in Figure 7.1, while the uncorrected distribution was shown in Figure 5.1. It is demonstrated that a Gaussian fit on the corrected m_W-distribution results in a mean value close to the world average M_W-value. The small residual shift with respect to the true M_W-value of $80.403\pm0.029\,\mathrm{GeV/c^2}$ is a consequence of the extra $LR_{comb}^{signSel}$-cut. Furthermore, the application of the light quark JES correction to the same events that will be used for the top quark mass estimate will induce an important cancelation of light quark JES systematic uncertainty on the measured M_{top} value. We will come back to this when discussing the systematics on the top mass measurement in Section 7.3.

7.1.1 Sensitive observables

Analogue to the way the selected event sample was enriched by signal events in Section 6.2, the most important aspect in the definition of a powerful Likelihood Ratio (LR) discriminant is to find observables able to separate 'signal' from 'background' contributions. For the selection of the correct jet combination, the following observables were found to be sensitive:

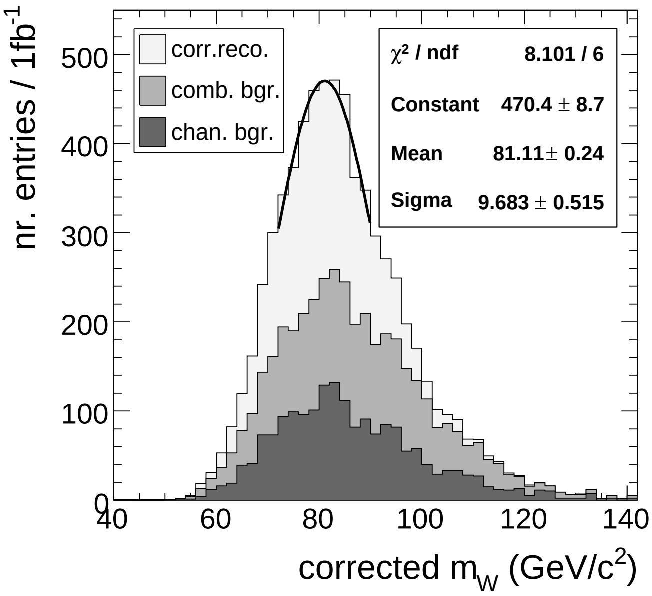


Figure 7.1: For all events passing the reconstructed object selection and the $LR_{comb}^{signSel}$-cut, the distribution of the reconstructed hadronic W-boson mass spectrum after applying the light quark JES correction obtained in Chapter 5.

- relative transverse momentum of the reconstructed hadronic top with respect to the mean value of both possible jet combinations (“$p_T^{had.\,top} / < p_T^{had.\,top} >$”)
- space angle between the reconstructed muon and the leptonic b-quark jet candidate (“$\alpha\,(\mu, b^{lep})$”)
- space angle between the reconstructed hadronic W-boson and the hadronic b-quark jet candidate (“$\alpha\,(W^{had}, b^{had})$”)

For each observable the inclusive p.d.f. for both correct and wrong jet combinations is shown in the left plots of Figure 7.2, while the $R_i(x_i)$ distributions (Eq. 6.4) with their respective fit functions are shown on the right. Only semi-muonic decaying signal $t\bar{t}$-events were employed in the construction of these pdf's. To further improve the discrimination power of the observables, the ‘signal’ histogram was only filled if for all the relevant jets to calculate the observable's value a well matching parton was found ($\alpha_{jp} < 0.2$). Otherwise the contribution is considered ‘background’.

The probability density functions (p.d.f.'s) for the first observable learn that on average a relatively higher reconstructed transverse momentum of the hadronic top quark is obtained for the correct jet combination. This will be particularly true for top quarks produced with a high p_T, of which the decay products received an important Lorentz-boost. As a consequence, also the reconstructed angles between the top quark decay products are expected to be lower for the correct jet assignment in these events. The p.d.f. distributions for all angular observables support this statement. Clearly, the jet combination leading to the smallest reconstructed $\alpha\,(\mu, b^{lep})$ or $\alpha\,(W^{had}, b^{had})$ value is observed to be more likely to be the correct choice. It is therefore not surprising

c_{ij}	$p_T^{had.\,t}$	$\alpha\left(\mu, b^{lep}\right)$	$\alpha\left(W^{had}, b^{had}\right)$
$p_T^{had.\,t}$	1	-0.34	-0.35
$\alpha\left(\mu, b^{lep}\right)$	-	1	0.11
$\alpha\left(W^{had}, b^{had}\right)$	-	-	1

Table 7.1: Correlations between the jet combination sensitive observables, estimated from the correct jet permutation and assuming a linear correlation.

to obtain in Table 7.1 rather large mutual correlation factors between the selected observables. As a consequence, an important correlation is expected between the final likelihood ratio discriminant value and e.g. the reconstructed transverse energy of the hadronic top quark.

7.1.2 Combining the information in a Likelihood Ratio discriminant

Starting from the discrimination power delivered by the individual sensitive observables, the same procedure will be followed to combine all information into a likelihood ratio discriminant as the one described in Subsection 6.2.2. Here, the combined LR value is defined by:

$$LR_{comb}^{jetComb} = \frac{LR'}{1 + LR'}, \tag{7.1}$$

with LR' defined in Eq. 6.8. In this case the addition of the ad-hoc term to account for the non-zero mutual correlations between the observables will be more important to somehow unweight the highly correlated observables. Where in the case of the signal enhancement LR-method the highest correlation factor found between two observables was only 9%, Table 7.1 now indicates a maximum of 35%. To illustrate the validity to assume linear correlations, the correlation between the $p_T^{had.\,t}$ and $\alpha\left(W^{had}, b^{had}\right)$ observable is drawn in Figure 7.3.

In Figure 7.4 the resulting $LR_{comb}^{jetComb}$ distribution is shown for both the correct and wrong jet combination. As only two possible jet combinations remain after the choice of light and b-quark jets, the same number of entries are filled to the 'signal' and 'background' histogram. The correct jet assignment is defined as the jet combination with the lowest sum of space angles between the b-quark jet candidates and the generated b-quarks. This plot shows the separation between correct and wrong jet solutions. However, it should be noted that still many events contain wrongly reconstructed or selected jets, what somehow limits the overall discrimination power of the $LR_{comb}^{jetComb}$-observables.

The correlation between the average transverse momentum of the hadronic top quark and the obtained $LR_{comb}^{jetComb}$-value is illustrated in Figure 7.5. In this plot this average p_T is demonstrated to ascend fast for increasing LR-values. This behaviour was already predicted when discussing the sensitive observables. Consequently, high-p_T (anti-)top quarks will be more likely to obtain a relatively high combined Likelihood Ratio value due to their boosted (anti-)top quark decays.

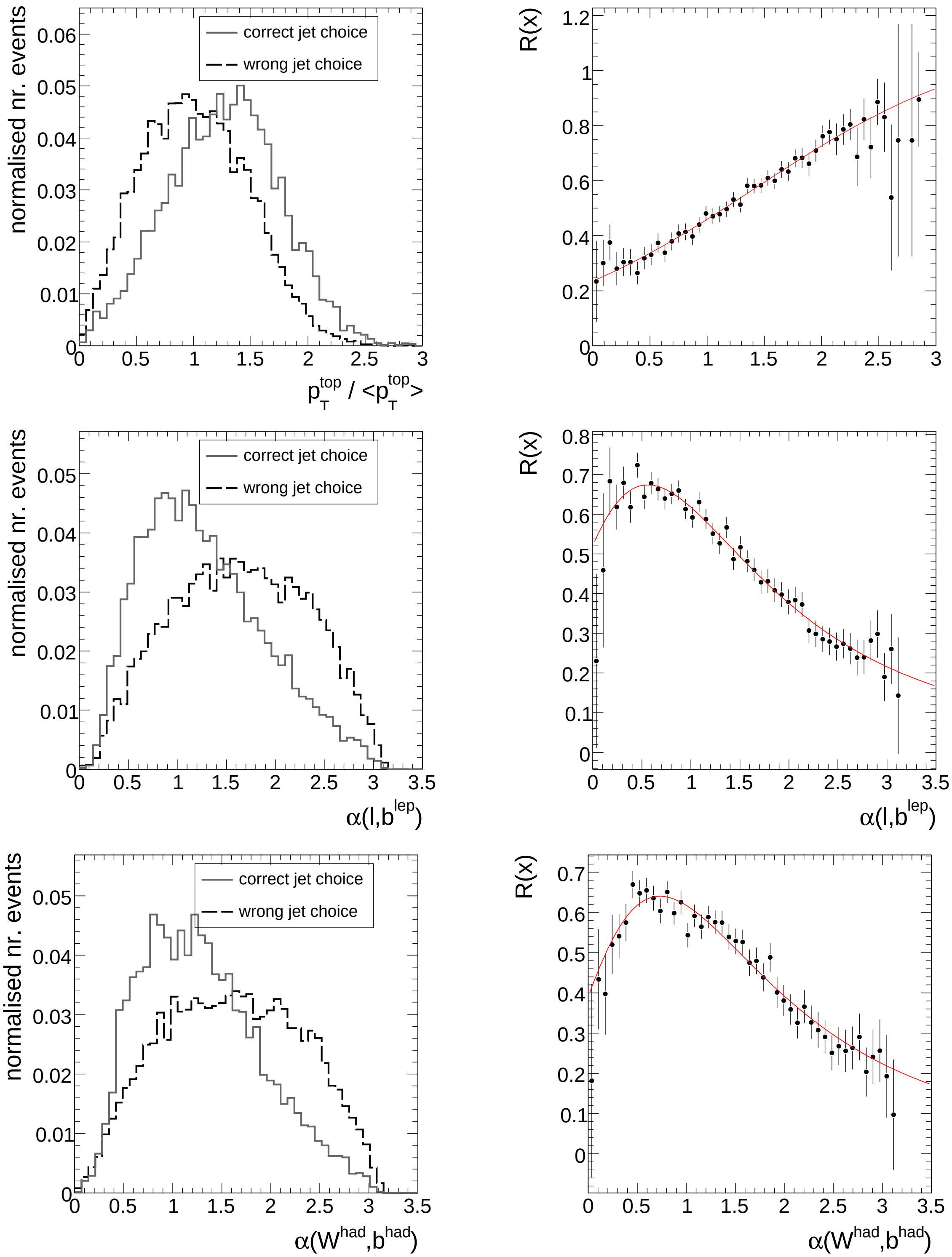


Figure 7.2: Distributions of the jet combination sensitive observables, along with their respective fitted $R_i = P_i^S/(P_i^S + P_i^B)$ distribution.

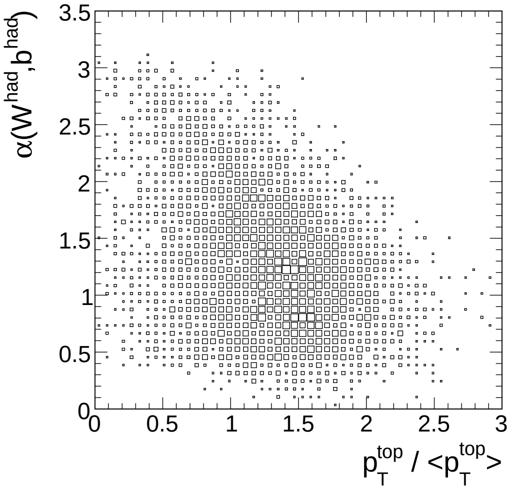


Figure 7.3: The linear correlation between the $p_T^{had.\,t}$ and $\alpha\left(W^{had}, b^{had}\right)$ observable, extracted from the correct jet permutations.

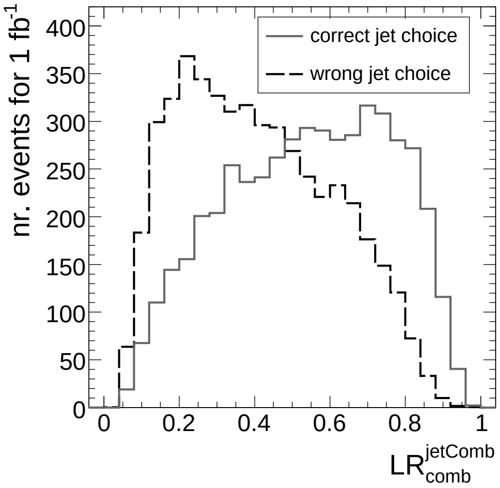


Figure 7.4: The combined LR discriminant value distributions for both correct and wrong b-quark jet combinations in semi-muonic decaying $t\bar{t}$-events.

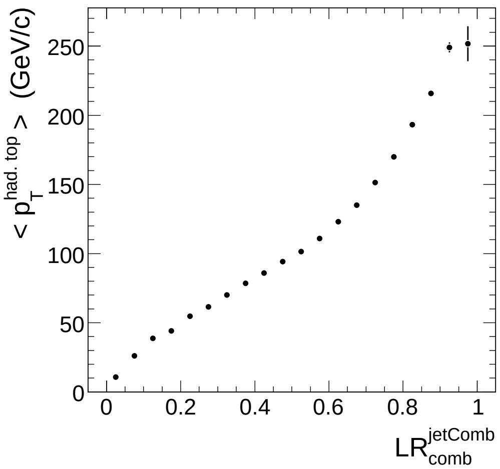


Figure 7.5: The average reconstructed hadronic top quark p_T versus the combined Likelihood Ratio value for correct jet assignments.

7.1.3 Performance

For each event, the jet combination with the highest $LR_{comb}^{jetComb}$-value is chosen as the correct jet association. Once this decision is made, it is important to have an estimate on how certain one may be about this choice. Or differently quoted, if in an event a maximum Likelihood Ratio value of x is found among the two jet combinations, what is the probability that the jet association corresponding to x is the correct one? To estimate this probability, another loop over the selected Monte-Carlo signal events is performed, resulting in the plot in Figure 7.6. This plot shows the distributions for the highest $LR_{comb}^{jetComb}$-value in the event. The solid grey distribution is obtained for

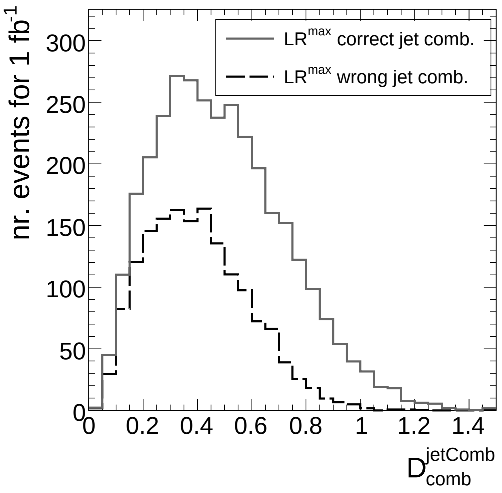


Figure 7.6: The $D^{jetComb}_{comb}$-distribution for the jet combination in an event with the highest $LR^{jetComb}_{comb}$-value, both for events where this choice was correct or wrong.

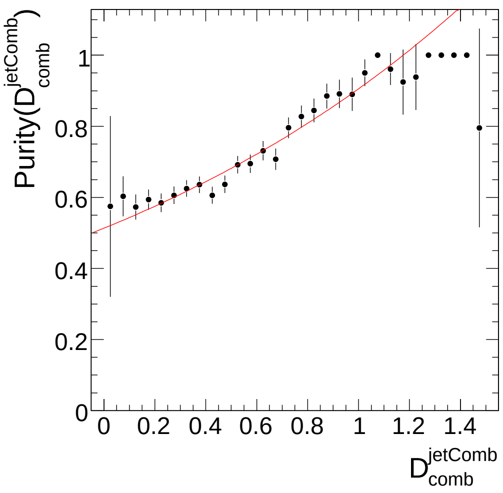


Figure 7.7: Bin-by-bin calculation of the purity the chosen jet combination corresponds to the correct jet assignment and well reconstructed hadronic top quark jets.

events where the highest LR-value corresponds to the correct jet combination, while wrong choices are represented by the black dashed histogram. In order to improve the histogram binning, the discriminator

$$D^{jetComb}_{comb} = -\log(1 - LR^{jetComb}_{comb}) \tag{7.2}$$

is plotted instead of the $LR^{jetComb}_{comb}$ observable. Nevertheless, this X-axis projection conserves the ranking of the original LR-values.

In 67% of the selected events the jet combination with the highest LR-value corresponds to the combination which matches best the generated quarks. On event-by-event basis, the probability for a correct choice, $P^{jetComb}$, increases significantly for an increasing discriminator value, as is observed in Figure 7.7. In order to reduce the amount of combinatorial background contributions in the later top quark mass measurements, a minimal $P^{jetComb}$ of 70 % is required. This value corresponds to a minimal $D^{jetComb}_{comb}$ discriminant value of 0.55. Additionally, Figure 7.8 shows the relation between the event sample's purity and event selection efficiency if an increasing cut on $D^{jetComb}_{comb}$ is applied. The dashed line indicates the point in the graph corresponding to the applied cut value 0.55. At this point, 68% of the events remain with an sample purity of 79%.

The event selection efficiencies after this extra cut are listed for all considered event subsamples in Table 7.2. For the semi-muonic jet multiplicity bins a constant reduction factor of 68% is obtained after this extra requirement, resulting in an overall selection efficiency of almost 2%. In 1 fb^{-1} of accumulated data, around 1577 semi-muonic signal events will pass all cuts. The fraction of channel background events in the remaining event sample is even more reduced, reflected by the increasing S/N ratio to an ultimate value of 6.79. In particular the $W(bb)$+jets events are almost all rejected. Only 12 such

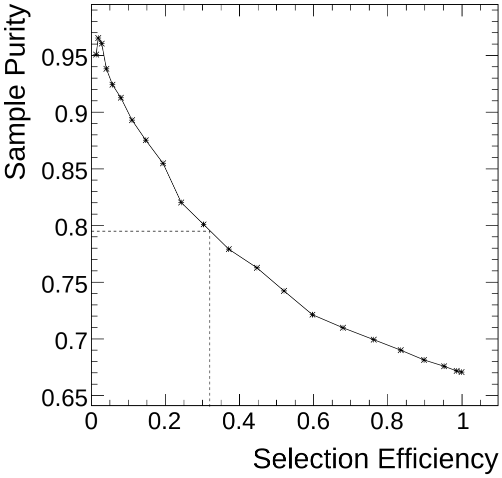


Figure 7.8: The event sample purity versus the event selection efficiency after an increasing cut on the $D_{comb}^{jetComb}$-discriminator. The point corresponding to a cut at $P^{jetComb}$ >70% (or $D_{comb}^{jetComb}$ >0.55) is indicated.

events are expected for $1\,\mathrm{fb}^{-1}$. When we consider the other $\mathrm{t\bar{t}}$-decays as signal events, the S/B-ratio becomes 150.

The effect of this extra event selection cut is visualised in Figure 7.9. In both plots the distribution of the fitted[1] mass of the hadronically decaying top quark, m_{top}^{fit}, is plotted for an integrated luminosity of $1\,\mathrm{fb}^{-1}$. A differentiation is made between correct reconstructions, combinatorial background and contributions remaining from $\mathrm{t\bar{t}}$-decays different from the semi-muonic channel. Only in the histogram on the right the extra requirement on $P^{jetComb}$ is imposed. This clearly results in an important reduction of the amount of combinatorial and channel background events. The use of this extra event information leads to an enhanced fraction of correct reconstructions. After the cut the contribution of events with the correct hadronic jets selected increases from 70% to 80% in a mass window of $40\,\mathrm{GeV/c^2}$ around the highest bin. Also the width of a Gaussian fit on the spectrum improves with the $P^{jetComb}$-cut.

From the result plotted in Figure 7.4 it is possible to define a probability that a set of three jets did originate from the decay of the hadronic top quark, given the event is a semi-muonic signal event: ‘P[correct reco | signal]’. Via the bin-by-bin calculation of the S/(S+B) ratio, illustrated in Figure 7.7, an estimate is obtained of this probability as a function of the calculated discriminant value. In the next section this information will be further employed in the construction of an event weight.

---

[1] A standard kinematic fit was applied to impose the hadronic W-boson mass constraint to the reconstructed event kinematics. For more details we refer to Section 6.1.

sample	after $LR^{signSel}_{comb}$	# after $P^{jetComb}$	reduction	overall sel.eff. (%)
signal				
$t\bar{t}0j$ (excl)	(1716±25)	(559±15)	-67%	1.99±0.05
$t\bar{t}1j$ (excl)	(1657±25)	(530±15)	-68%	2.10±0.06
$t\bar{t}2j$ (excl)	(921±19)	(306±11)	-67%	2.10±0.08
$t\bar{t}3j$ (excl)	(290±13)	(95±8)	-67%	1.61±0.13
$t\bar{t}4j$ (incl)	(274±13)	(87±7)	-68%	0.97±0.08
incl. semi-μ	(4858±44)	(1577±26)	-68%	1.90±0.07
background				
$t\bar{t}0j$ (excl)	(133±7)	(43±4)	-67%	0.027±0.003
$t\bar{t}1j$ (excl)	(236±10)	(65±5)	-73%	0.044±0.004
$t\bar{t}2j$ (excl)	(232±10)	(55±5)	-76%	0.065±0.006
$t\bar{t}3j$ (excl)	(111±8)	(23±4)	-79%	0.068±0.011
$t\bar{t}4j$ (incl)	(142±9)	(35±5)	-75%	0.068±0.009
incl. other-$t\bar{t}$	(854±20)	(220±10)	-74%	0.096±0.005
$W4j$ (excl)	(7±4)	(3±3)	-65%	0.001±0.001
$W5j$ (excl)	(20±4)	(4±2)	-82%	0.008±0.004
$W6j$ (incl)	(13±4)	(3±2)	-75%	0.010±0.007
$Wbb2j$ (excl)	(23±2)	(3±1)	-87%	0.018±0.004
$Wbb3j$ (excl)	(11±2)	(0±0)	-100%	0.±0.
incl. $W(bb)$	(73±8)	(12±4)	-83%	0.009±0.003
S/B	(5.24±0.13)	(6.79±0.34)		

Table 7.2: Overview of the number of events before and after applying the extra $P^{jetComb}$ cut for all studied `AlpGen` samples. All indicated statistics are calculated for an integrated luminosity of 1 fb^{-1}. The third column gives the relative reduction factors for each sample, while in the last column the overall event selection efficiency is listed after all cuts applied.

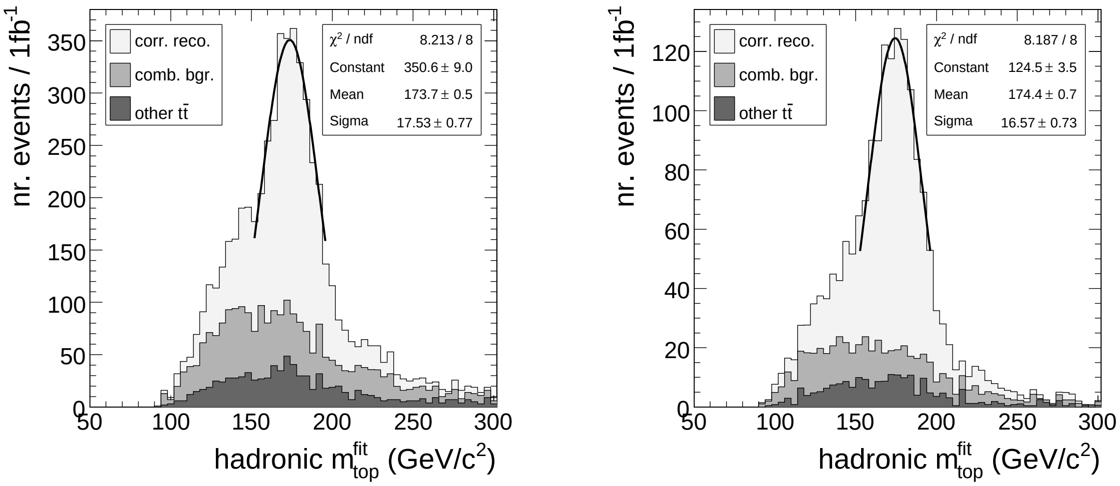


Figure 7.9: The fitted hadronic top quark mass distribution for an integrated luminosity of 1 fb^{-1} (left). Only events passing the $P^{jetComb} > 0.7$ cut are used in the right plot.

7.2 Defining the different Top Quark Mass Estimators

In this section three different estimators of the theoretical top quark mass will be introduced. While the first measurement is obtained via a simple Gaussian fit on the inclusive hadronic top quark mass spectrum, an event-by-event ‘ideogram’ approach was applied for the other two. In a way, the first measurement will serve as a reference to quantify the improvements in robustness and total uncertainty of the ideogram based estimators. Also, it should be remarked that, due to the negligible amount of $W(bb)$+jets events left after the full event selection (cfr. Table 6.3), these channel background contributions are no longer considered in the following.

7.2.1 Gaussian fitted hadronic m_{top} distribution

For all events remaining after the reconstructed objects selection cuts and the cut on $LR_{comb}^{signSel}$, the best jet combination was defined as the event solution with the highest $D_{comb}^{jetComb}$-value. If for this jet combination the discriminator value corresponds to a $P^{jetComb}$ higher than 70%, the event is selected for further analysis. Hence, the described procedure leads to a unique value for the fitted hadronic top quark mass for each selected event.

From the hadronic m_{top}^{fit} distribution shown in the right plot of Figure 7.9, a top quark mass estimator can now be defined as the mean value $\hat{M}_{top}$ of a Gaussian function, $G(m_{top}|\hat{M}_{top})$, fitted around the peak of the histogram. Analogue to the method used in Section 5.2 to fit the W-boson mass spectrum, the full fit range is chosen that value between 40 and 60 GeV/c^2 which returns a minimal $|\chi^2/ndf - 1|$-value. Ultimately, a value of $\hat{M}_{top} = 174.37 \pm 0.73$ GeV/c^2 was found for 1 fb^{-1} of data in an optimal m_{top}-window of 44 GeV/c^2 around the modal bin. This illustrates that already for a relatively limited amount of analysed data it is possible to achieve a small statistical uncertainty. Also the bias with respect to the generated top quark mass of 175 GeV/c^2 is acceptable given the sensitivity of the estimator to e.g the jet energy calibration. In this context the importance of the light quark JES correction factor (cfr. Chapter 5) that is applied to the light quark jet candidates should be mentioned. When describing the main systematic effects on the top quark mass measurements we will come back to this.

7.2.2 Construction of the Event Ideogram

Rather than developing top quark mass estimators on samples of events, an event-by-event likelihood technique is pursued in this work. The method starts with the definition of an event ideogram or likelihood function [152]. It reflects the relative compatibility of the reconstructed event kinematics with the hypothesis that one heavy object with mass m_{top} decays into three jets of which two originate from the decay of a W-boson.

In a first approach this ideogram function is assumed to be Gaussian around the fitted hadronic top quark mass m_{top}^{fit}, and with a width $\sigma_{m_{top}^{fit}}$, equal to the uncertainty on the fitted top quark mass. This uncertainty can be determined via error propagation starting from the covariance matrices on the kinematics of the three fitted jets. The resulting ideogram function $I(\overline{p}_j|m_{top})$ can then be written as:

$$I(\{\overline{p}_j\}|m_{top})\ dm_{top} \ = \ exp\left(-\frac{1}{2}\cdot\chi^2(\{\overline{p}_j\}|m_{top})\right), \tag{7.3}$$

with:

$$\chi^2(\{\overline{p}_j\}|m_{top}) = \left(\frac{m_{top} - m_{top}^{fit}}{\sigma_{m_{top}}^{fit}}\right)^2 . \tag{7.4}$$

$I(\{\overline{p}_j\}|m_{top})$ is hence defined in the space of the reconstructed top quark mass m_{top}. The function expresses how well the measured event kinematics $\{\overline{p}_j\}$ agree with a reconstructed top quark mass m_{top} given the result from the kinematic fit m_{top}^{fit} and the calculated uncertainty $\sigma_{m_{top}}^{fit}$.

Another way to construct this ideogram function is to explicitly scan the probability $P(\{\overline{p}_j\}|m_{top})$ by forcing a reconstructed top quark mass to the event by means of an extra constraint in the kinematic fit setup. In order to cover the whole reconstructed top quark mass range, 340 m_{top}-values were scanned between 85 and 424 GeV/c^2. The hypothesis of a Gaussian ideogram is not needed in this approach, but the computing time is increased by an important factor.

In Figure 7.10 the ideogram function is shown for several selected events, and as an illustration a comparison is made between both approaches. The maximum of the scanned function $P(\{\overline{p}_j\}|m_{top})$ is fixed at unity in order to facilitate the comparison of the obtained ideogram shapes[2]. For each event, the most likely reconstructed top quark mass is found to be identical for both ideogram definitions. This, simply because for the

---

[2] A rescaling of the ideogram function by a constant factor A will leave the ultimately measured top quark mass, as well as its statistical uncertainty, invariant (cfr. Eq. 7.5 and Eq. 7.8).

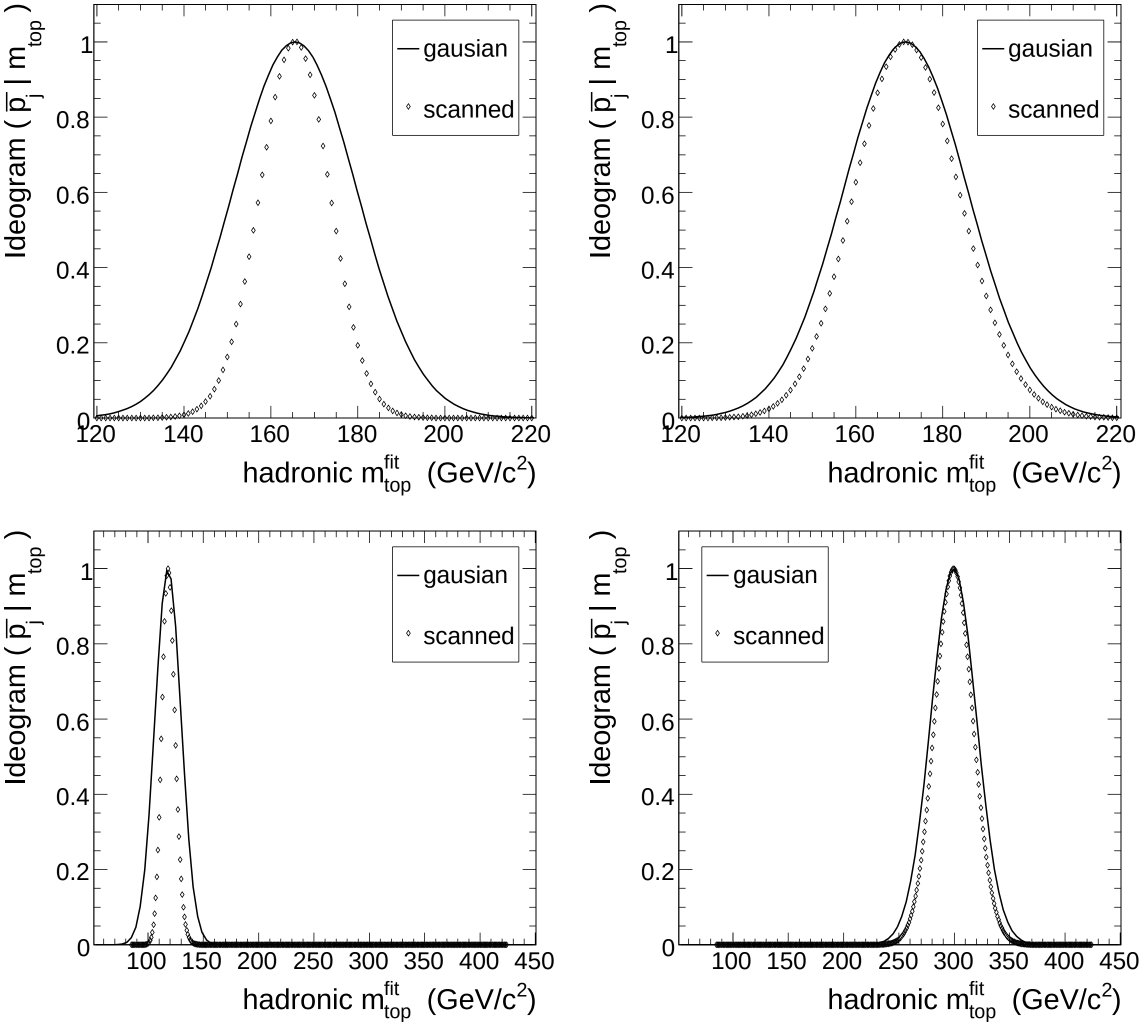


Figure 7.10: For some selected signal $t\bar{t}$ events, the reconstructed Gaussian (full line) and scanned (dashed line) ideogram is compared.

fitted m_{top}^{fit}-value no additional corrections to the measured b-quark jet four-momentum are needed during the kinematic fit, what results in the maximal P_{χ^2}-value. In general, the scanned ideogram is also Gaussian, be it with a width narrower compared to the parametrised ideogram. This can be understood from the fact that a kinematic fit is able to reduce the uncertainties on the initially measured object's kinematics. As only in the case of the fully scanned ideogram the b-quark jet four-vector is included in the fit, the Gaussian ideogram is expected to be more broad than the scanned one. These observations are valid in the whole reconstructed m_{top}-range, as is demonstrated by the bottom plots in Figure 7.10.

The next paragraph will elaborate on the transfer of the reconstructed event information to the theoretical space in which M_{top}, the top quark pole mass, is defined.

This convolution from one space to the other will however be independent on how the ideogram was constructed. As a consequence, both the parametrised and scanned ideogram approach will result in two different top quark mass estimators. In the discussion of the systematic uncertainties in Section 7.3 the advantage of each is illustrated.

7.2.3 Extraction of M_{top} via a Maximum Likelihood technique

To obtain information about the top quark's pole mass M_{top}, the reconstructed ideogram needs to be convoluted with a theoretical expected probability density function or 'template' function $\mathcal{T}(m_{top}|M_{top})$. This transition can be written as:

$$\mathcal{L}_i(M_{top}) = \int \mathcal{I}(\{\overline{p}_j\}|m_{top}) \cdot \mathcal{T}(m_{top}|M_{top})\ dm_{top}\ , \tag{7.5}$$

where one integrates over the kinematic relevant range of m_{top} to obtain a likelihood function $\mathcal{L}_i(m_{top})$ for each event i in the true top quark mass dimension.

Several contributions should be added in the expected template $\mathcal{T}(m_{top}|M_{top})$: a term for the correct jet combinations $\mathcal{S}(m_{top}|M_{top})$, one for the combinatorial background $\mathcal{B}_{comb}(m_{top})$ and a process background contribution $\mathcal{B}_{proc}(m_{top})$. In case of the correct jet assignments the theoretically expected shape is a Breit-Wigner function [153]:

$$BW(m_{top}|M_{top},\Gamma_{top}) = \frac{1}{2\pi}\frac{\Gamma_{top}}{(m_{top}-M_{top})^2+(\Gamma_{top}/2)^2} \equiv \mathcal{S}(m_{top}|M_{top}), \tag{7.6}$$

which is dependent of the pole mass M_{top} and the top quark's decay width Γ_{top}. For this latter parameter, a value of $1.4\,\mathrm{GeV/c^2}$ is taken, as predicted by the Standard Model for a top quark mass of $175\,\mathrm{GeV/c^2}$ and $\alpha_s = 0.118$ [3].

Different than for the signal contribution, Monte-Carlo simulated events are used to parameterise the shapes of the combinatorial and process background contributions. Both probability density functions can easily be extracted from the reconstructed top quark mass distribution shown in the right plot of Figure 7.9. These densities for $\mathcal{B}_{comb}(m_{top})$ and $\mathcal{B}_{proc}(m_{top})$, together with their respective fit function are illustrated in Figure 7.11. The two background contributions are assumed to be independent on the value of M_{top}[3].

For each event and for each value of M_{top} a template function can now be defined as:

$$\begin{aligned}\mathcal{T}(m_{top}|M_{top}) &= P^{signal}\cdot\left[P^{jetComb}\cdot\mathcal{S}(m_{top}|M_{top}) + (1-P^{jetComb})\cdot\mathcal{B}_{comb}(m_{top})\right]\\ &+ (1-P^{signal})\cdot\mathcal{B}_{proc}(m_{top}),\end{aligned} \tag{7.7}$$

with:

---

[3] Although this is an approximation, this still allows the definition of top quark mass estimators with a linear response to a changing Monte-Carlo generated top quark mass. This will be illustrated in paragraph 7.2.4 .

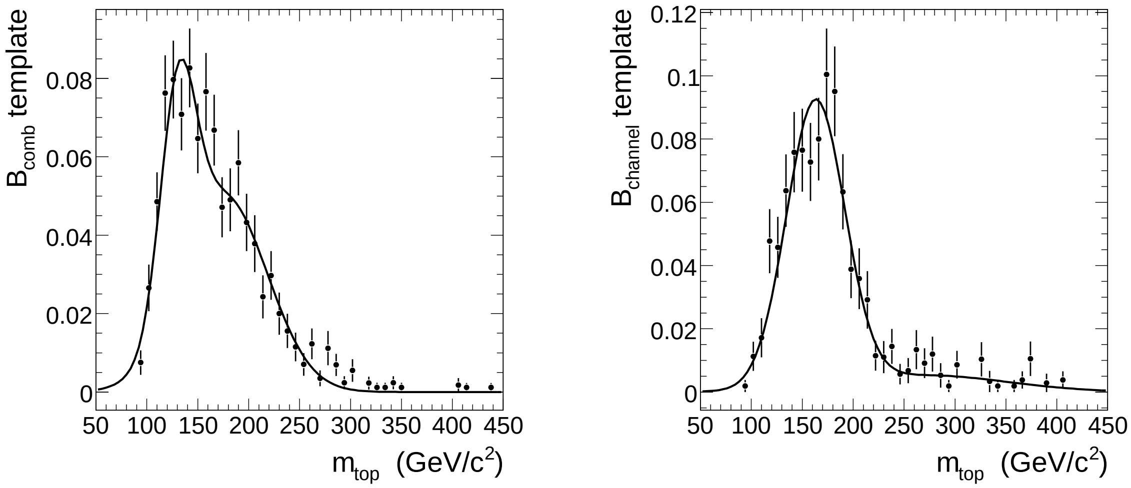


Figure 7.11: The fitted combinatorial (left) and process (right) background density function for all the events passing the $LR^{signSel}_{comb}$ >0.7 cut.

- P^{signal}: The probability that the event is a semi-muonic signal event. For each event, this probability is evaluated from the fit function shown in Figure 6.6 and the $LR^{signSel}_{comb}$-value calculated for the given event.

- $P^{jetComb}$: The probability that the selected jet combination is the one matching best the directions of the generated quarks. Analogue, for each event, this value is deduced using the fit function illustrated in Figure 7.7 and the event's $D^{jetComb}_{comb}$-value.

Both values are hence fixed for a given event, and determine the weights of the signal and background contributions in the combined template function $\mathcal{T}(m_{top}|M_{top})$. Two examples of such a template function are drawn in Figure 7.12. Where in the left plot an event with a high probability to be well reconstructed is considered, the right plot illustrates the template function obtained for an event just passing the likelihood requirements. In both cases the Breit-Wigner term is calculated for a M_{top}-value equal to 175 GeV/c^2.

Once the template and ideogram function is defined for a given event, Eq. 7.5 allows to calculate the event likelihood function $\mathcal{L}_i(M_{top})$. This one-dimensional likelihood is only function of the parameter to estimate, the true top quark mass M_{top}. It expresses the probability that the measured four-momenta $\{\overline{p}_j\}$ would be produced from the hadronic decay of a top quark with a pole mass value M_{top}. The probability that a set of N selected events with kinematics $\{\overline{p}^1_j\}$, $\{\overline{p}^2_j\}$, ... $\{\overline{p}^N_j\}$ are detected for a given M_{top}-value then becomes the product of the individual event likelihoods $\mathcal{L}_i$:

$$\mathcal{L}(\{\overline{p}^1_j\},\, \{\overline{p}^2_j\},\, \dots \{\overline{p}^N_j\};\; M_{top}) \;=\; \prod_{i=1}^{N} \mathcal{L}_i(M_{top}) \;\equiv\; \mathcal{L}(M_{top}) \;. \tag{7.8}$$

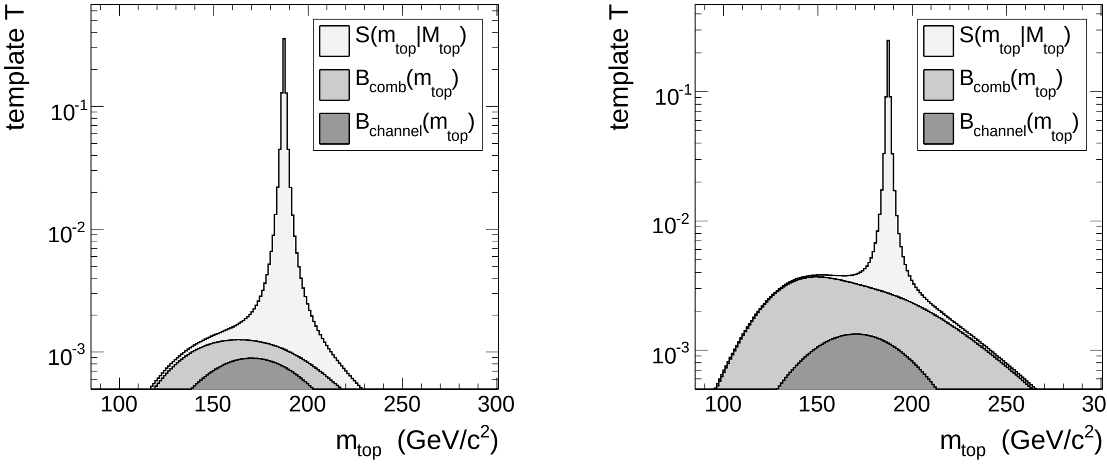


Figure 7.12: A template function $\mathcal{T}(m_{top}|M_{top})$ for a very well reconstructed event (left) and an event more likely to be combinatorial background (right). The different contributions to the template function are visualised.

The maximum likelihood (ML) technique [143] can now be employed to estimate the most likely top quark mass from the selected event sample. According to this method, the ML estimator $\hat{M}_{top}$ is the value of M_{top} for which the likelihood $\mathcal{L}$ is a maximum. In practise it is easier to maximise the logarithm of $\mathcal{L}$.

$$\left.\frac{d\ln\mathcal{L}}{dM_{top}}\right|_{M_{top}=\hat{M}_{top}} = 0 \tag{7.9}$$

This ML estimator has the properties to be consistent and efficient. Moreover, in the limit $N \to \infty$ the variance of $\hat{M}_{top}$ is equal to the minimum variance bound. From this fact, combined with the central limit theorem, it can be derived that the likelihood function is expected Gaussian over the relevant range of M_{top} close to $\hat{M}_{top}$, while the log likelihood is a parabola. The standard deviation of this Gaussian is an estimate of the uncertainty of the ML estimator $\hat{M}_{top}$. As a consequence, at the point 1σ away from the peak, the log likelihood value has increased by 0.5 from its minimum. This attribute is employed in the estimation of the statistical uncertainty on the top quark mass estimates.

In Figure 7.13 this strategy is demonstrated. Rather than the total likelihood $\mathcal{L}$, the χ^2-distribution,

$$\chi^2(M_{top}) = -2\,\ln\mathcal{L}(M_{top})\ , \tag{7.10}$$

is plotted. Evidently, the procedure to estimate the Maximum Likelihood estimator $\hat{M}_{top}$ and its statistical uncertainty stays invariant under this transformation.

For the ~1.8k of events in a $1\,\mathrm{fb}^{-1}$ data-like sample remaining after the jet combination probability cut (cfr. Table 7.2), the above described maximum likelihood technique is applied to estimate the true top quark mass M_{top}. The resulting $\chi^2(M_{top})$ distribution for the Gaussian and scanned ideogram approach are shown in Figure 7.14.

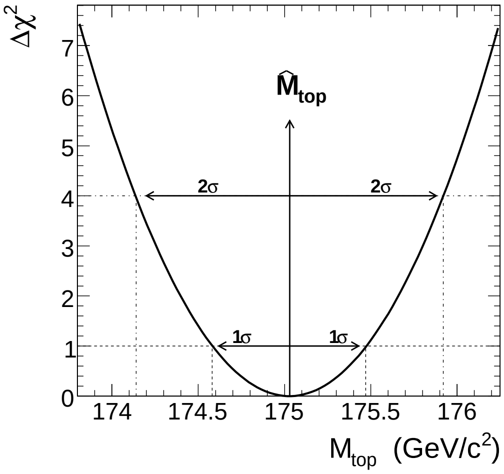


Figure 7.13: Graphical extraction of statistical uncertainty via the $\Delta\chi^2$ distribution of the ML top quark mass estimator.

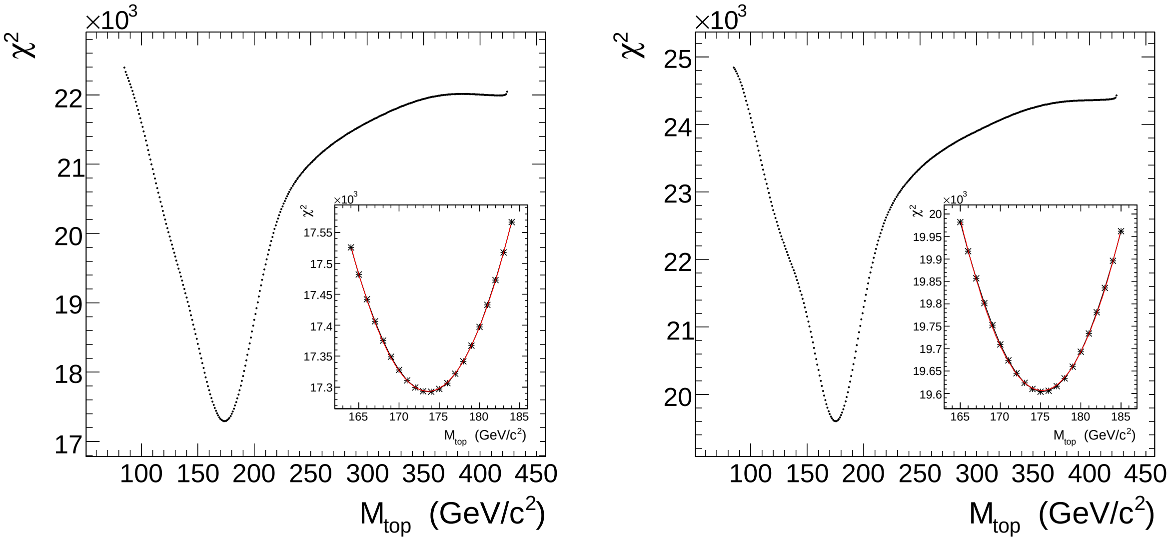


Figure 7.14: χ^2-distribution obtained for the top quark mass estimation based on a Gaussian (left) and the fully scanned (right) ideogram approach.

In both cases a smooth curve is obtained, with a clear and unique minimum in the expected M_{top} region. The insets illustrate the parabolic behaviour of the total likelihood function around the minimum, and consequently, the validity of our way of estimating the statistical uncertainty on M_{top}. Ultimately, the Gaussian ideogram leads to an estimated ML value of $\hat{M}_{top} = 173.07 \pm 0.65\,\mathrm{GeV/c^2}$, while for fully scanned approach this value becomes $\hat{M}_{top} = 174.40 \pm 0.56\,\mathrm{GeV/c^2}$. Hence, apart from the statistical un-

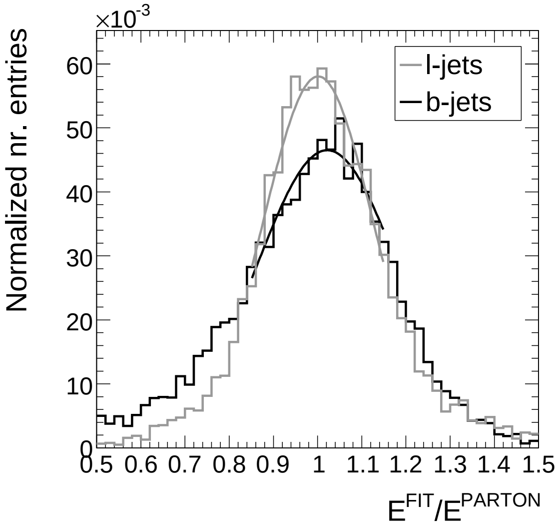


Figure 7.15: The expected JES shift on the input jets of the mass measurement.

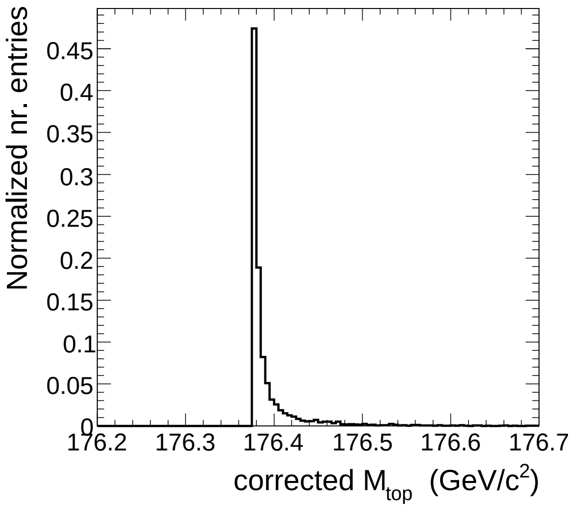


Figure 7.16: Propagation of the 2% b-quark JES overcalibration towards the generated top quark mass.

certainty, also the bias with respect to the generated top quark mass of 175 GeV/c^2 is reduced impressively when the scanned ideogram definition is selected.

To assure ourselves that this reduction in bias is indeed a merit of the ideogram method, the expected bias on M_{top} should somehow be quantified. The main source of such shifts is a miscalibrated jet energy scale. Therefore, the average relative shift on both the light and b-quark jet energy is determined and illustrated in Figure 7.15. Only well matched jets ($\alpha_{jp} < 0.2$) are included from semi-muonic events that pass all previously described event selection cuts. After a Gaussian fit on the spectrum, the light quark JES is observed to have no residual shift with respect to the parton energy[4], while for the b-quark jet energy an overcalibration of 2% was found. In order to estimate an M_{top} bias that takes into account this b-JES miscalibration, an ‘expected’ generated top quark mass can be calculated after applying the same b-JES shift to the b-quarks and recalculate the generated top quark mass. The result is shown in Figure 7.16. Hence, instead of a generated top quark mass of 175 GeV/c^2, the mean ‘unbiased’ M_{top}-value becomes 176.4GeV/c^2. Because this corrected M_{top}-value is even larger than the true generated value, our conclusion that the scanned ideogram method results in the best top quark mass measurement from the bias point of view is still valid.

In the next paragraph some further tests on the estimator’s robustness are described.

7.2.4 Quality and Robustness of the Estimators

As a first check, the bias with respect to the generated top quark mass of 175 GeV/c^2 is reduced impressively when the scanned ideogram definition is selected. Also the

---

[4] This is a direct result of the applied light quark JES correction (cfr. Chapter 5), and especially the event-by-event use of the kinematic fit to fix the light quark jet fourmomenta to the W-boson mass constraint.

statistical uncertainty is observed to be smaller for this technique. Also, the linearity of the estimators has been determined and illustrated for the three top quark mass estimators in the left plots of Figure 7.17. For five different values of the generated top quark mass, identically the same analysis was performed to investigate the estimator's behaviour. Any slope different than unity would translate in a bias depending on the generated top mass, and hence be an indication for the estimator's instability or correlations in the event selection or weighting. All five Monte-Carlo samples were generated with `PYTHIA`, and simulated, digitised and reconstructed using the parametrised fast simulation program. Only semi-muonic events were simulated for this linearity check.

For the measurement relying on the Gaussian fit on the inclusive hadronic top quark mass spectrum, a slope of 0.887 ± 0.025 is found. In the case of the ideogram methods, the slopes are much closer to and compatible with unity: 1.016 ± 0.017 and 1.007 ± 0.002 for respectively the Gaussian and scanned approach.

An analogue conclusion is obtained from the pull distributions obtained via the bootstrapping resampling technique [154], and illustrated in the right plots of Figure 7.17. The bootstrap is a procedure that involves the choice of random samples with replacement out of a general large data set and to analyse each sample in the same way. Sampling with replacement means that every sample is returned to the data set after sampling. So a particular data element from the original data set could appear multiple times in a given bootstrap sample. In this case, for each sample 750 of the remaining events are chosen randomly, and the pull Z^i is calculated:

$$Z^i = \frac{\hat{M}^i_{top} - \overline{M}_{top}}{\delta \hat{M}_{top}^{\ \ i}}. \tag{7.11}$$

Hereby, $\hat{M}^i_{top}$ and $\delta \hat{M}_{top}^{\ \ i}$ denote the M_{top}-estimate and its uncertainty for sample i, while $\overline{M}_{top}$ represents the average M_{top}-value over all samples (20k). If the residuals $\hat{M}^i_{top} - \overline{M}_{top}$ and the uncertainties are well estimated, the pull distribution is expected Gaussian with zero mean and unit width. For all three top quark mass estimators these requirements are found reasonably well fulfilled. In the case of the Gaussian ideogram the pull indicates a 10% overestimation of the statistical uncertainty[5], while in the case of the Gaussian fit on the spectrum the obtained precision is slightly underestimated. However, in general the procedure to determine the uncertainty on M_{top} can be considered reliable for all three estimators.

The exact numbers, together with the expected statistical uncertainty on the top quark mass estimators for $1\,\mathrm{fb}^{-1}$ of integrated luminosity, are shown in Table 7.3. On the last row the uncertainty for each of the top quark measurements is corrected for the pull. After this correction, the ideogram based M_{top}-estimators result in a relatively smaller statistical precision. Also, one should keep in mind the worse linearity for the Gaussian fit on the spectrum, and the relatively large bias and pull correction for the Gaussian ideogram approach. Hence, from statistical point of view, the maximum likelihood measurement based on a scanned ideogram can be posted as optimal.

---

[5] The reason for this low pull is the overestimation of the top quark uncertainty, due to the overestimated b-quark jet resolutions. In the case of the scanned ideogram, the uncertainties on the hadronic b-quark fourmomentum are adjusted in the kinematic fit procedure and consequently better estimated.

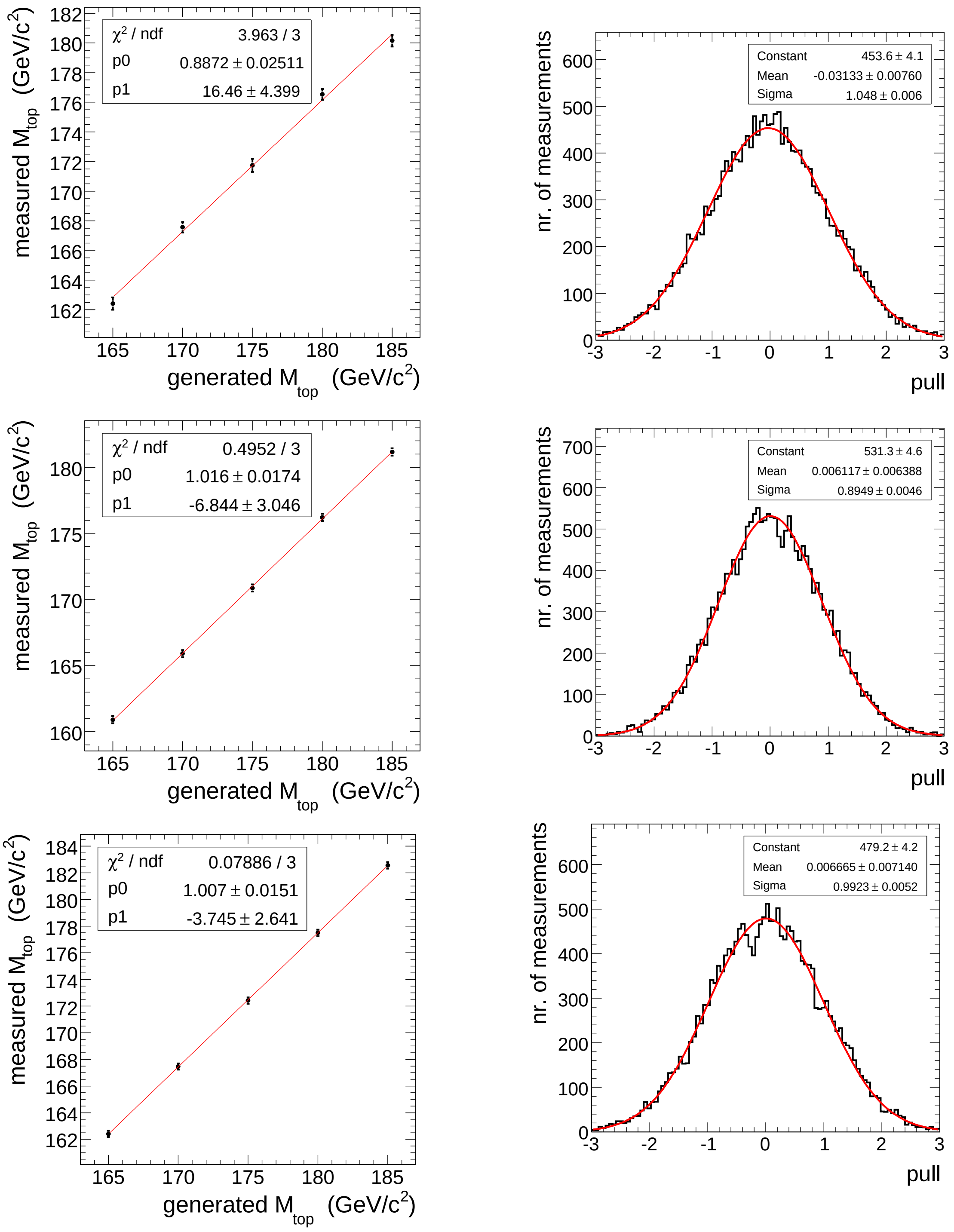


Figure 7.17: For the different top quark mass estimators, the linearity versus the generated top quark mass (left) and the pull distribution (right). The upper plots show the result for the Gaussian fit on the inclusive top quark mass spectrum, while on the second and last row the robustness of respectively the Gaussian and scanned ideogram method are illustrated.

	Gaussian Fit	Gaussian Ideogram	Full Scan Ideogram
bias (GeV/c^2)	0.63	1.93	0.60
uncertainty (GeV/c^2)	0.73	0.65	0.56
slope	0.887	1.016	1.007
width pull-distribution	1.048	0.895	0.992
corr. uncer. (GeV/c^2)	0.73	0.58	0.56

Table 7.3: Overview of the statistical properties of the three top quark mass estimators defined in the text.

7.3 Study of the Systematic Uncertainties on M_{top}

Several systematic effects could induce a shift on the top quark mass estimators. They originate from our understanding of the detector performance, the robustness of the reconstructed objects (jets in particular), and the general description of the proton collisions in the simulation.

In this section exactly the same procedure will be applied to extract the sensitivity of the various top quark mass estimators towards systematic influences as employed in the light quark JES (Section 5.3) and cross-section measurement (Section 6.4). For the three top quark mass estimators, all resulting systematic shifts are given in Table 7.4. If the systematic effect on the top quark mass is compatible with zero within its statistical uncertainty, then for this systematic effect the number is quoted which corresponds to the statistical precision of the test. In these cases the shift is marked by an asterisk in Table 7.4.

Only $t\bar{t}$ events were employed in the calculation of the listed values. Considering the extremely small $W(bb)$ event rate remaining after both LR-discriminant cuts (cfr. Table 7.2), the addition or rescaling of these event contributions has a negligible influence on the proposed top quark mass estimators. Therefore, the systematic effect due to the uncertainty in the channel background level is excluded from Table 7.2.

Furthermore, for CPU reasons only semi-muonic decaying $t\bar{t}$ events were used to calculate the systematic shifts due to the uncertainty in the pile-up collision rate, the PDF's and the `PYTHIA` modelling. With a final event selection S/B-ratio of 6.7, the introduction of the same effect in the background processes would only be a second order effect.

The systematic effects are determined on the three estimators described in the previous section: a simple Gaussian fit on the fitted top quark mass spectrum, a convolution technique with a Gaussian parameterisation of the ideogram and a convolution technique with a full probability scan. In the following we will discuss the main observations and conclusions that can be derived from this study of the main systematic effects.

	Uncertainty on M_{top} (GeV/c^2)		
	Gaussian Fit	Gaussian Ideogram	Scanned Ideogram
Pile-up collisions	0.61	0.78	0.65
Underlying event	0.27	0.20	0.16
Perturbative QCD radiation	0.66	0.50	0.16
Hadronisation	0.40	0.11	0.16
Parton Density Functions	0.36*	0.17*	0.20*
Jet Energy Scale Shift (5%)	3.3	3.8	3.8
Jet Energy Scale Smearing	1.2	0.41*	0.38*
b-tagging	0.43*	0.40*	0.34*
Misalignment/Miscalibration	1.4	0.40*	0.36*
Total systematics (without JES shift)	2.0	1.3	0.96
Total systematics (with JES shift)	3.9	4.0	3.9
Statistical uncertainty for 1 fb^{-1}	0.73	0.58	0.56
Total uncertainty for 1 fb^{-1}	4.0	4.0	4.0

Table 7.4: Overview of the investigated systematic effects on the different top quark mass estimators. Each shift marked with an asterisk corresponds to effects for which the statistical uncertainty on the systematic shift exceeded the determined value of the shift itself. In this case the uncertainty on the shift is listed.

7.3.1 Pile-up collisions

The sensitivity of the different top quark mass estimators towards the inclusion of pile-up collisions was tested. Herefore, the same `PYTHIA` generated event samples, a nominal one and a sample superposed with low luminosity pile-up collisions, were employed. The resulting shifts on M_{top} indicate that a comparable and limited systematic shift can be expected for all three top quark mass estimators. Analogue to the observations for the measurement of the JES correction factor and the $t\bar{t}$ cross-section, Figure 7.18 illustrates the strong dependence of this systematic shift with respect to the number of reconstructed vertices in the event. Only selecting events with one primary vertex would almost completely cancel the effect. From the plot the scanned ideogram technique is clearly identified as the top quark mass estimator with the lowest sensitivity towards the inclusion of pile-up events.

7.3.2 `PYTHIA` modelling and PDF's

The guidelines described in [70] and summarised in Chapter 2 are applied to estimate the sensitivity of a physics measurement towards the theoretical uncertainties in the `PYTHIA` modelling and in the PDF fits. For the evaluation of their systematic effect

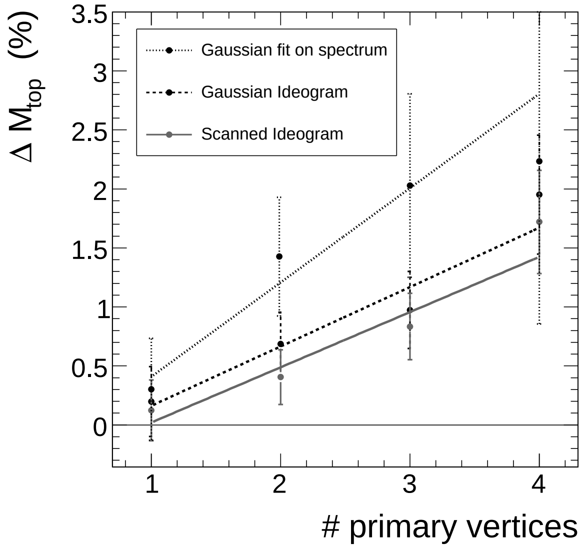


Figure 7.18: The shift on M_{top} due to the inclusion of low luminosity pile-up collisions evaluated as a function the number of reconstructed primary vertices in the event. The result for the three top quark mass estimators is compared.

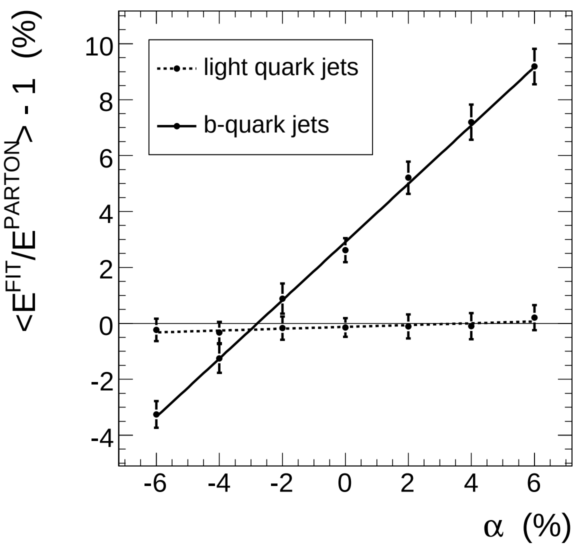


Figure 7.19: The Gaussian fitted mean of the $(E^{FIT}/E^{PARTON}-1)$ distribution versus the applied inclusive JES shift. Only well matched jets were used from signal events that passed all LR criteria.

on the different M_{top} estimations, exactly the same simulated event samples were employed as for the determination of the systematic shifts on the $t\bar{t}$ cross-section (cfr. paragraph 6.4.2). It is observed in Table 7.4 that in general all three measurements are robust to the uncertainties in the `PYTHIA` modelling and PDF fits. This conclusion is in particular true for the scanned ideogram method, while for both other estimates the hard radiation effect is dominant. Again however, it should be noticed that this latter effect is expected to be reduced with the use of the `AlpGen` hard event simulation, as in this generator the hard radiation is included in the matrix element calculation. In any case, from the listed systematic shifts it can be assumed that non of these effects are expected to give rise to a dominant systematic effect on the top quark mass measurement.

7.3.3 Jet energy scale

On the contrary, the top quark mass measurement is extremely sensitive to the uncertainties in the jet energy scale. Analogue to Section 6.4.4, a differentiation is made in this paragraph between an uncertainty on the average and on the spread of the JES measurement.

The analysis response to a shift of the JES by a fixed factor (cfr. Eq. 6.17) will be different for light and b-quark jets. First of all because of the possibility to apply the light quark JES correction factors derived from the W-boson mass constraint imposed on semi-muonic $\mathrm{t\bar{t}}$ data (cfr. Chapter 5). In Figure 6.11 it was illustrated that the application of this calibration method did indeed reduce significantly the impact of a light quark JES shift. As an input to this plot, only well matched and calibrated jets were used.

However, the jet four-momenta, comprised in the jet combinations that are selected for the top quark mass measurements, were adjusted by the kinematic fit procedure (imposing event-by-event the W-boson mass constraint to the two light quark jet candidates). Because of the use of this kinematic fitting technique, the light quarks jet momenta entering the top quark mass measurements are expected to be even more robust towards an uncertainty in the JES. This is illustrated by Figure 7.19, in which the fitted jet four-momenta were used from jet combinations that passed all combined LR criteria. Consequently, the obtained systematic shift on the each of the top quark mass estimators due to a change in the JES should be interpreted as a test of the measurement's sensitivity towards an uncertainty on the b-quark JES.

Analogue to Section 6.4.4 it is presumed that the JES uncertainty will be known to 5% at the time $1\,\mathrm{fb}^{-1}$ of data is collected. This estimate is rough, and depends mainly on our understanding of the detector response. The plots in Figure 7.20 illustrate the correlation between the assumed JES uncertainty and the resulting systematic shift on the different M_{top} estimators. A slightly smaller slope is obtained in the case the top quark mass is inferred from the Gaussian fit on the spectrum. As a result, the assumption of a 5% JES uncertainty leads to a $\Delta M_{top} = 3.2\,\mathrm{GeV/c^2}$ systematic uncertainty in the case of the Gaussian fit, and $3.8\,\mathrm{GeV/c^2}$ for both ideogram approaches. This small but significant difference in sensitivity can be explained by means of both Figure 7.21 and 7.22. The first plot illustrates that the average shift on the fitted hadronic top quark mass m_{top}^{FIT} after a b-quark JES shift of +5% increases with the transverse energy of the hadronic b-quark jet $E_T^{hadr.\,b}$. However, already in Section 4.1.4 the relative jet transverse energy resolution was showed to decrease with increasing E_T (cfr. Figure 4.16). As a consequence, also the uncertainty on the reconstructed top quark mass, δm_{top}^{FIT}, will decrease for higher $E_T^{hadr.\,b}$ values, which is illustrated in Figure 7.22. Because events with a top quark mass contribution with a low uncertainty will have a higher weight in the maximum likelihood technique, the ideogram based top quark mass estimators favour the jet contributions with high $E_T^{hadr.\,b}$-values. Consequently, these methods become more sensitive to the applied b-quark JES shift.

The robustness of the top quark mass estimators with respect to an increase of the JES smearing is more straightforward to understand. Only for the Gaussian fit on the spectrum this effect is found significant ($1.2\pm0.7\,\mathrm{GeV/c^2}$). This is expected, as one of the main advantages of the use of an event ideogram is that it will account for the resolution on the fitted top quark mass. This uncertainty δm_{top}^{FIT} is calculated from the fitted jet covariance matrices. Events that suffer more from the extra JES smearing will be more difficult to fit and will consequently obtain a larger reconstructed top quark mass uncertainty and hence a lower weight in the final top quark mass measurement.

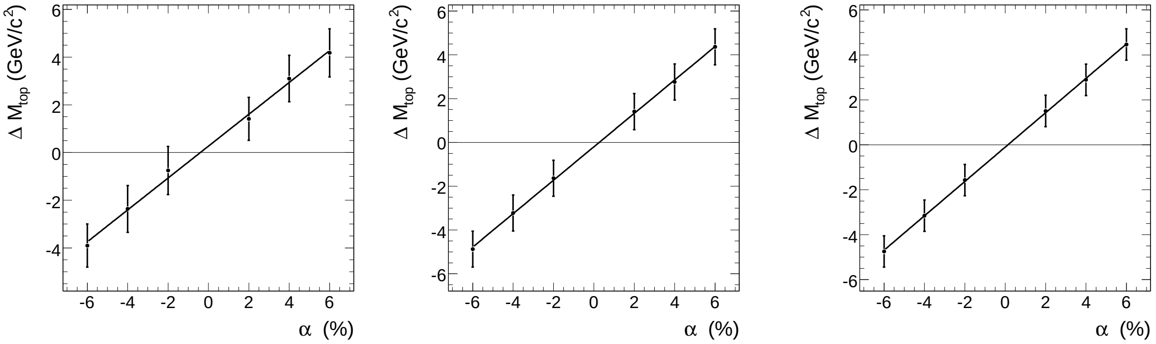


Figure 7.20: The top quark mass estimator's systematic uncertainty for several fixed jet energy scale shifts. The left plot shows the result for the Gaussian fit on the M_{top} spectrum, while for the middle and right plot respectively the Gaussian parametrised and scanned ideogram technique was employed to estimate M_{top}.

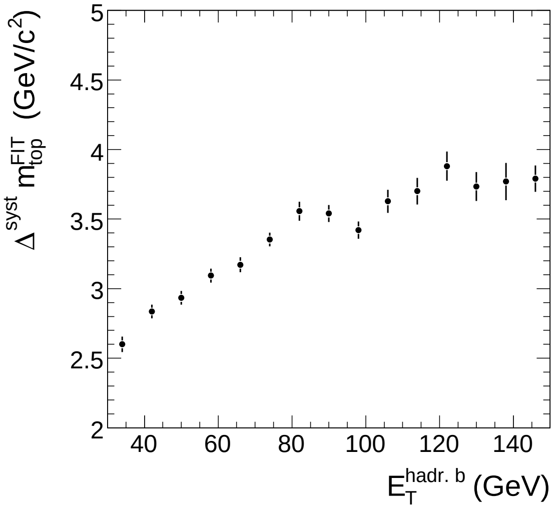


Figure 7.21: For all jet combinations passing the LR criteria, the average m_{top}^{FIT}-shift after a b-quark JES shift of 5% is evaluated as a function of the hadronic b-quark jet's transverse energy.

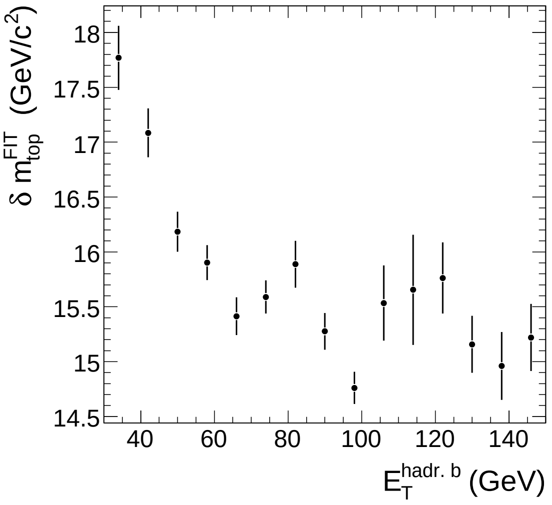


Figure 7.22: For the same jet combinations as in Figure 7.21, the uncertainty on the fitted hadronic top quark mass is plotted as a function of the transverse energy of the hadronic b-jet.

7.3.4 b-tagging efficiency

In Section 6.4.5 it was illustrated that the expected 8% uncertainty on the b-tagging efficiency measurement had a relative important impact on the obtained event selection efficiency. This resulted in a significant systematic uncertainty contribution to the cross-section estimate. What is however relevant for the top quark mass measurement is the effect of the precision of the b-tagging performance on the mistag rate. Events in which the jet is falsely tagged as a b-quark jet might enhance the fraction of combinatorial background in the sample of selected jet combinations. Nevertheless, in Table 7.4 it is observed that this variation of the mistag rate has a negligible systematic uncertainty in all three mass measurements. The following arguments help in understanding this observation:

- Relative strong *b*-tagging requirements are imposed on the *b*-quark jet candidates both in the physics object selection and in the combined LR discriminant $LR_{comb}^{signSel}$, what decreases the mistag probability.
- Mistagged jets might be in reality light quark jets resulting from the hadronic W-decay. Such event solutions will generally be discarded as they will obtain a low P_{χ^2}-value for the kinematic fit imposing the W-boson mass constraint to the light quark jets.
- After the jet combination LR cut the purity of correct jet combinations becomes 79%. This high value ensures that the effect of an increase in the combinatorial background level is limited.

It can hence be concluded that all three top quark mass estimators are robust to a variation of the *b*-tagging efficiency and the corresponding mistag rate.

7.3.5 Miscalibration/misalignment

For the study of the systematic effect introduced on M_{top} by the uncertainty on the detector calibration and misalignment constants, the same `AlpGen` samples were used as in Section 6.4.6. In Section 3.2.5 more information is given on the precision achievable on these constants with $100\,\mathrm{pb}^{-1}$ of accumulated data. Analogue to the JES smearing, the systematic bias on the top quark mass measurement due to the transition from a perfect to a miscalibrated/misaligned detector is observed to be only significant in the case M_{top} is inferred from the Gaussian fit on the spectrum ($1.4\pm0.8\,\mathrm{GeV/c^2}$).

7.3.6 Combination

In Table 7.4 all described systematic effects are added quadratically, and hence are assumed to be uncorrelated. The resulting total systematic uncertainty on all three top quark mass estimators is calculated with and without the inclusion of the main systematic effect, the JES shift of 5%. Where this total systematic uncertainty is found constant for all three measurements ($\sim 4\,\mathrm{GeV/c^2}$), the estimator based on the scanned ideogram is clearly favoured when the JES shift systematic effect is not considered. Combined with the fact that the scanned ideogram method was found to be the one with the best statistical properties, this technique can be seen as the most optimal and robust among all three measurements. Moreover, because the second most dominant uncertainty in this measurement arises from the systematic shift due to the inclusion of pile-up collisions, which should be easy to reduce with a cut on the number of reconstructed vertices in the event.

This leaves us with the strong dependence of the top quark mass measurement with respect to the accuracy on the JES measurement, and more specific, on the *b*-quark JES scale. With the presented top quark mass estimators, the goal to measure the top quark mass with an overall precision of $1\,\mathrm{GeV/c^2}$ will only be achieved if the *b*-quark JES is known to the level of 1.3%.

Chapter 8

Conclusions

Since the discovery of the top quark at the Tevatron collider near Chicago in 1995, the exciting and extended field of top quark physics opened. In Section 1.2.4 the relevance and importance of the exact determination of the top quark properties were argumented. Precision measurements in the top quark sector will invoke not only a further constrain on the SM Higgs boson parameter space, top quark physics is also known to be a perfect window to search for new physics.

With the start of the Large Hadron Collider in the summer of 2008, a new era will open in the field of top quark physics. At the LHC an enormous amount of top quark events is expected to be produced and detected by the Compact Muon Solenoid experiment. However, as we will stress in the following, only this richness in statistics will not be sufficient for a precise estimate of the $t\bar{t}$ production cross-section or the top quark mass. Where in the first measurement in particular the determination of the channel background level is crucial, a perfect understanding of the detector response will be critical for an accurate top quark mass measurement. In respectively Section 8.2 and 8.3 we will briefly sketch how both issues might be addressed. In the first section, we give an overview of the main conclusions taken from the different studies and analyses presented in this thesis. Finally, in Section 8.4 the implications of these results are discussed.

We remind that all described analyses envisage measurement on $1\,\mathrm{fb}^{-1}$ of data, and that in these analyses only semi-muonic $t\bar{t}$ decays were considered as signal events.

8.1 Overview of the main results and observations

8.1.1 The Monte-Carlo generators comparison

An appropriate estimate of the top quark potential is only achievable with a decent description of the expected physics, which is simulated by dedicated Monte-Carlo event generators. Chapter 2 elaborated on the complex tasks of these programs. The leading order multi-parton matrix element generator `AlpGen` was introduced, and its description of the W+jets and $t\bar{t}$+jets kinematics were respectively compared to CDF data and to the prediction of a LO and a NLO simulation. It was shown that, in contrast to pure tree-level LO generators such as `TopRex`, `Alpgen` is capable to reproduce well the observable's shapes predicted by the NLO MC@NLO program or measured in data. In general, for initial and final state radiation jets, a harder jet E_T spectrum is obtained by `AlpGen` compared to a LO description in which all of the extra jets are produced by the parton shower. As a consequence, issues such as the combinatorial background or the rate of selecting QCD events will depend strongly on this description. The use of `AlpGen` samples with a high extra jet multiplicity is hence recommended in order to start the analysis from a more realistic description of the expected physics. However, in this thesis only the employed $\mathrm{t\bar{t}}$ and $W(bb)$ samples were produced with this generator, the QCD events were still generated with `PYTHIA`. A proper simulation of the QCD events would result in an increase of its selection efficiency due to the higher efficiency for events to pass the jet requirements.

8.1.2 The jet performance study

After a description of the Large Hadron Collider and the CMS detector in Chapter 3, Chapter 4 described the reconstruction of the physics objects starting from the detected digitised information. Hereby, the jet clustering and calibration was emphasised. A method based on several quality markers was introduced to quantify the jet reconstruction performance from the analysis point of view. For three jet algorithms, the iterative cone, the k_T-algorithm and the midpoint cone algorithm, the respective parameter space was scanned for the most optimal algorithm setting. This resulted for the cone based algorithms in an optimal cone opening angle of around 0.5, while for the k_T-algorithm the best performance was found for a R-parameter of 0.6. Also the impact of the magnetic field and the reconstruction in the calorimeter was investigated. In general both effect will result in an increase of the optimal opening angle and R-parameter. All these observations are summarised in Table 4.2. Based on these figures the iterative cone algorithm with a cone opening angle of 0.5 was chosen to be used in the top quark analyses.

8.1.3 The determination of the l-JES calibration

After the application of these Monte-Carlo based, jet E_T and η dependent, jet energy scale corrections to the reconstructed jet four-momenta, several rather straightforward requirements on the reconstructed physics objects were posed to enhance the fraction of semi-muonic decaying $\mathrm{t\bar{t}}$ events in the selected event sample (cfr. Table 4.3). The QCD event rate expected after these cuts was examined and showed to be of the order of a factor three higher than the rate expected for the $W(bb)$-events, but still more than a factor three lower than the signal event rate. After all requirements on the physics objects, a S/B-ratio of 2.56 was found for a signal event selection efficiency of around 20%.

With the remaining events a method to calibrate the inclusive light quark jet energy scale (JES) was introduced in Chapter 5. In this technique the two selected light quark jets are assumed to be the jets originating from the hadronic W-boson decay products. The distribution of the invariant mass of this system is expected to be centred around the world average W-boson mass. Any shift of the Gaussian fitted mean value with respect to the true W-boson mass is considered as a miscalibration of the inclusive light quark JES.

The resulting estimator of the inclusive light quark JES correction factor is found unbiased with respect to the Monte-Carlo true calibration factor. For $1\,\mathrm{fb}^{-1}$ a total uncertainty on this estimator of the order of 1% might be realistic, if both the systematic effects originating from the uncertainty on the channel background shape and level, and the influence of pile-up collisions are well understood. However, a cut on the number of reconstructed primary vertices was proposed as a highly effective tool to decrease the sensitivity of the method towards pile-up collisions. In all three top quark analyses (ΔC, $\sigma_{semi-\mu}$ and M_{top} estimation) it was illustrated that this extra requirement allowed to even cancel this systematic effect. Hence, even if the pile-up collision rate is underestimated, a powerful technique to suppress its influence is available. Such an underestimation is far from unlikely. For the inclusion of low luminosity pile-up collisions, an average number of 3.5 minimum bias events were added to the hard event using the CMS FastSim program. To illustrate, the inclusion of pile-up collisions resulted in a ∼10% increase in the number of reconstructed jets with an $E_T > 30\mathrm{GeV}$ directed in the tracker acceptance. It should however be kept in mind that the kinematics of the minimum bias events were extrapolation from the CDF and DØ measurements. This induces large uncertainties in the pile-up description and consequently in e.g. the number of jets and the JES.

Although only an inclusive JES miscalibration factor was derived, the presented analysis is considered as an important show case of the strength of using $\mathrm{t\bar{t}}$ events for detector commissioning tasks. Currently, important work is going on to investigate the differentiation of this method as function of the jet transverse energy or its pseudorapidity. Additionally, an analogue jet calibration technique is being developed in which both the world average W-boson and top quark mass constraints are imposed on the kinematics of the hadronic decaying top quark in semi-leptonic $\mathrm{t\bar{t}}$ events. This way also a b-quark JES correction can be extracted from the data [155].

8.1.4 The $\mathrm{t\bar{t}}$ production cross-section measurement

In order to further suppress the fraction of background events, a Likelihood Ratio discriminator was constructed in Chapter 6, combining the information of several observables that are able to distinguish between semi-muonic signal events and background contributions. After a cut on this combined LR-discriminant value a S/B-ratio of 5.2 is obtained, an increase by a factor of almost 50 with respect to the initial ratio (for $\epsilon^{tot}_{signal}=5.85$). With these remaining events an estimator of the semi-muonic production cross-section was constructed. Due to the high signal purity in the selected event sample, this measurement was defined in terms of a simple counting experiment. With $1\,\mathrm{fb}^{-1}$ of accumulated data, a total relative uncertainty of 26 % was found, by far dominated by the systematic uncertainty arising from the channel background level (16% for a rescaling factor of 5). Possible strategies to address this main systematic effect will be sketched in the following section. Additionally, the cross-section estimate was shown rather sensitive to the uncertainty on the b-tagging efficiency, the alignment and calibration of the detector, and the JES smearing. Where an increase of the LR-cut value was illustrated to be highly efficient in the reduction of the b-tagging systematic effect (cfr. Figure 6.17), both other effects are more difficult to tackle. Especially, the impact of an increase of the JES resolution on the jets selection efficiency should be studied carefully. However, studies are on-going to measure the JES resolution from the data, what would reduce the uncertainty and hence the systematic effect on the $\sigma_{semi-\mu}$ estimate.

8.1.5 The measurement of the top quark mass

In Chapter 7 several top quark mass estimators were constructed and compared using the events remaining after the cut on the signal event enhancing LR-discriminator. Each of the techniques is based on the complete reconstruction of the hadronic top quark decay and hence the determination of the initial hadronic top quark four-momentum. After the selection of two light and two b-quark jet candidates, a last ambiguity remains in the identification of the b-quark jet originating from the hadronic top quark decay. Another time the power of a Likelihood Ratio technique was employed to combine the information of several topological observables which can differentiate the correct from the wrong combination. After a cut on the resulting LR-discriminator a purity of correct jet assignment of 76% was found. Only these event solutions are used as an input to the top quark mass estimators.

Three different top quark mass estimators are defined: a simple fit on the reconstructed top quark mass spectrum and two event-by-event likelihood methods which convolute the resolution function of the event or so-called ideogram with the expected theoretical template. In this theoretical template both the probability of having the correct jet combination and the probability that the selected event is indeed a semi-muonic decaying $\mathrm{t\bar{t}}$ event is accounted for.

For each of the three estimators, the statistical properties were studied and are summarised in Table 7.3. The results indicate a slope of unity between the generated and estimated top quark mass for both ideogram based methods, while a significant lower value was found for the Gaussian fit on the spectrum. A study of the width of the pull distribution on the other hand indicated that only in the case of the Gaussian ideogram method the statistical uncertainty on the top quark mass estimate is somehow underestimated. Also the bias with respect to the generated M_{top}-value was found significantly higher for this method. Combining these observations clearly favours the scanned ideogram technique. For $1\,\mathrm{fb}^{-1}$ of data, a statistical uncertainty in the order of $0.6\,\mathrm{GeV/c}^2$ is expected, which will anyhow be negligible with respect to the total systematic uncertainty on each of the three estimators (cfr. Table 7.4).

The uncertainty on the jet energy scale is, as suspected, identified as the main systematic uncertainty on the inferred top quark mass estimates (3.3-3.8 $\mathrm{GeV/c}^2$). More precisely, only the uncertainty on the b-quark JES is relevant, as both the application of the data-driven light quark JES measurement presented in this work and the use of a kinematic fit almost completely cancels the effect of a light quark JES shift. The ideogram based methods are more sensitive to the JES uncertainty compared to the fit on the spectrum, due to the higher weights of high-E_T events in the combined likelihood and the relatively higher impact of a fixed shift of the JES for increasing jet transverse energy. In the calculation of this systematic uncertainty on M_{top}, the JES is assumed to be known to the level of 5%, as predicted in [88, 117]. In Section 8.3 some reflections are given on how this systematic uncertainty might be lowered, which is crucial to reach the goal of measuring the top quark mass with a precision of 1 GeV/c^2.

When excluding the systematic effect due to the jet energy scale uncertainty from the calculation of the total systematic uncertainty, clearly the scanned ideogram technique is found to be the most optimal estimator. For this method, only the inclusion of pile-up collisions will give rise to another significant systematic shift (0.65 GeV/c^2). However, also here this effect is easily reduced by a cut on the number of reconstructed primary vertices. Hence, apart from the sensitivity towards the JES uncertainty, this method can be considered robust towards all other tested theoretical, phenomenological and detector related sources of uncertainty.

Finally, a total uncertainty on the M_{top}-estimate of 4.0 GeV/c^2 was derived for all three methods, dominated by the jet energy scale of heavy jets.

8.2 The extraction of the channel background level and shape from data

Only by actually measuring the channel background event rate and shape in the data itself this prominent source of uncertainty on several measurements, such as on the cross-section estimate presented in this work, can be reduced significantly. In this section a sketch is given of how to apply in CMS the method used by CDF to measure the QCD level [156]. An alternative to this measurement is the so-called DØ Matrix Method [157]. Both ideas should be fully exploited and compared.

The CDF technique is based on the idea to isolate sidebands in a given observable's phase-space that are populated almost exclusively by QCD events, and another region with an important contribution of 'W-like' events ($\mathrm{t\bar{t}}$ and $W(bb)$). The aim is to find typically two observables that can be considered uncorrelated and which contain a strong separation power between QCD and W-like events. In this example we selected as observables the transverse momentum of the isolated muon and the transverse momentum of the jet with fourth highest E_T in the event. Also, a relaxed single muon trigger stream was used, which only required a reconstructed muon with a p_T exceeding 16 GeV/c, without any isolation cut applied.

The resulting distributions of the selected and triggered events over the two dimensional plane are illustrated in Figure 8.1. While in the left plot the result is shown for QCD events, the W-like contributions fill the right plot. Clearly, QCD events tend to populate the regions close to the origin and the sidebands of the plane. Four quantities can be identified[1]:

- N_A: the number of triggered events with no isolated muon and less than four reconstructed jets
- N_B: the number of triggered events with no isolated muon but with the fourth highest E_T jet in the event passing the $E_T^{4th\,jet}$ >30 GeV/c requirement

---

[1] Due to the limited amount of simulated QCD events no b-tagging requirement is included in the event selection. However, in principle two b-tagged jets could be required in data to stay as close as possible to the event selection applied in the analyses.

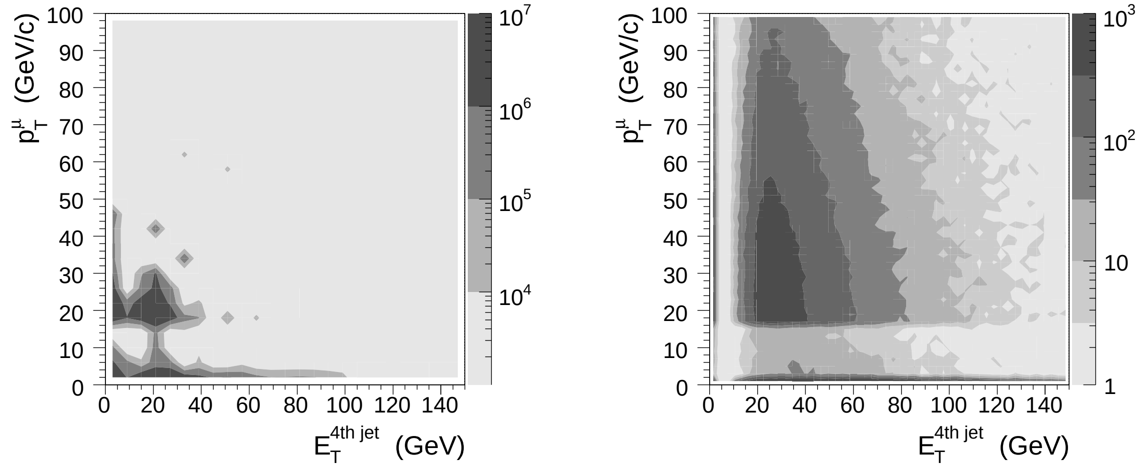


Figure 8.1: The population of QCD (left) and W-like (right) events over the ($E_T^{4th\,jet}$, $p_T^{iso\,\mu}$) plane. All events were required to pass the relaxed single muon trigger stream with a p_T^μ threshold of 16 GeV/c (without additional isolation requirements).

- N_C: the number of triggered events with an isolated muon exceeding $p_T^{iso\,\mu}$ >20 GeV/c and less than four reconstructed jets
- N_D: the number of triggered events remaining after both above requirements

Considering the large QCD event rate passing the relaxed trigger, the W-like events contribution in the regions A, B and C are negligible compared to the QCD event rate. Assuming both the muon and jets related observable to be uncorrelated , the number of QCD events in region D can be estimated as:

$$N_D^{QCD} = \frac{N_B \, . \, N_C}{N_A}.$$

This method will allow a precise estimate of the fraction of QCD events that is expected after a given event selection. With larger samples of simulated QCD events this approach can be tested.

However, more creativity is needed to measure for example the QCD W-boson mass spectrum from the data. A possible procedure to extract this m_W^{QCD} distribution is sketched below:

1. Construct a first m_W-spectrum using only events with 4 high E_T jets (of which two are b-tagged) and a high p_T lepton. This distribution will dominated by QCD events.

2. Construct a second m_W-spectrum using the same event selection plus the additional requirement for muon isolation. A much higher fraction of W-like event is present in this distribution.

3. Subtract the second from the first m_W distribution, to obtain a spectrum which almost exclusively contains QCD events.

Only events will be employed that remain after the relaxed single muon trigger requirements. This procedure exploits the assumption that no correlation is expected between the muon isolation and the reconstructed m_W value, which is expected to be a good approximation, but anyhow can be verified carefully when larger simulated QCD event samples are available.

8.3 Reflections on the jet energy scale uncertainty

Any top quark mass measurement based on the full reconstruction of the initial hadronic top quark four-momentum, unavoidably introduces a strong correlation between the measured jet energy scale and the obtained top quark mass. In Chapter 7 it was however illustrated that the light quark jet energy scale uncertainty is almost completely cancelled when exploiting the topological constraint of the W-boson mass in the top decay chain. Consequently, by far the largest contribution to the M_{top} uncertainty arises from the uncertainty on the b-quark jet energy scale. In order to achieve the aimed precision of $1\,\mathrm{GeV/c^2}$, improvements are required both in the top quark mass analyses as in the JES measurement.

Improvements in the JES measurement

Analogue to CDF and DØ, CMS will implement a factorised, data-driven JES calibration chain, as described in Section 4.1.3. Starting with an offset correction to address e.g. the extra energy deposits arising from pile-up collisions or detector noise, other data-driven methods are developed to apply a relative η and absolute p_T correction factor to the jets. From the di-jet and γ+jet calibration techniques, a 5% overall uncertainty on the JES can be expected for $1\,\mathrm{fb}^{-1}$ [88, 117]. Additional to these correction factors, parton level or flavour dependent corrections will be measured to further reduce the JES uncertainty and resolution. The calibration factor obtained from the method presented in Chapter 5 for example will result in a light-quark jet correction factor up to the parton level.

For the heavy quark flavours, calibration constants can also be derived from semi-muonic $\mathrm{t\bar{t}}$ events by imposing the world average top quark mass as an extra constraint [2]. In principle, these b-quark jet corrections might even be applied to a separate $\mathrm{t\bar{t}}$ subsample. Especially in the first period of the experiment, when the focus lies on the definition and the understanding of the proposed top quark mass estimators, and consequently no combination of the Tevatron and LHC top quark mass results are envisaged, this option might be preferred.

Heavy flavour corrections might also be derived from Monte-Carlo simulated events. In this case data-driven methods should be applied to correct the reconstructed jet up to the level of a generated particle jet. Hence only for the extraction of the correction factors to move from the generated b-quark jet to the parton level Monte-Carlo simulated events are used[3]. In the top quark mass measurements, all b-quark jets in

---

[2] Additionally, Zbb/Zqq events are seen as a promising alternative, which is currently under study.

[3] Of course, the derivation of these MC based calibration factors should be accompanied with a systematic study to check the sensitivity towards e.g. the employed jet fragmentation model.

semi-muonic $\mathrm{t\bar{t}}$ decays might initially be calibrated as if they were light quark jets, using data-driven techniques, but with a Monte-Carlo correction applied to correct for the difference in light and heavy generated particle level jets.

Apart from the extraction of more accurate calibration factors from the data, the JES uncertainty might also be highly sensitive to the definition of the input objects in the jet clustering algorithms. In all presented studies ECAL+HCAL calorimeter towers were used. However, as soon the detector response is well understood, the transition to particle flow objects is expected to improve impressively the jet resolution and consequently the uncertainty on the JES [101].

Finally, the comparison of the output of several jet clustering algorithms might be a powerful strategy in the reduction of the systematic effect due to the JES and perturbative QCD uncertainties. In [151] a strategy to make the top quark mass estimators more robust against systematic uncertainties was exploited. An additional event selection requirement was defined, according to which only events are accepted in which the direction of the four selected hadronic $\mathrm{t\bar{t}}$-decay products is independent of the jet clustering algorithm that was used. Although the first results of this extra requirement indicated a minor improvement on the final M_{top}-uncertainty, this idea might be a powerful tool at start-up or in other data-analyses.

Reducing the sensitivity of the mass measurement

The uncertainty on the jet energy scale will not be constant over the phase-space of a given jet. Forward and relatively low E_T jets are expected to have worse energy resolution compared to central, high transverse momentum jets. In the presented studies, however, an overall JES uncertainty of 5% was accounted for. Once more advanced (data-driven) studies become available and data-taking starts, this accurancy will be estimated for several phase-space regions. Consequently, only selecting events with jets in the most precise determined JES regions might reduce the total uncertainty on M_{top} significantly. After such a more stringent jet selection, a balance should be found between the statistical and systematic uncertainty.

8.4 Future implications of the presented top quark measurements

In this thesis the potential of measuring the top quark mass and production cross-section with the CMS detector and using $1\,\mathrm{fb}^{-1}$ of data was investigated. Estimates were given of the total uncertainties expected in both measurements ($\Delta M_{top} \simeq 4\,\mathrm{GeV/c^2}$; $\Delta\sigma_{semi-\mu}/\sigma_{semi-\mu} \simeq$26%), numbers that were highly dominated by systematic effects. Time will be needed to fully understand and reduce the impact of these systematic effects, which is in particular the case for the JES. Nevertheless, it should be emphasised that the scanned ideogram M_{top}-estimator was shown to be robust with respect to all tested sytematic effects but the b-quark JES, what allows to fully concentrate on this latter effect and employ the large statistics for the reduction of this systematic uncertainty.

The total uncertainty on both the top quark mass and cross-section measurements derived from the study of semi-muonic decaying $\mathrm{t\bar{t}}$ events will also be reduced by combining the results obtained in different $\mathrm{t\bar{t}}$-decay channels or processes, and by merging the CMS and ATLAS results [34, 147, 158, 159]. However, today many different procedures are used in the calculation of the systematic uncertainties, and different integrated luminosities are studied, what makes a comparison difficult. For example, in many analyses the effect of the channel background level in the determination of the cross-section is underestimated, as is the JES uncertainty in the top quark mass measurement. Nevertheless, some important conclusions can be made. Firstly, the fully leptonic $\mathrm{t\bar{t}}$-decay for example is expected to be the most promising channel for a cross-section measurement in the early stage of CMS ($\Delta\sigma/\sigma \simeq 11\%$ [158]), due to its limited amount of channel background events. Secondly, for the measurement of the top quark mass, the highest precision will be obtained in the semi-leptonic decay channel, although for high integrated luminosities the J/Ψ-based method becomes compelling ($\Delta M_{top} \simeq 1.5\,\mathrm{GeV/c^2}$ for $20\,\mathrm{fb^{-1}}$ [34]).

The top quark mass precision obtained in this thesis is in good agreement with the semi-leptonic result determined by ATLAS [159]. Also for ATLAS the main systematic uncertainty on M_{top} arises from the b-quark JES uncertainty. In [159], a systematic shift on M_{top} of $0.7\,\mathrm{GeV/c^2}$ was found due to the systematic effect of a very optimistic 1% b-JES uncertainty. Rescaling this shift to the 5% uncertainty similar to the one applied in this work results in a top quark mass uncertainty of $3.5\,\mathrm{GeV/c^2}$, which is of the same magnitude as what is determined in Chapter 7.

Work is also needed to determine the correlation between the diverse range of methods and the different channels explored at the LHC to extract the top quark mass and cross-section. Where in the case of a combination of the top quark mass results obtained with the J/Ψ and the scanned ideogram technique the correlation can be considered negligible, the correlation between the measurements from different $\mathrm{t\bar{t}}$-decays is much more difficult to estimate. However, with a proper understanding of the detector performance a combined precision on the top quark mass of $1\,\mathrm{GeV/c^2}$ and a cross-section estimate better than $\sim$10% should be feasible.

These precision measurements will be of major importance to further constrain the uncertainty on the Standard Model Higgs boson mass and to search for new physics, as motivated in Section 1.2.4. Combined with an improvement on the W-boson mass uncertainty to $\Delta M_W \simeq 15\,\mathrm{MeV/c^2}$ [4], a top quark mass uncertainty of $1\,\mathrm{GeV/c^2}$ would yield a relative Higgs boson mass of 25% [161]. The measurement of the cross-section might also be translated in a consistency check of the top quark mass. An uncertainty on the cross-section of 10% could result in an uncertainty of $3.5\,\mathrm{GeV/c^2}$ on the top quark mass (cfr. Section 1.2.3). It will be important to compare the direct with the indirect measurement of the top quark mass as they are related to the Higgs boson

---

[4] For $10\,\mathrm{fb^{-1}}$, the combination of the results obtained for the $W \rightarrow e\nu$ and $W \rightarrow \mu\nu$ decay channels in CMS is expected to decrease the total uncertainty on the W-boson mass to the order of $20\,\mathrm{MeV/c^2}$ [160]. After a merge of the measurements performed by ATLAS and CMS a $15\,\mathrm{MeV/c^2}$ is within reach.

mass via radiative loop correections. Differentiating the cross-section measrement as a function of the $t\bar{t}$ invariant mass is considered as an important window on new physics [43].

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] W.-M. Yao *et al.*, J. Phys. G **33** (2006) 1–1232.

[4] S. Weinberg, Phys. Rev. Lett. **19** (1967) 1264–1266.

[5] A. Salam and J. C. Ward, Phys. Lett. **13** (1964) 168–171.

[6] P. A. M. Dirac, Proc. Roy. Soc. Lond. **A118** (1928) 351.

[7] K. Schulten, *Notes on Quantum Mechanics*. 2000. `http://www.ks.uiuc.edu/Services/Class/PHYS480/qm_PDF/chp10.pdf`.

[8] E. Noether and M. A. Tavel, *Invariant Variation Problems*, `physics/0503066`.

[9] S. Weinberg, *The Quantum Theory of Fields. Vol. 1: Foundations*, Cambridge, UK: Univ. Pr. (1995) 609 p.

[10] G. 't Hooft and M. J. G. Veltman, Nucl. Phys. **B44** (1972) 189–213.

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

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

[13] S. P. Martin, *A supersymmetry primer*, `hep-ph/9709356`.

[14] M. Tytgat, *Introduction to the Standard Model of Electroweak Interactions*, Lectures given at the Joint Belgian-Dutch-German Summer School, Bonn.

[15] T. Appelquist, H.-C. Cheng, and B. A. Dobrescu, Phys. Rev. **D64** (2001) 035002.

[16] L. Randall and R. Sundrum, Phys. Rev. Lett. **83** (Oct, 1999) 3370–3373.

[17] A. Leike, Phys. Rept. **317** (1999) 143–250.

[18] C. T. Hill and E. H. Simmons, Phys. Rept. **381** (2003) 235–402.

[19] K. D. Lane, *Technicolor 2000*, hep-ph/0007304.

[20] F. Abe *et al.*, Phys. Rev. D **50** (Sep, 1994) 2966–3026.

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

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

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

[24] R. Bonciani, S. Catani, M. L. Mangano, and P. Nason, Nucl. Phys. **B529** (1998) 424–450.

[25] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, JHEP **01** (2007) 013.

[26] CDF Collaboration, *Combination of CDF top quark pair production cross section measurements with up to 760 pb*$^{-1}$, http://www-cdf.fnal.gov/physics/new/top/confNotes/cdf8148_ttbar_xs_combo.ps CDF Note 8148, April, 2006.

[27] CDF Collaboration, *Combined* $\mathrm{t\bar{t}}$ *Production Cross Section in the Lepton+Track and Dilepton Final States using 1 fb*$^{-1}$ *of data*, http://www-d0.fnal.gov/Run2Physics/WWW/results/prelim/TOP/T60/T60.pdf D0 note 5477-CONF, September, 2007.

[28] CDF Collaboration, *A Limit on the Top Quark Width and the Lifetime using the Template Method in the Lepton plus Jets Channel at CDF II*, http://www-cdf.fnal.gov/physics/new/top/confNotes/cdf8953_topwidth_1fb-1.pdf CDF Note 8953, August, 2007.

[29] D. Collaboration, Physics Letters B **639** (2006) 616.

[30] S. L. Glashow and J. Iliopoulos, Phys. Rev. D **3** (Feb, 1971) 1043–1045.

[31] C. A. Nelson, B. T. Kress, M. Lopes, and T. P. McCauley, Phys. Rev. D **56** (Nov, 1997) 5928–5944.

[32] CDF II Collaboration, A. Abulencia *et al.*, Phys. Rev. **D75** (2007) 052001.

[33] C. S. Hill, J. R. Incandela, and J. M. Lamb, Phys. Rev. **D71** (2005) 054029.

[34] R. Chierici and A. Dierlamm, *Determination of the top mass with exclusive events* $t \to Wb \to \ell\,\nu\,J/\Psi\,X$, CMS-NOTE-2006-058.

[35] CDF Collaboration, *A Combination of CDF and D0 results on the mass of the top quark*, hep-ex/0703034.

[36] CDF Collaboration, *Measurement of the top mass with in situ jet energy scale calibration in the all-hadronic channel using the Template Method with 1.9 fb*$^{-1}$, http://www-cdf.fnal.gov/physics/new/top/confNotes/cdf9165_cdfAllHadTMT2D.ps CDF note 9165-CONF.

[37] CDF Collaboration, *Top Mass Measurement in the Lepton + Jets channel using a modified matrix element method and in-situ JES calibration with 1.9 fb*$^{-1}$, `http://www-cdf.fnal.gov/physics/new/top/confNotes/cdf9196_mtm25_1900invpb_pub.eps` CDF note 9196-CONF.

[38] D0 Collaboration, *Measurement of the top quark mass with the matrix element method using the lepton+jets 1 fb*$^{-1}$ *data set*, `http://www-d0.fnal.gov/Run2Physics/WWW/results/prelim/TOP/T42/T42.pdf` DO note 5362-CONF.

[39] L. Lyons, D. Gibaut, and P. Clifford, Nuclear Instruments and Methods in Physics Research A **270** (July, 1988) 110–117.

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

[41] *The SLAC Large Detector experiment*, `http://www2.slac.stanford.edu/VVC/detectors/sld.html`.

[42] *The LEP Electroweak Working Group*, `http://lepewwg.web.cern.ch/LEPEWWG/`.

[43] R. Frederix and F. Maltoni, *Top pair invariant mass distribution: a window on new physics*, Comments: 30 pages, 17 figures.

[44] H. Y. Zhou, K. Y. Ping, G. R. Lu, H. Wang, and Y. C. Xing, Nucl. Phys. B, Proc. Suppl. **75** (1999), no. hep-ph/9808316, 302–304.

[45] H.-J. Zhang, *Top-Quark FCNC Decay t-¿cgg in Topcolor-assisted Technicolor Model*, `arXiv:0712.0151 [hep-ph]`.

[46] C. T. Hill and S. J. Parke, Phys. Rev. D **49** (May, 1994) 4454–4462.

[47] R. Casalbuoni, P. Chiappetta, D. Dominici, A. Fiandrino, and R. Gatto, Z. Phys. **C69** (1996) 519–524.

[48] T. Appelquist and G. Triantaphyllou, Phys. Rev. Lett. **69** (1992) 2750–2753.

[49] D0 Collaboration, *Search for* $t\bar{t}$ *Resonance in the Lepton+Jets Final State in* $p\bar{p}$ *Collisions at* $\sqrt{s} = 1.96\,TeV$, `http://www-d0.fnal.gov/Run2Physics/WWW/results/prelim/TOP/T53/T53.pdf` D0 Note 5443-CONF, July, 2007.

[50] D. Benedetti *et al.*, *Search for* $H \rightarrow b\bar{b}$ *in Association with a ttbar Pair at CMS*, tech. rep., Jun, 2006. CMS-NOTE-2006-119.

[51] D. P. Roy, Phys. Lett. **B459** (1999) 607–614.

[52] S. P. Mehdiabadi, *Inclusive search for SUSY in top final states at CMS*, Tech. Rep. hep-ph/0610086, Oct, 2006.

[53] S. Lowette, J. D'Hondt, J. Heyninck, and P. Vanlaer, *Offline Calibration of b-Jet Identification Efficiencies*, CMS-NOTE-2006-013.

[54] M. A. Dobbs *et al.*, *Les Houches guidebook to Monte Carlo generators for hadron collider physics*, `hep-ph/0403045`.

[55] T. Sjostrand, *Monte Carlo generators*, `hep-ph/0611247`.

[56] T. Sjostrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 Physics and Manual*, `hep-ph/0603175`.

[57] G. Corcella, I. G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M. H. Seymour, and B. R. Webber, *HERWIG 6.5: an event generator for Hadron Emission Reactions With Interfering Gluons (including supersymmetric processes)*, `hep-ph/0011363`.

[58] D. Stump *et al.*, JHEP **10** (2003) 046.

[59] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, Eur. Phys. J. **C14** (2000) 133–145.

[60] *On-line PDF Plotting and Calculation*, `http://durpdg.dur.ac.uk/hepdata/pdf3.html`.

[61] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, Eur. Phys. J. **C28** (2003) 455–473.

[62] J. Pumplin *et al.*, JHEP **07** (2002) 012.

[63] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, JHEP **07** (2003) 001.

[64] F. Caravaglios, M. L. Mangano, M. Moretti, and R. Pittau, Nucl. Phys. **B539** (1999) 215–232.

[65] M. L. Mangano, M. Moretti, and R. Pittau, Nucl. Phys. **B632** (2002) 343–362.

[66] F. Maltoni and T. Stelzer, JHEP **02** (2003) 027.

[67] T. Stelzer and W. F. Long, Comput. Phys. Commun. **81** (1994) 357–371.

[68] Y. L. Dokshitzer, Sov. Phys. JETP **46** (1977) 641–653.

[69] V. V. Sudakov, Sov. Phys. JETP **3** (1956) 65–71.

[70] P. Bartalini, R. Chierici, and A. De Roeck, *Guidelines for the estimation of theoretical uncertainties at the LHC*, CERN-CMS-NOTE-2005-013.

[71] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber, JHEP **11** (2001) 063.

[72] S. Hoche *et al.*, *Matching parton showers and matrix elements*, `hep-ph/0602031`.

[73] F. E. Paige and S. D. Protopopescu, *ISAJET 5.30: A MONTE CARLO EVENT GENERATOR FOR p p AND anti-p p INTERACTIONS*, To appear in Proc. of the 1986 Summer Study on the Physics of the Superconducting Super Collider, Snowmass, Colo., Jun 23 - Jul 11, 1986.

[74] S. Frixione and B. R. Webber, *The MC@NLO 3.3 event generator*, `hep-ph/0612272`.

[75] M. Pierini and M. Spiropulu, *Study* $\mathrm{t\bar{t}}$*+jets as a function of jet* p_T *and multiplicity for 2007 1fb*$^{-1}$ *Monte Carlo production at CMS*, CMS Internal Note IN-2007/038.

[76] S. R. Slabospitsky and L. Sonnenschein, Comput. Phys. Commun. **148** (2002) 87–102.

[77] CDF Collaboration, T. Aaltonen *et al.*, *Measurement of the cross section for W-boson production in association with jets in ppbar collisions at sqrt*$(s) = 1.96$ *TeV*, `arXiv:0711.4044 [hep-ex]`.

[78] B. Andersson, Acta Phys. Polon. **B32** (2001) 3993–4011.

[79] C. Peterson, D. Schlatter, I. Schmitt, and P. M. Zerwas, Phys. Rev. D **27** (Jan, 1983) 105–111.

[80] OPAL Collaboration, G. Alexander *et al.*, Z. Phys. **C69** (1996) 543–560.

[81] *The Large Hadron Collider Project*, `http://lhc.web.cern.ch/lhc/`.

[82] *The ATLAS Experiment*, `http://atlas.web.cern.ch/Atlas/index.html`.

[83] *The CMS Experiment*, `http://cms.cern.ch/`.

[84] *The ALICE Experiment*, `http://aliceinfo.cern.ch/`.

[85] *The LHC-b Experiment*, `http://lhcb.web.cern.ch/lhcb/`.

[86] *The relativistic heavy ion collider*, `http://www.bnl.gov/rhic/default.htm`.

[87] A. Ball, M. Della Negra, L. Fo, and A. Petrilli, *CMS physicsTechnical Design Report*. Technical Design Report CMS. CERN, Geneva, 2006. revised version.

[88] M. Della Negra, L. Foà, A. Hervé, and A. Petrilli, *CMS physics Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 2006.

[89] I. Belotelov, O. L. Buchmüller, I. González-Caballero, A. Heister, T. Lampén, P. Martinez, C. Martínez-Rivero, F. Matorras, M. Thomas, and V. Valuev, *Simulation of Misalignment Scenarios for CMS Tracking Devices*, CMS-NOTE-2006-008.

[90] T. Lampén, M. Weber, N. de Filippis, and A. Schmidt, *Misalignment Scenarios for the Startup Conditions of the CMS Tracker*, CMS Internal Note IN-2007/061.

[91] N. de Filippis, T. Lamén, F.-P. Schilling, A. Schmidt, and M. Weber, *Update of Misalignment Scenarios for the CMS Tracker*, CMS Internal Note IN-2007/061.

[92] *The CSA07 exercise*, `http://twiki.cern.ch/twiki/bin/view/CMS/CSA07`.

[93] *Calibration and alignment constants used in reconstruction for CSA07 and their Scenarios*, http://twiki.cern.ch/twiki/bin/view/CMS/DetectorPerformanceCSA07AlCaConstants.

[94] CMS Collaboration, S. Dasu *et al.*, *CMS. The TriDAS project. Technical Design Report, Vol. 1: The Level-1 Trigger*, CERN-LHCC-2000-038.

[95] CMS Collaboration, P. Sphicas (Ed.), *CMS: The TriDAS project. Technical Design Report, Vol. 2: Data Acquisition and High-Level Trigger*, CERN-LHCC-2002-026.

[96] C. L. et al., CMS Analysis Note AN-2007/009.

[97] G. Collaboration, Nuclear Instruments and Methods in Physics Research A **506** (July, 2003) 250–303.

[98] M. Hansen *et al.*, *The Top Quark Analysis Framework*, CMS Internal Note IN-2007/068.

[99] *The Top Quark Analysis Framework TWiki-page*, `http://twiki.cern.ch/twiki/bin/view/CMS/TWikiTopQuark`.

[100] C. Grandi, D. P. Stickland, L. Taylor, A. Petrilli, and A. Herv, *CMS Computing Model*,.

[101] *Particle Flow and Tau Identification Physics Object Group (work in process)*, `https://twiki.cern.ch/twiki/bin/view/CMS/EFlow`.

[102] R. Demina *et al.*, *Calorimeter cell energy thresholds for jet reconstruction in CMS*, CERN-CMS-NOTE-2006-020.

[103] S. V. Chekanov, *Jet algorithms: A mini review*, `hep-ph/0211298`.

[104] UA1 Collaboration, G. Arnison *et al.*, Phys. Lett. **B132** (1983) 214.

[105] G. C. Blazey *et al.*, *Run II jet physics*, `hep-ex/0005012`.

[106] S. D. Ellis and D. E. Soper, Phys. Rev. **D48** (1993) 3160–3166.

[107] CDF - Run II Collaboration, A. Abulencia *et al.*, Phys. Rev. **D75** (2007) 092006.

[108] CDF - Run II Collaboration, A. Bhatti, F. Chlebana, *et al.*, Phys. Rev. **D74** (2006) 071103.

[109] J. M. Butterworth, J. P. Couchman, B. E. Cox, and B. M. Waugh, Comput. Phys. Commun. **153** (2003) 85–96.

[110] D0 Collaboration, N. Buchanan and S. Hagopian, *Jet energy scale determination for the Run II D0 calorimeter*, Prepared for 9th ICATPP Conference on Astroparticle, Particle, Space Physics, Detectors and Medical Physics Applications, Villa Erba, Como, Italy, 17-21 Oct 2005.

[111] CDF and D0 Collaboration, K. Hatakeyama, PoS **TOP2006** (2006) 014.

[112] G. L. e. a. S. Esen, CMS Analysis Note AN-2007/055.

[113] A. Heister *et al.*, *Measurement of jets with the CMS detector at the LHC*, CERN-CMS-NOTE-2006-036.

[114] R. Harris, *Jet Calibration from Dijet Balancing*, CERN-CMS-ANALYSIS-NOTE-2005-034.

[115] V. Konoplyanikov, O. Kodolova, and A. Ulyanov, *Jet calibration using gamma + jet events in the CMS detector*, CERN-CMS-NOTE-2006-042.

[116] J. D'Hondt, S. Lowette, J. Heyninck, and S. Kasselmann, *Light quark jet energy scale calibration using the W mass constraint in single-leptonic t anti-t events*, CERN-CMS-NOTE-2006-025.

[117] J. Rohlf and C. Tully, *Recommendations for Jet and Missing Transverse Energy Reconstruction Settings and Systematics Treatment*, CERN-CMS-IN-NOTE-2006-025.

[118] C. Buttar *et al.*, *Les Houches physics at TeV colliders 2005, standard model, QCD, EW, and Higgs working group: Summary report*, `hep-ph/0604120`.

[119] C. Rubbia, A. Astbury, B. Aubert, and A. C. Benvenuti, *A 4π solid angle detector for the SPS used as a proton-antiproton collider at a centre of mass energy of 540 GeV*, Tech. Rep. CERN-SPSC-78-19. SPSC-P-92-S, CERN, Geneva, 1978.

[120] G. Arnison *et al.*, Phys. Lett. B **122** (Jan, 1983) 103–116. 31 p.

[121] CDF Collaboration, Phys. Rev. D **56** (1997), no. FERMILAB-PUB-97-031-E, 1357–1362.

[122] *The CMS hadron calorimeter projectTechnical Design Report.* Technical Design Report CMS. CERN, Geneva, 1997.

[123] H. Pi, P. Avery, D. Green, J. Rohlf, and C. Tully, *Measurement of Missing Transverse Energy With the CMS Detector at the LHC*, Tech. Rep. CMS-NOTE-2006-035.

[124] S. Cucciarelli, M. Konecki, D. Kotlinski, and T. Todorov, *Track reconstruction, primary vertex finding and seed generation with the Pixel Detector*, Tech. Rep. CMS-NOTE-2006-026.

[125] A. Strandlie and W. Wittek, *Propagation of Covariance Matrices of Track Parameters in Homogeneous Magnetic Fields in CMS*, CMS-NOTE-2006-001.

[126] W. Adam, B. Mangano, T. Speer, and T. Todorov, *Track Reconstruction in the CMS tracker*, CMS-NOTE-2006-041.

[127] R. B. R. Kalman, *New Results in LInear Filtering and Prediction Theory*, Transaction of the ASME, Series D. J. Basic Eng., Vol 83, 1961.

[128] F. P. Schilling, *Track Reconstruction and Alignment with the CMS Silicon Tracker*, Tech. Rep. CMS-CR-2006-061, CERN, 2006.

[129] T. Speer, K. Prokofiev, R. Frühwirth, W. Waltenberger, and P. Vanlaer, *Vertex Fitting in the CMS Tracker*, CMS-NOTE-2006-032.

[130] R. Frühwirth, W. Waltenberger, and P. Vanlaer, *Adaptive Vertex Fitting*, Tech. Rep. CMS-NOTE-2007-008, CERN, 2007.

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

[132] C. Weiser, *A Combined Secondary Vertex Based B-Tagging Algorithm in CMS*, CMS-NOTE-2006-014.

[133] J. D'Hondt, S. Lowette, and J. Heyninck, *Electron and muon reconstruction in single leptonic ttbar events*, tech. rep., 2006. CMS-NOTE-2006-024.

[134] B. Mohr, *Measurement of the Top Quark Mass with a Matrix Element Method in the Lepton Plus Jets Channel at CDF*, hep-ex/0605083.

[135] D0 Collaboration, V. M. Abazov *et al.*, Phys. Rev. **D75** (2007) 092001.

[136] LEP Collaboration, *A Combination of Preliminary Electroweak Measurements and Constraints on the Standard Model*, hep-ex/0412015.

[137] BABAR Collaboration, B. Aubert *et al.*, Phys. Rev. Lett. **92** (2004) 071802.

[138] BABAR Collaboration, B. Aubert *et al.*, Phys. Rev. Lett. **93** (2004) 011803.

[139] BABAR Collaboration, B. Aubert *et al.*, Phys. Rev. **D69** (2004) 111103.

[140] The LEP Working Group for Higgs Boson Searches, R. Barate *et al.*, Phys. Lett. **B565** (2003) 61–75.

[141] J. D'Hondt *et al.*, *Fitting of Event Topologies with External Kinematic Constraints in CMS*, CMS-NOTE-2006-023.

[142] *Event Shapes - Sphericity and aplanarity*, http://cepa.fnal.gov/psm/simulation/mcgen/lund/pythia_manual/pythia6.3/pythia6301/node213.html.

[143] R. Barlow, *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*. Wiley, 1989.

[144] CDF and D0 Collaboration, S. Cabrera, *Top quark pair production cross section measurements in* $p\bar{p}$ *collisions at* $\sqrt{s} = 1.96\ TeV$, Presented at 14th International Workshop on Deep Inelastic Scattering (DIS 2006), Tsukuba, Japan, 20-24 Apr 2006.

[145] F. Larios and F. Penunuri, *FCNC production of same sign Top quark pairs at the LHC*, hep-ph/0311056.

[146] S. Kraml and A. R. Raklev, AIP Conf. Proc. **903** (Sep, 2006) 225. 20 p.

[147] J. D'Hondt, J. Heyninck, and S. Lowette, *Measurement of the cross section of single leptonic ttbar events*, CMS-NOTE-2006-064.

[148] D0 Collaboration, M. Abazov *et al.*, Phys. Lett. B **626** (Apr, 2005) 45–54.

[149] *PDF systematics estimation in the online LHAPDF manual*, `http://projects.hepforge.org/lhapdf/manual#tthFtNtAAB`.

[150] C. Saout, A. Scheurer, F.-P. Schilling, and A. Schmidt, *Impact of Tracker Misalignment on the CMS b-Tagging Performance and Considerations for a robust Algorithm*, CMS Analysis Note AN-2007/047.

[151] J. D'Hondt, J. Heyninck, and S. Lowette, *Top quark mass measurement in single leptonic ttbar events*, CMS-NOTE-2006-066.

[152] DELPHI Collaboration, Phys. Lett. **B511** (2001) 159.

[153] G. Breit and E. Wigner, Phys. Rev. **49** (Apr, 1936) 519–531.

[154] B. Efron, Ann. Stat. **7** (1979), no. 1, 1–26.

[155] P. V. M. J. D'Hondt, CMS Analysis Note AN-2007/029.

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

[157] B. Abbott *et al.*, Phys. Rev. D **61** (Mar, 2000) 072001.

[158] M. Davids *et al.*, *Measurement of top-pair cross section and top-quark mass in the di-lepton and full-hadronic channels with CMS*,.

[159] I. Borjanovic *et al.*, Eur. Phys. J. C **39** (Mar, 2004) s63–s90. 47 p.

[160] V. Buge, C. Jung, G. Quast, A. Ghezzi, M. Malberti, and T. Tabarelli de Fatis, *Prospects for the Precision Measurement of the W Mass with the CMS Detector at the LHC*, CMS-NOTE-2006-061.

[161] S. Haywood *et al.*, *Electroweak physics*, hep-ph/0003275.

Summary

This thesis presents the top quark physics potential inferred from the study of semi-muonic decaying $\mathrm{t\bar{t}}$-events produced in proton-proton collisions by the CMS detector at the LHC. In these top quark pair events one of the W-bosons decays into a muon, while the other decays hadronically, $\mathrm{t\bar{t}} \rightarrow b\,W\,\bar{b}\,\bar{W} \rightarrow b\,q\,\bar{q}'\,\bar{b}\,\mu\,\bar{\nu_{\mu}}$. Apart from a comparison of several Monte-Carlo event generators to motivate the use of the `AlpGen` program in Chapter 2, a study of the jet clustering performance was introduced in Chapter 4, before arriving to the three main analyses. Based on several quality markers the most optimal jet clustering configuration for the iterative cone, midpoint cone and k_T-algorithm were determined, and the influence of the magnetic field and the detector response was examined. The iterative cone algorithm with opening angle 0.5 was selected to be used throughout the work.

In a first main analysis, top quark events are demonstrated to be a powerful tool in the calibration of the jet energy scale (JES). The possibility to select semi-muonic $\mathrm{t\bar{t}}$-events with a high purity allows the definition of an unbiased estimator of the inclusive light quark JES. Based on the topological constraint that requires the invariant mass of the two light quark jets to be equal to the world average W-boson mass, the relative jet miscalibration factor can be extracted from a Gaussian fit on the reconstructed m_W-spectrum. For $1\,\mathrm{fb}^{-1}$ of accumulated data, a total uncertainty on the measurement of the calibration factor ΔC of 1.4% is found, by far dominated by systematic uncertainties. Especially the effect of pile-up collisions and the uncertainty on the channel background level and shape were found significant, and should be studied carefully. It was shown that a cut on the number of reconstructed primary vertices was highly efficient to reduce the systematic shift due to pile-up collision. In the last chapter of this thesis however, a possible technique to measure the channel background level and shape in data was sketched.

Apart from the use of top quarks for commissioning and calibration issues, precision measurements of quantities such as the $\mathrm{t\bar{t}}$ cross-section and its mass are envisaged in CMS. For both physics observables an analysis method is presented in this thesis based on $1\,\mathrm{fb}^{-1}$ of data. The cross-section measurement is treated as a simple counting experiment, an approach favoured by the high signal event purity in the selected sample. In order to enhance this purity, a combined Likelihood Ratio discriminator was developed. A cut on this observable resulted in a final S/B ratio of 5.2 for a signal event selection efficiency of 5.8%. When considering all $\mathrm{t\bar{t}}$-decays as signal and including the expected QCD event rate in the calculation, this ratio becomes ∼20. Also the cross-section estimate is by far dominated by systematic uncertainties:

$$\Delta\sigma^{tot}_{semi-\mu} \;=\; 2.1\,(stat.) \,\oplus\, 23.5\,(syst.) \,\oplus\, 10\,(lumi.)\;\% \;=\; 26.0\,\%\;,$$

with a leading contribution arising from the uncertainty on the channel background level (16 % for a rescaling factor of 5). Although an increased threshold on the signal event enhancing LR-discriminant allows to reduce $\Delta\sigma^{tot}_{semi-\mu}$ significantly, such an optimisation of the method should occur in parallel with a precise determination of the channel background level from data itself.

For the measurement of the top quark mass three estimators were defined. Each method starts from the four jets selected as the hadronic $\mathrm{t\bar{t}}$ decay products. A kinematic fit imposing the W-boson mass constraint to the two light quark jet candidates is applied. In a first method the invariant mass spectrum of the hadronic decaying top quark is fitted with a Gaussian, and the mean value of the fit is considered as an estimate of the true top quark mass M_{top}. In order to choose the three hadronic decay products among the four selected jets, another Likelihood Ratio discriminant was constructed. This M_{top} measurement is compared with the result obtained by two ideogram based techniques. Especially the method in which the scanned ideogram is convoluted with a template function to pass from the reconstructed to the theoretical top quark mass space yields the most optimal results. Both the statistical properties and the robustness towards systematic effects are found more optimal for this mass estimator. The largest uncertainty in the presented top quark mass measurement is expected from the uncertainty on the b-quark jet energy scale:

$$\Delta^{tot} M_{top} \;=\; 0.56\,(stat.) \,\oplus\, 0.96\,(syst.) \,\oplus\, 3.8\,(JES\,shift)\ \mathrm{GeV/c^2} \;=\; 4.0\ \mathrm{GeV/c^2},$$

Similarly to the approach presented by DØ and CDF, the systematic effect due to the uncertainty on the JES should be factorised in several contributions to isolate the difference in jet response between light and heavy quark flavours. Via an in-situ light quark JES calibration based on the W-boson mass constraint, only the uncertainty on this difference will remain a significant source of uncertainty. However, the ultimate goal to measure the top quark mass with a precision of $1\,\mathrm{GeV/c^2}$ will require a perfect understanding of the whole detector response.

All methods studied in this thesis will be relevant to understand the top quark events produced in the proton-proton collisions at the LHC.

Samenvatting

Dit doctoraatswerk beoogt een schatting van de meetprecisie haalbaar in de top quark sector van het Standaard Model, vertrekkende van semi-muonisch vervallende $\mathrm{t\bar{t}}$-gebeurtenissen geproduceerd bij proton-proton botsingen in de CMS detector aan de LHC. In deze gebeurtenissen vervalt één van beide W-bosonen naar een muon, terwijl het andere desintegreert naar twee quarks, $\mathrm{t\bar{t}} \rightarrow b\,W\,\bar{b}\,\bar{W} \rightarrow b\,q\,\bar{q}'\,\bar{b}\,\mu\,\bar{\nu_\mu}$. Naast de motivatie om het `AlpGen` programma te gebruiken als gebeurtenis generator in Hoofdstuk 2, introduceerden we in Hoofdstuk 4 een jetstudie alvorens tot de belangrijkste analyses te komen. Deze jetstudie baseert zich op de uitkomst van enkele kwaliteitsindicatoren om de beste configuratie van drie jet klusteringsalgoritmes te bepalen. Het zogenaamde 'iterative cone' algoritme met een openingshoek van 0.5 werd geselecteerd voor de verdere studies.

In een eerste belangrijke analyse demonstreerden we het belang van top quark gebeurtenissen in de calibratie van de jet energie schaal (JES). De mogelijkheid om met grote zuiverheid semi-muonische $\mathrm{t\bar{t}}$ gebeurtenissen te selecteren, laat toe een robuste schatter te definiëren voor de jet energie schaal van lichte quarks. Hierbij eisten we dat de invariante massa distributie van de twee lichte quark jets in de gebeurtenissen een Gaussisch gefitte piekwaarde heeft gelijk aan de goed gekende W-boson massa. Elke afwijking van de gefitte waarde is te wijten aan een miscalibratie van de JES, en kan geschat worden door de JES met een constante factor te variëren. De analyse van $1\,\mathrm{fb}^{-1}$ aan gegevens resulteert in een totale onzekerheid op de geschatte JES correctiefactor van 1.4%. Vooral de systematische onzekerheid te wijten aan 'pile-up' gebeurtenissen en in de verwachte hoeveelheid aan achtergrond gebeurtenissen bepalen deze precisie. Er werd aangetoond dat een snede op het aantal gereconstrueerde primaire vertices in de gebeurtenis toelaat om het effect van pile-up sterk te reduceren. Ook werd in het laatste hoofdstuk een methode geschets om de hoeveelheid process achtergrond direct te meten in de data.

Naast het gebruik van top quark gebeurtenissen voor de calibratie van de detector, beoogt CMS ook precisie metingen van bijvoorbeeld de top quark massa M_{top} en de productie werkzame doorsnede $\sigma_{semi-\mu}$. Voor beide observabelen werd in dit doctoraatswerk een analyse gepresenteerd gebaseerd op $1\,\mathrm{fb}^{-1}$ aan gegevens. Door de grote signaal/achtergrond (S/A) verhouding is het mogelijk de werkzame doorsnede te meten via een tel experiment. Om deze verhouding nog verder op te drijven, werd een Likelihood discrimant opgesteld. Een snede hierop resulteerde in een waarde voor S/A van 5.2 met een efficiëntie om signaal te selecteren gelijk aan 5.8%. Wanneer alle andere $\mathrm{t\bar{t}}$ vervalkanalen ook beschouwd worden als signaal en de QCD achtergrond in rekening wordt gebracht, bekomen we een waarde S/A = ∼20. Ook de precisie op de

$\sigma_{semi-\mu}$-meting is sterk gedomineerd door systematische effecten:

$$\Delta\sigma^{tot}_{semi-\mu} \;=\; 2.1\,(stat.) \,\oplus\, 23.5\,(syst.) \,\oplus\, 10\,(lumi.)\;\% \;=\; 26.0\,\%\;,$$

waarbij vooral de onzekerheid op de hoeveelheid achtergrond (16% voor een onzekerheidsfactor van 5) bepalend is. Ondanks het feit dat de totale onzekerheid kan gereduceerd worden door een strengere snede op de Likelihood discriminant, moet deze optimalisatie van de methode parallel gebeuren met een meting van de hoeveelheid achtergrond in de data zelf.

Voor de meting van de top quark massa werden drie schatters gedefinieerd. Elk van deze methodes vertrekt van de vier jets die werden geselecteerd als hadronische $\mathrm{t\bar{t}}$ vervalprodukten. Een kinematische fit werd gebruikt om de W-boson massa op te leggen als invariante massa voor beide lichte quark jets. In een eerste methode werd het invariant top quark massa spectrum Gaussisch gefit, en werd de piekwaarde beschouwd als meting van M_{top}. Om te bepalen welke drie van de vier jets afkomstig zijn van het hadronische top quark verval, werd een tweede Likelihood discriminant opgesteld. Deze M_{top} meting werd vergeleken met het resultaat bekomen met twee ideogram gebaseerde meettechnieken. Vooral de methode waarbij het gescande ideogram werd geconvolueerd met een transferfunctie om over te gaan van de experimentele naar de theoretische M_{top}-ruimte leverde de beste resultaten op. Zowel de statistische eigenschappen van de schatter als de robustheid ten opzichte van systematische effecten bleken optimaal voor deze techniek. De grootste onzekerheid in de meting van de top quark massa is verwacht voor de onzekerheid op de b-quark jet energie schaal:

$$\Delta^{tot}M_{top} \;=\; 0.56\,(stat.) \,\oplus\, 0.96\,(syst.) \,\oplus\, 3.8\,(JES\,shift)\ \mathrm{GeV/c^2} \;=\; 4.0\ \mathrm{GeV/c^2},$$

Net zoals gepresenteerd in de DØ en CDF collaboraties, zal een factorisatie van de JES correctiefactor toelaten dit effect te reduceren. Zo kan de JES met op data gebaseerde technieken gecorrigeerd worden als functie van de jet richting en energie, en kan een b-jet specifieke correctieterm bepaald worden uit Monte-Carlo gebeurtenissen. Een meting van de top quark massa met een totale nauwkeurigheid van $1\,\mathrm{GeV/c^2}$ zal dus een zeer goed begrip van de detector respons vereisen.

Alle methoden en technieken gepresenteerd in dit doctoraatswerk zullen uitermate nuttig zijn voor de studie van top quark gebeurtenissen geproduceerd bij proton-proton botsingen aan de LHC.

Het denken mag zich nooit onderwerpen,
noch aan een dogma,
noch aan een partij,
noch aan een hartstocht,
noch aan een belang,
noch aan een vooroordeel,
noch aan om het even wat,
maar uitsluitend aan de feiten zelf,
want zich onderwerpen betekent het einde van alle denken.

Henri Poincaré
21 november 1909

Uit een redevoering ter gelegenheid van de 75ste verjaardag van de Université Libre de Bruxelles.