Data-driven multi-jet and V+jets background estimation methods for top quark pair production at CMS

Thèse présentée par

Grégory Hammad

En vue de l'obtention du grade de

Docteur en Sciences

(Faculté des Sciences)

de l'Université Libre de Bruxelles

Bruxelles, Belgique

Doctoral examination commission

Prof. Dr. Thomas Hambye (ULB), *chair*
Prof. Dr. Catherine Vander Velde (ULB), *supervisor*
Prof. Dr. Laurent Favart (ULB), *secretary*
Prof. Dr. Freya Blekman (VUB)
Dr. Eric Chabert (IPHC)
Dr. Andrea Giammanco (UCL)

Research funded by a Ph. D. grant of the *Fonds pour la formation à la Recherche dans l'Industrie et dans l'Agriculture* (FRIA).

Introduction

Whether it is believed that the world is made of four elements, air, fire, water and wind, or made of elementary particles, it proceeds from the same underlying idea ; the Nature is revealed through its ordering into fundamental elements and the interplay between them. This idea reached its apogee in high energy physics in the late 1970's with the advent of the so-called Standard Model, a theory embedding the building blocks of all the visible matter in the Universe, six quarks and six leptons, and describing their interactions. During the following decades, predictions of the Standard Model have been compared with experimental results for interactions occurring at higher and higher energies and have been found in agreement to an astonishing level of precision. However, the Standard Model cannot be considered as the final theory in particle physics as it exhibits several shortcomings ; it does not include gravity and fails in providing satisfactory answers to some experimental observations. A possible way to overcome this issue is to assume that the Standard Model is a low-energy approximation of a more fundamental theory. Although able to describe particle interactions occurring at energies up to about 100 GeV, at higher energies, the Standard Model must be superseded by a new theory. Many such extensions of the Standard Model have been proposed, including Super-Symmetry or theories involving extra spatial dimensions. The top quark is the heaviest particle present in the Standard Model and has a mass close to 170 GeV$/c^2$, suggesting that any deviation from the predictions of the Standard Model are more likely to be visible in processes involving top quarks. Signs of new physics may appear, for example, as new production or decay channels and therefore lead to an excess in the top quark production with respect to the Standard Model predictions.

So far, the signs of new physics have been too faint to be discovered at the preceding collider energies but, with the recent advent of the new proton accelerator, the *Large Hadron Collider* (LHC), of the European Organization for Nuclear Research (CERN) and its unprecedented centre-of-mass energy of 7 TeV, physicists have for the first time the opportunity to search for signs of new physics at an energy scale of the order of the TeV in the partonic centre-of-mass frame .

The study presented in this thesis focuses on top quark pair production during proton-proton collisions at LHC, in the *Compact Muon Solenoid* (CMS) detector. This study exploits the top quark pair semi-muonic decay channel where one of the W boson coming from the top quark decay gives a pair of quarks whereas the other W boson decays into a muon and a neutrino in the final-state. As other processes produced at the LHC could also lead to the same final-state, called background processes, any accurate measurement of the top quark pair production involves a precise estimate of the background contribution. Unfortunately, the cross-sections of the relevant background processes for the top quark pair production suffer from large theoretical uncertainties. Methods have thus been developed in this thesis to estimate, from data, the two main background processes, namely the multi-jet production and the vector boson production associated with jets. The cross-section of the former is almost 10^6 times larger than the top quark pair production cross-section. Although it is possible to reject this background to a large extent, the number of background events passing the selection criteria remains important and needs to be estimated. For this purpose, the so-called ABCD method has been used. Its principle consists in defining three control regions dominated by the multi-jet background. The vector boson production cross-section is only ~ 450 times greater than the top quark pair cross-section but this background is more difficult to reject than the multi-jet one and cannot be estimated with the ABCD method because of the presence of a W boson produced in the final state, as for top quark pairs. Another method has thus been developed in this thesis to deal with this particular background ; the method is based on the number of jets originating from b-quark hadronization in the final state. Both methods have been validated and their performances evaluated with Monte-Carlo simulations.

In Chapter 1, the Standard Model of elementary particles is introduced ; the elementary particles and their interactions are described. The concept of spontaneous symmetry breaking and its application to generate particle masses are explained. The concept of running coupling constant, essential when modelling strong interactions, is also discussed, followed by a review of the shortcomings and possible extensions of the Standard Model. Chapter 2 focuses on some of the theoretical aspects of proton collisions ; such collisions are complex because protons are composite objects. They are made of quarks and gluons, called partons. This chapter presents also the simulation of the mechanisms involved during the evolution of the partons participating to the interactions as well as the evolution of the proton remnants. The last part of this chapter is dedicated to the Monte-Carlo generators used in this thesis to simulate proton-proton collisions. Methods to link the simulation of the hard interaction between proton constituents and their subsequent evolution are emphasized. The experimental context of this study is provided in Chapter 3 by introducing the CERN and the LHC. It is followed by a description of the CMS detector. Chapter 4 presents a survey of the algorithms needed to reconstruct, from data, the different physics objects used in this analysis as muons or jets of particles. In Chapter 5, some of the top quark properties are reviewed together with the top quark pair production and decay mechanisms. The agreement between the theoretical predictions for these properties and their measurements are also discussed. The relevant background processes for the top quark pair production are described with emphasis on the theoretical uncertainties related to their simulations. Criteria recommended by the CMS collaboration to select top quark pair events, based on the knowledge of the physics objects present in the final state, are presented, together with a possible improvement developed during this thesis work, concerning the jet selection. Finally, the last two chapters are dedicated to methods developed, using Monte-Carlo simulations, to estimate, from data, the level of background left after the top quark pair event selection ; Chapter 6 introduces the ABCD method, used to estimate the multi-jet background while Chapter 7 presents the new method to estimate the production of vector boson associated with jets. In both cases, the principles of the methods are described and the performances of the estimations are evaluated together with their statistical and systematic uncertainties. The final chapter contains the conclusions of the results obtained in this thesis.

Contents

“There is no absolute knowledge.
And those who claim it,
whether they are scientists or dogmatists,
open the door to tragedy.
All information is imperfect.
We have to treat it with humility.
That is the human condition.”

— Jacob Bronowski, *The Ascent of Man.*

1

The Standard Model of elementary particles

Elementary particle physics describes Nature at the sub-atomic scale. Since the beginning of the twentieth century, it is understood that, at this scale, quantum mechanics is required. In this framework, Nature is described in terms of probability ; *God plays dices!* A particle is related to a probability amplitude, whose absolute square give the probability to find this particle at different places and different states. Generally, this amplitude is time and position dependent. On the other hand, since the ninetieth and the early twentieth century, several fundamental forces of Nature were also well described, with respect to experimental observations, in terms of fields. In this description, the field maps the value of the force to each space-time point. This was the case for the electrical and magnetic fields, whose dynamics were described by the Maxwell's equations.

Nowadays, elementary particles and their interactions are described by a relativistic quantum field theory called the *Standard Model* (SM), to which this chapter is dedicated. Within this theory, all known matter in the Universe is made of spin-$1/2$ particles, the fermions, whose interactions consist in the exchange of spin-1 particles, the bosons, seen as the mediators of these interactions. Over the last decades, the predictive power of the

Standard Model has been tested up to unprecedented precision. But despite its success, some theoretical predictions and experimental observations do not find any explanation in the framework of the Standard Model.

In section 1.1, the building blocks of the Standard Model, the fermions and bosons, are reviewed. Then, in section 1.2, the concept of symmetry is introduced, from which interactions naturally arise. Although the Standard Model proved to be a predictive theory, it does not predict mass to the particles. As explained in section 1.3, mass terms can be derived as a consequence of a spontaneous symmetry breaking mechanism. Section 1.4 introduces the concept of renormalization, which is required by the theory in order to predict physically observable quantities. Finally, some of the shortcomings of the Standard Model are reviewed in section 1.5, as well as possible extensions of the Standard Model.

1.1 Elementary particles and fundamental interactions

The Standard Model [1] involves 12 different elementary particles and their associated anti-particles as fundamental building blocks of our Universe. These particles, the fermions, are grouped into two categories, 6 quarks[1] and 6 leptons[2], according to their behaviour under the different fundamental interactions. Quarks and leptons are further divided into three families, also called *generations*, with similar properties and increasing masses. Each family of quarks comprises two quarks of different types, with fractional electrical charge and each lepton family is composed of a lepton with integer electrical charge and its associated neutral lepton, called *neutrino*. The different types of quarks and leptons are called *flavour*. These informations are summarized in Table 1.1. All stable and visible matter in the Universe is made of first generation particles ; the *up* and *down* quark are the elementary constituent of the proton (uud) and the neutron (udd) and therefore the constituents of the nuclei of all chemical elements present in the Mendeleiev's periodic table. Electrons are associated to these nuclei to form atoms and are involved in the chemical bounds between them. Electron neutrinos are emitted during nuclear radioactive β-decay of nuclei.

[1] Murray Gell-Mann $[1929, -]$, 1969 Physics Nobel Prize, took this name from the book *Finnegan's Wake* by James Joyce

[2] Léon Rosenfeld $[1904, 1974]$, a Belgian physicist, coined the name lepton in 1948 from the Greek $\lambda\epsilon\pi\tau\acute{o}\nu$, meaning thin, small.

	Q_e	1^{st} generation	2^{nd} generation	3^{rd} generation
Quarks	$+2/3$	Up (u)	Charm (c)	Top (t)
	$-1/3$	Down (d)	Strange (s)	Bottom (b)
Leptons	-1	Electron (e)	Muon (μ)	Tau (τ)
	0	Electron neutrino (ν_e)	Muon neutrino (ν_μ)	Tau neutrino (ν_τ)

Table 1.1: Overview of the particles of the three fermion generations included the Standard Model. Q_e is the electrical charge, in unit of the elementary charge, *e*.

The Standard Model provides a mathematical formulation for three of these fundamental interactions :

1. the electro-magnetic interaction (EM) describes the interaction between electrically charged particles in terms of exchange of massless neutral photons (γ).
2. the weak interaction concerns all particles, quarks and leptons. This fundamental interaction is mediated by three different massive bosons : the $W^{\pm}$ and the Z bosons.
3. the strong interaction is responsible for the interaction between particles carrying a colour charge like the quarks. The mediators of this interactions, the eight massless gluons (g), carry also a colour charge and therefore interact with each other. Leptons are colourless and thus do not undergo strong interactions.

Historically, electro-magnetic interactions were first described by a unified theory, the *Quantum Electrodynamic* (QED). Later on, a successful description of the electro-magnetic and weak interactions in a unified way was achieved : the electro-weak theory. Similarly, strong interactions are described by the *Quantum Chromodynamic* (QCD). The fourth fundamental interaction, Gravity, is not included in the Standard Model. A lot of efforts are made to develop a satisfactory quantum formulation of the gravitational interaction that could be included in an extended new Standard Model for particle physics. However, at the current experiments, Gravity can safely be neglected compared to the other interactions.

1.2 Symmetries and interactions

In the Standard Model, the electro-magnetic, weak and strong forces are described in terms of quantum fields, called *gauge fields* whose physical manifestations are the bosons. In this section, we shall illustrate how these gauge fields arise from symmetry principles.

In quantum mechanic, a physical state is represented by a so-called *state vector* in a complex vector space, the Hilbert space. Following Dirac's notation for these state vectors, the probability for a particle to evolve from a physical state described by $|\psi\rangle$ to another physical state described by $|\phi\rangle$ is given by :

$$|\langle\psi|\phi\rangle|^2 \tag{1.1}$$

which remains invariant under any transformation R of a group whose representation U in the Hilbert space correspond to a unitary operator, $U^\dagger U = 1$:

$$|\psi'\rangle = U|\psi\rangle \;\Rightarrow\; |\langle\psi'|\phi'\rangle|^2 = |\langle\psi|\,U^\dagger U\,|\phi\rangle|^2 = |\langle\psi|\phi\rangle|^2 \tag{1.2}$$

Since the probability for a particle to evolve from a physical state to another one remains identical under any transformation like R, the equations of motions must also remain invariant under such transformations. These equations are deduced from the Euler-Lagrange equation which is derived from the *Principle of Least Action*. According to this principle, the motion of a particle between the times t_1 and t_2 follows the path which minimizes the action defined as the integral over the time of the Lagrangian density $\mathcal{L}$ between t_1 and t_2. Therefore, the invariance of the equations of motions can be retained by requiring the invariance of the Lagrangian density.

In the case of a free moving spin-$1/2$ particle of mass m, such as fermions, the Lagrangian density is given by :

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

where ψ is the Dirac representation of spin-$1/2$ particle wave-functions, called spinor, $\bar{\psi}$ its adjoint spinor, defined as $\bar{\psi} = \psi^\dagger\gamma^0$ with $\psi^\dagger$ being the hermitian conjugate of ψ and γ^μ, the Dirac matrices. The index μ runs over the four space-time coordinates[3].

In order to study the effects of requiring the invariance of the Lagrangian density, let us consider a local[4] phase transformation, U[5], also called *gauge* transformation, which

[3]Throughout this thesis, the Einstein summation convention, also called Einstein notation, is used : $a^\mu b_\mu = a^0b_0 + a^1b_1 + a^2b_2 + a^3b_3$

[4]From the physic's point of view, the existence of an invariance under a transformation of parameter θ means that this parameter cannot be measured and has no physical meaning. Therefore, it is reasonable to let this parameter be a function of the space-time coordinates to retain the full generality of the reasoning.

[5] A representation U of this transformation in the Hilbert space can be obtained by exponentiation of the vector $\vec{\theta}$:

$$\psi' = U(\vec{\theta}(x_\mu))\psi = e^{i\vec{\theta}(x_\mu)\cdot\frac{\vec{\tau}}{2}}\psi \tag{1.4}$$

depends on a continuous parameter $\vec{\theta}$, such as :

$$\psi \;\rightarrow\; \psi' = U(\vec{\theta}(x_\mu))\psi \tag{1.6}$$

where x_μ are the space-time coordinates. This transformation should leave the Lagrangian density invariant.

The adjoint spinor transforms as :

$$\bar{\psi}' = U^{-1}(\vec{\theta}(x_\mu))\bar{\psi} \tag{1.7}$$

and therefore, the last part ($m\bar{\psi}\psi$) of the Lagrangian density (Eq. 1.3) remains invariant. However, the derivative of ψ :

$$\partial_\mu\psi' = U(\vec{\theta}(x_\mu))\partial_\mu\psi + \partial_\mu U(\vec{\theta}(x_\mu))\psi \tag{1.8}$$

breaks this invariance. This issue can be solved by using a covariant derivative, D_μ instead of the usual space-time derivative ∂_μ :

$$D_\mu = \partial_\mu - ig\frac{\vec{\tau}}{2}\cdot\vec{A}_\mu \tag{1.9}$$

where $\vec{A}$ is a new vector field, called *gauge field* and g is a constant, called coupling constant, which defines the strength of the interaction between the gauge field and the spin-$1/2$ particle. This new vector field is introduced so that the covariant derivative transforms as required in order to retain the local gauge invariance of the Lagrangian :

$$D_\mu\psi' = U(\vec{\theta}(x_\mu))D_\mu\psi \tag{1.10}$$

The gauge field compensates the modifications induced by the local gauge transformation :

$$\vec{A}'_\mu = U(\vec{\theta}(x_\mu))\vec{A}_\mu U^{-1}(\vec{\theta}(x_\mu)) - \frac{i}{g}\left[\partial_\mu U(\vec{\theta}(x_\mu))\right]U^{-1}(\vec{\theta}(x_\mu)) \tag{1.11}$$

---

where $\vec{\tau}$ represents the generators of the group formed by these transformations. The generator properties are defined by the commutator algebra (Lie algebra) :

$$[\tau_a, \tau_b] = if_{abc}\tau_c \tag{1.5}$$

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

By substituting the new covariant derivative in Equation (1.3), one obtains :

$$\mathcal{L} = i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi + g\bar{\psi}\gamma^{\mu}\psi\frac{\vec{\tau}}{2}\cdot\vec{A}_{\mu} \tag{1.12}$$

The new term corresponds to the coupling of the gauge field to the spin-$1/2$ particle.

More generally, the number of gauge fields that needs to be introduced is equal to the number of generators of the symmetry group which is considered.

1.2.1 $SU(2)_L \times U(1)_Y$ symmetry and electro-weak interactions

In order to describe both the electro-magnetic and the weak interactions in a unified way, four gauge fields need to be introduced to account for the four associated bosons : the $W^{\pm}$ and Z^0 bosons and the photon, γ. As the number of gauge fields is equal to the number of independent generators of the symmetry group, this restraints the choice for a possible symmetry group.

Based on the $SU(2) \times U(1)$ group, the unification of the electro-magnetic and the weak interactions into a non-Abelian gauge theory, also known as the *Yang-Mills* theory, was first successfully achieved, independently, by S. Weinberg [2] and by A. Salam [3], based on the previous attempts made by S. Glashow [4] and by A. Salam and J. Ward [5, 6]. However, at that time, Yang and Mills had already pointed out that unbroken non-Abelian gauge theories involve massless gauge bosons. Furthermore, the renormalizability of such theories was still an open question.

Salam and Weinberg overcame the issue of massive gauge bosons by using spontaneous symmetry breaking mechanisms which allow to retain the invariance of the Lagrangian of the gauge theory while giving mass to the gauge bosons. The proof of the renormalizability of Yang-Mills theory was provided by 't Hooft and Veltman [7]. The Weinberg-Salam theory is often referred to as the *Electro-weak* theory.

The covariant derivative of this theory is given by :

$$D_{\mu} = \partial_{\mu} - ig\vec{T}\cdot\vec{W}_{\mu} - i\frac{g'}{2}YB_{\mu} \tag{1.13}$$

where $\{W^a_{\mu}\}_{a\in[1,3]}$ are the $SU(2)$ gauge fields associated the $W^{\pm}$ and Z^0 bosons, $\{T_a = \tau_a/2\}_{a\in[1,3]}$ the generators of the $SU(2)$ group, formed with 2×2 Pauli matrices (τ_a), B_{μ} is the $U(1)$ gauge field associated to the photon and Y the generator of the $U(1)$ group.

The universal coupling to the gauge fields are quantified by the constant g and g'. The representations of $\vec{T}$ and Y are the weak isospin and weak hypercharge operators. They have been defined such as to correspond to the electro-magnetic charge operator Q via Gell-Mann Nishijima formula :

$$Q = T_3 + \frac{1}{2}Y \tag{1.14}$$

where T_3 is the 3^{rd} component of the weak isospin.

Experimental studies have revealed that weak interactions distinguish between left-handed and right-handed fermions[6] : only left-handed particles undergo weak interactions. But both left- and right-handed particles may carry an electrical charge and then be affected by the electro-magnetic interaction. Consequently, within the electro-weak theory, leptons are represented by left-handed $SU(2)_L$ doublets :

$$L_e = \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L, \quad L_\mu = \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L, \quad L_\tau \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L \tag{1.15}$$

with quantum numbers $(T_3, Y) = (\pm 1/2, -1)$ and right-handed $SU(2)_L$ singlets :

$$R_{e,\mu,\tau} = e_R, \mu_R, \tau_R \tag{1.16}$$

with quantum numbers $(T_3, Y) = (0, -2)$. Since the neutrinos were originally considered as massless, right-handed neutrinos were not introduced in the original formulation of the electro-weak theory.

Quarks are represented in a similar way by left-handed $SU(2)_L$ doublets :

$$L_q^1 = \begin{pmatrix} u \\ d' \end{pmatrix}_L, \quad L_q^2 = \begin{pmatrix} c \\ s' \end{pmatrix}_L, \quad L_q^3 = \begin{pmatrix} t \\ b' \end{pmatrix}_L \tag{1.17}$$

with quantum numbers $(T_3, Y) = (\pm 1/2, 1/3)$ and right-handed $SU(2)_L$ singlets :

$$R_u^{1,2,3} = u_R,\ c_R,\ t_R \quad \text{and} \quad R_d^{1,2,3} = d_R,\ s_R,\ b_R \tag{1.18}$$

with quantum numbers $(T_3, Y) = (0, 4/3)$ and $(T_3, Y) = (0, -2/3)$ respectively. However, unlike for leptons, the weak eigenstates of the lower components of the quark doublets are not the mass eigenstates ; the quark mixing matrix that maps both sets of eigenstates is

[6]The property of being left- or right-handed is often referred to as the chirality, which coincide with the helicity for massless fermions

the so-called Cabibbo-Kobayashi-Maskawa (CKM) 3×3 unitary matrix :

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{1.19}$$

Historically, N. Cabibbo first formulated the idea of quark mixing in 1963 [8] to account for weak decays of strange particles ($u \leftrightarrow s$ transitions). His work was extended 10 years later by M. Kobayashi and T. Maskawa who demonstrated that mixing among three quark doublets yields a non-trivial phase that could account for the observed Charge-Parity (CP) symmetry violation [9]. In 2008, M. Kobayashi and T. Maskawa were awarded the Physics Nobel Prize for this work.

It is also not possible to identify any of the gauge fields $\vec{W}_\mu$ and B_μ with the electromagnetic field A_μ for example as two leptons from the same $SU(2)$ doublet carry different electrical charge. Instead,the observed interacting fields, the mass eigenstates, A_μ for the photon and $W^\pm_\mu$, Z_μ for the $W^\pm$ and Z^0 bosons respectively, arise from combinations of the gauge fields :

$$A_\mu = \frac{gW^3_\mu + g'YB_\mu}{\sqrt{g^2 + g'^2}} \quad W^\pm_\mu = \frac{W^1_\mu \mp W^2_\mu}{\sqrt{2}} \quad Z_\mu = \frac{gW^3_\mu - g'YB_\mu}{\sqrt{g^2 + g'^2}} \tag{1.20}$$

1.2.2 $SU(3)_C$ symmetry and strong interactions

The *Quark Model* was introduced in 1964 by Gell-Mann [10] and by Zweig [11, 12] to describe experimentally observed particles like Δ^{++}. In this model, such particle is described by a configuration involving three identical quarks ($\Delta^{++} = uuu$), although the existence of such configuration is forbidden by the Pauli's exclusion principle. On the other hand, allowed quark configurations such as qq or $\bar{q}\bar{q}$ were missing in the observed hadron spectrum. Both problems were solved by the introduction of a new quantum number ; the *colour*. This colour charge can take three distinct values : R(ed), B(lue) and G(reen). Due to this colour charge, quarks interact via strong interactions, mediated by massless gluons. Later on, in 1969, Feynman introduced the *Parton Model* [13] to explain experimental observations in electron-proton collisions made by the experiments of the *Standford Linear Accelerator Center* (SLAC). Similarly to the Rutherford's experiment, angular distributions of the scattered electrons revealed that protons have an internal structure. In this model, hadrons are made of weakly interacting point-like constituents, which will later by identified with the quarks. Finally, under the impulsion of Gross, Wilczek, Fritzsch and Weinberg among others, the *Quark Model* was promoted to a gauge theory, the *Quantum Chromodynamic*, which successfully accounted for both the weakness and the strength of strong interactions at high and low energy respectively. This was done by introducing the concept of *asymptotic freedom* with the help of an energy-dependent coupling strength. This concept conduced physicists to to perform QCD calculations within the framework of the perturbative theory : the so-called *perturbative QCD* (pQCD).

The mathematical formulation of the QCD theory is based on the symmetry group $SU(3)$. A fundamental representation of this group is formed with the three colour charges of the quarks. Requirement of non-abelian invariance of this theory under $SU(3)$ transformations leads to the introduction of $3^2 - 1 = 8$ gauge fields whose mass eigenstates are the gluons. The associated covariant derivative is given by :

$$D_\mu = \partial_\mu + ig\vec{T}\cdot\vec{G}_\mu \tag{1.21}$$

where $\{G^a_\mu\}_{a\in[1,8]}$ are the gauge fields and $\{T_a = \lambda_a/2\}_{a\in[1,8]}$ the generators of the $SU(3)$ group, formed with the 3×3 Gell-Mann matrices (λ_a). A review of one of the most remarkable features of the strong interactions, the *asymptotic freedom*, is given in section 1.4. The application of the QCD in proton-proton collisions is the subject of the chapter 2

1.3 Spontaneous Symmetry Breaking

To take into account the behaviour of left-handed and right-handed particles during electroweak interactions, it is necessary to separate the left and right components of the Dirac representations of spin-$1/2$ particle wave-functions by using the chirality operator[7] γ^5 :

$$\begin{aligned}\psi = \left[\frac{1}{2}(1-\gamma^5) + \frac{1}{2}(1+\gamma^5)\right]\psi &= \frac{1}{2}(1-\gamma^5)\psi + \frac{1}{2}(1+\gamma^5)\psi \\ &= \psi_L + \psi_R\end{aligned} \tag{1.22}$$

where $\psi_{L/R}$ are Weyl representation of spin-$1/2$ particles, known as the Weyl spinors.

[7] The chirality operator is defined as the product of the four Dirac matrices : $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$

Rewriting the Dirac Lagrangian, defined in Equation (1.3), using the Weyl spinors, a mixed mass term is obtained :

$$m\psi\bar{\psi} = m(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R) \tag{1.23}$$

This term breaks the required $SU(2)$ invariance because ψ_L and ψ_R are respectively a doublet and a singlet under $SU(2)$ transformations. Therefore, the fermions have to be treated as massless particles, in contradiction with experimental observations. Similarly, weak boson mass terms also break the gauge invariance. Moreover, the introduction by hand of such terms, at the cost of an explicit symmetry breaking, would nevertheless lead to a non-renormalizable theory and therefore spoil its predictive power. In order to establish a unified theory of the electro-magnetic and weak interactions, with massive fermions and massive weak gauge bosons, Weinberg and Salam used the mechanism of spontaneous symmetry breaking (SSB) which solves the problem of generating massive particles without explicitly violating the gauge invariance.

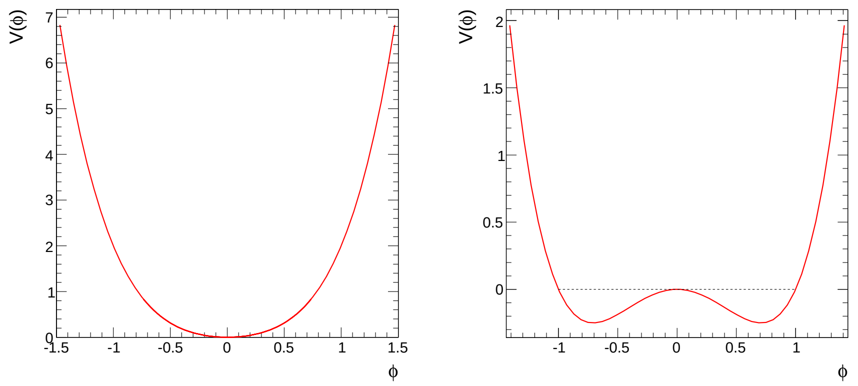
Spontaneous symmetry breaking occurs when the symmetry of a given law of physics does not hold for the solution of this law without explicit symmetry breaking input. Whence the name *spontaneous*. The concept of SSB was first developed in condensed matter physics and then transferred to quantum field theory, mainly due to the work of Nambu and G. Jona-Lasinio [14], in analogy with the breaking of gauge symmetry in the theory of superconductivity, developed by J. Bardeen, L. N. Cooper and J. R. Schrieffer in 1957, the so-called BCS theory [15]. In 1964, this concept of SSB was formulated by P. Higgs [16] and R. Brout and F. Englert [17], independently, in the context of local gauge symmetry in order to give mass to the gauge bosons. This formulation requires the introduction of a new scalar field. Within the high-energy physics community, this mechanism is often referred to as the *Higgs mechanism*.

Within the Standard Model, it is possible to generate masses for the $W^{\pm}$ and Z^0 bosons by introducing a $SU(2)$ doublet of complex scalar fields :

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \sqrt{\frac{1}{2}}\begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.24}$$

An invariant Lagrangian under $SU(2)$ transformations for such fields is given by :

$$\mathcal{L} = (D^\mu\phi)^\dagger(D_\mu\phi) - V(\phi) \text{ with } V = \mu^2\phi^\dagger\phi + \lambda\left(\phi^\dagger\phi\right)^2 \tag{1.25}$$



(a) **(b)**

Figure 1.1: One-dimensional representation of the potential V of the scalar field ϕ in the case $\mu^2 > 0$ (a) and $\mu^2 < 0$ (b).

where D_μ is the electro-weak covariant derivative introduced in section 1.2.1 and V is the potential of the scalar field ϕ. In the case $\mu^2 > 0$, the potential has its minimum at $\phi = 0$. The interesting case is $\mu^2 < 0$ for which the minimum is found at a finite value of $|\phi|$, ϕ_0, defined by the following relation :

$$\phi_0^\dagger \phi_0 = \frac{-\mu^2}{2\lambda} \equiv \frac{\upsilon^2}{2} \tag{1.26}$$

where υ is called the vacuum expectation value of the field ϕ. The set of values of $|\phi|$ for which the potential is minimum can be seen as a four-dimensional hypersphere. In a one-dimensional case, this set of values is reduced to two distinct values $\pm\upsilon/\sqrt{2}$ (Fig. 1.1). The ground-state does not respect the original symmetry of the Lagrangian anymore ($\phi \rightarrow -\phi$) ; the symmetry is said to be spontaneously broken. An arbitrary direction can be chosen in the four-dimensional space such that $\phi_1 = \phi_2 = \phi_4 = 0$ and $\phi_3^2 = \upsilon^2$. An expansion of the field is made around its minimum ϕ_0 :

$$\phi = \phi_0 + h = \sqrt{\frac{1}{2}} \begin{pmatrix} 0 \\ \upsilon + h \end{pmatrix} \tag{1.27}$$

where the four initial fields have been reduced to only one real scalar field, h, using a gauge transformation. The remaining field h is called the *Higgs* field whose physical manifestation is the Higgs boson.

It is now possible to rewrite the potential V for the scalar potential as following :

$$\begin{aligned} V(\phi) &= \frac{\mu^2}{2}(v+h)^2 + \frac{\lambda}{4}(v+h)^4 \\ &= V_0 + \lambda v^2 h^2 + \mathcal{O}(h^3) \end{aligned} \tag{1.28}$$

with $V_0 = \mu^2 \phi_0^\dagger \phi_0 + \lambda(\phi_0^\dagger \phi_0)^2 = -\lambda v^4/4$. The Higgs boson mass m_h now simply reads :

$$m_h = 2\lambda v^2 \tag{1.29}$$

Mass terms for the gauge bosons are generated by coupling of the electro-weak gauge fields to the vacuum Higgs field ϕ_0. These terms, can be made explicit by substituting ϕ_0 into the Lagrangian defined by Eq. (1.25). The relevant term, containing the covariant derivative as defined in Eq. (1.13), is the following :

$$\begin{aligned} (D^\mu \phi_0)^\dagger (D_\mu \phi_0) &= \left| \left(-i\frac{g}{2}\vec{\tau} \cdot \vec{W}_\mu - i\frac{g'}{Y} 2B_\mu \right) \phi_0 \right|^2 \\ &= \frac{1}{8} \left| \begin{pmatrix} gW_\mu^3 + g'YB_\mu & g(W_\mu^1 - iW_\mu^2) \\ g(W_\mu^1 + iW_\mu^2) & -gW_\mu^3 + g'YB_\mu \end{pmatrix} \begin{pmatrix} 0 \\ v \end{pmatrix} \right|^2 \\ &= \left(\frac{1}{2vg} \right)^2 \left[(W_\mu^1)^2 + (W_\mu^2)^2 \right] + \frac{1}{8} v^2 \left(W_\mu^3, B_\mu \right) \mathcal{M} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix} \end{aligned} \tag{1.30}$$

where $\mathcal{M}$ is a 2×2 matrix whose one of the eigenvalues is zero. By expressing the matrix $\mathcal{M}$ in an orthonormal basis made of its eigenvectors, it is possible to rewrite the previous equation as :

$$\begin{aligned} (D^\mu \phi_0)^\dagger (D_\mu \phi_0) = &\left(\frac{1}{2vg} \right)^2 \left[(W_\mu^1)^2 + (W_\mu^2)^2 \right] \\ &+ \frac{1}{8} v^2 \left[gW_\mu^3 - g'YB_\mu \right]^2 \\ &+ 0 \left[gW_\mu^3 + g'YB_\mu \right]^2 \end{aligned} \tag{1.31}$$

Now, in terms of physical electro-magnetic and weak fields, A_μ and $W_\mu^\pm$, Z_μ as defined in Eq. 1.20, one obtains :

$$(D^\mu \phi_0)^\dagger (D_\mu \phi_0) = M_W^2 W_\mu^+ W^{\mu,-} + \frac{1}{2} M_Z^2 Z_\mu Z^\mu + \frac{1}{2} M_A^2 A_\mu A^\mu \tag{1.32}$$

with $M_W = \frac{1}{2} v g$, $M_Z = \frac{1}{2} v \sqrt{g^2 + g'^2 Y}$ and $M_A = 0$.

As desired, after spontaneous symmetry breaking, one is left with three mass eigenstates, obtained by mixing the gauge fields, among which one is massless.

In a similar fashion, it is also possible to generate mass terms for the fermions with the same $SU(2)$ doublet of complex fields introduced in Eq. (1.24). More details can be found in [1].

1.4 Renormalization and running coupling constant

The Standard Model Lagrangian describes all the matter particles, the fermions, and their interactions, mediated by the gauge bosons. But working out predictions from the Lagrangian remains a highly technical and difficult task. Nevertheless, it is possible to simplify this procedure by associating to this Lagrangian a set of diagrams and rules, known as Feynman diagrams and Feynman rules respectively. Each Feynman diagram represents a contribution to the Lorentz-invariant scattering amplitude, related to the matrix element, $\mathcal{M}$, whose modulus-squared is proportional to the differential cross-section of the process :

$$d\sigma = \frac{1}{F}|\mathcal{M}|^2 d\phi \tag{1.33}$$

where F is a flux factor and $d\phi$ is Lorentz-invariant phase-space element, imposing four-momentum conservation.

These diagrams consist of lines representing particles and vertices where particles are either created or annihilated. Lines entering and leaving the diagram, called external lines, represent particles in the initial and final state of the process respectively. For these particles, the energy-momentum relation is satisfied. Internal lines represent virtual particles appearing in the intermediate stage of the process, for which the energy-momentum relation is not satisfied. At each vertex, four-momentum is conserved.

Feynman rules define how to calculate the contribution of each diagram to the scattering amplitude ; it associates for each vertex, a vertex factor which is proportional to the coupling constant, α, corresponding to the considered interaction and particles. It also associates a propagator factor for each internal lines, which is proportional to $1/(p^2 - m^2)$, where p and m are respectively the four-momentum and the mass of the virtual particle.

In principle, to calculate $\mathcal{M}$, it is necessary to take into account all of the infinite number of possible Feynman diagrams, including more and more loops of internal lines (Fig. 1.2).

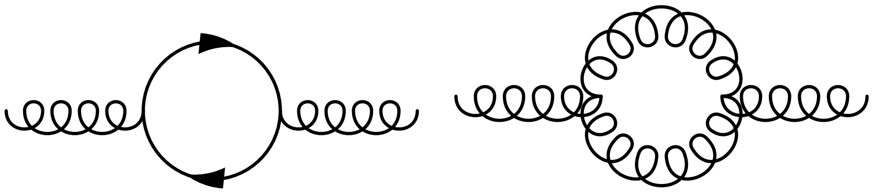


Figure 1.2: Example of fermion (a) and boson (b) loops introduced in Feynman diagrams accounting for higher-order QCD corrections in series expansions of a physical quantity in terms of α_s

However, the contribution of a diagram with n vertices is the order of α^n. Therefore, if $\alpha << 1$, the more vertices a diagram has, the less it contributes to the calculation of $\mathcal{M}$ and a perturbative calculation can be performed for any observable. The Feynman diagrams with the smallest number of vertices possible depict thus the process at leading-order and the diagrams with loops represent the higher-order quantum corrections.

Although the four-momentum is conserved at each vertex, the four-momentum p of the particle involved in a loop is still undetermined ; the magnitude of the loop four-momentum could be zero or infinite or have any intermediate value. As this particle is virtual, it is necessary to integrate the contribution of the corresponding Feynman diagram over all possible values of p. Unfortunately, it turns out that this integral diverge when $p \to \infty$. This type of divergences is called *ultra-violet* divergences. A way out was found by using the concept of renormalizability ; divergences coming from any order of a series expansion in terms of a given parameter λ can be included into an associated unobservable bare parameter λ_0.

In what follows, we shall review the renormalization mechanism for the expansion of a physics quantity in terms of the strong coupling constant, as illustrated in [18]. This will allow us to introduce the concept of running coupling constant and its application in the context of strong interactions : the *asymptotic freedom*.

Consider a dimensionless physical observable R, which depends on a single energy scale Q. A new energy scale μ_R, the renormalization scale, is introduced at which the subtraction of these divergences is performed. Since R is dimensionless, it can only depends on ratio of the form Q^2/μ_R^2 and α_s. It follows also that since divergences are absorbed by a redefinition of the coupling constant α_s, this so-called *renormalized* coupling constant depends on μ_R. However, μ_R remains an arbitrary parameter, so keeping the non-renormalized coupling constant fixed, R must be independent of the choice of μ_R. this condition can be expressed by :

$$\mu_R^2 \frac{d}{d\mu_R^2} R(Q^2/\mu_R^2, \alpha_s) \equiv \left[\mu_R^2 \frac{\partial}{\partial \mu_R^2} + \mu_R^2 \frac{\partial \alpha_s}{\partial \mu_R^2} \frac{\partial}{\partial \alpha_s} \right] R(Q^2/\mu_R^2, \alpha_s) = 0 \tag{1.34}$$

Defining

$$t = ln \left(\frac{Q^2}{\mu_R^2} \right) \text{ and } \beta(\alpha_s) = \mu_R^2 \frac{\partial \alpha_s}{\partial \mu_R^2} \tag{1.35}$$

one can rewrite Equation (1.34) as the following :

$$\left[-\frac{\partial}{\partial t} + \beta(\alpha_s) \frac{\partial}{\partial \alpha_s} \right] R(e^t, \alpha_s) = 0 \tag{1.36}$$

Now, by introducing the so-called *running coupling constant* $\alpha_s(Q^2)$ defined by the following relation :

$$t = \int_{\alpha_s(\mu_R^2)}^{\alpha_s(Q^2)} \frac{dx}{\beta(x)} \tag{1.37}$$

and differentiating this equation, one obtains the so-called *renormalization-group* equation (RGE) :

$$\frac{\partial}{\partial t} \alpha_s(Q^2) = Q^2 \frac{\partial}{\partial Q^2} \alpha_s(Q^2) = \beta(\alpha_s(Q^2)) \tag{1.38}$$

It implies that $R(1, \alpha_s(Q^2))$ is a solution of Equation (1.36) and therefore that all the scale dependence of R enters through the running coupling constant $\alpha_s(Q^2)$. It turns out that the knowledge of $R(1, \alpha_s(Q^2))$, calculated at fixed-order in perturbation theory, is sufficient as it is possible to predict its evolution with the energy scale Q via Equation (1.37).

In the previous paragraph, we have seen that the independence of any physical observable with respect to the arbitrary cut-off, μ_R, which regulates the inherent divergences, might be retained by introducing an energy scale-dependent coupling constant, whose evolution is controlled by the so-called β function through the renormalization-group equation. In the context of strong interactions, this β function has the following perturbative expansion :

$$\beta(\alpha_s) = -b\alpha_s^2(1 + b'\alpha_s + b''\alpha_s^2 + O(\alpha_s^3)) \tag{1.39}$$

where the function coefficients $\{b, b', b''\}$ are extracted from higher-order loop corrections to the bare vertices. Assuming that at the energy scale Q^2, $\alpha_s(Q^2)$ can be treated perturbatively and by truncating its series expansion at the first order and using the RGE (1.38), one obtains :

$$\begin{aligned} Q^2 \frac{\partial}{\partial Q^2} \alpha_s(Q^2) &= \frac{\partial}{\partial t} \alpha_s(Q^2) \\ &= \beta(\alpha_s(Q^2)) \\ &= -b\alpha_s^2(Q^2) \end{aligned} \tag{1.40}$$

which, after integration, leads to :

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu_R^2)}{1 + \alpha_s(\mu_R^2) bt} \tag{1.41}$$

The value of the running coupling constant decreases as the inverse of $ln(Q^2)$ when Q^2 increases. This dependence in Q^2 is defined by the renormalization-group equation but its absolute value remains unknown and needs to be determined experimentally. The current standard approach in QCD is to define the value of α_s at $\mu = M_Z$, where M_Z is the mass of the Z boson, which is large enough to ensure the validity of the perturbative approach. Historically, this scale was set at the energy where the running coupling constant becomes of the order of unity. At this scale, the series expansion does not converge any more and the perturbative approach breaks down. This scale Λ_{QCD} is defined as the following :

$$ln\left(\frac{Q^2}{\Lambda_{QCD}^2}\right) = -\int_{\alpha_s(Q^2)}^{\infty} \frac{dx}{\beta(x)} \tag{1.42}$$

It is now possible to rewrite Equation (1.41) in terms of Λ_{QCD} as the following :

$$\alpha_s(Q^2) = \frac{1}{b\; ln(Q^2/\Lambda_{QCD}^2)} \tag{1.43}$$

Typically, the value of Λ_{QCD} is of the order 200 MeV. But depends on the order at which the series expansion of α_s is truncated, the number of active flavours and the renormalization scheme [18]. The running behaviour of the strong coupling constant is illustrated in Figure 1.3 ; its value at next-to-leading order in QCD corrections ranges from 0.206 for $Q = 5$ GeV to 0.106 for $Q = m_{top}$, the mass of the top quark. The coupling strength to the gluonic fields asymptotically vanishes at high energy. This property of the strong interactions is known as the *asymptotic freedom* ; quarks do not interact any more and

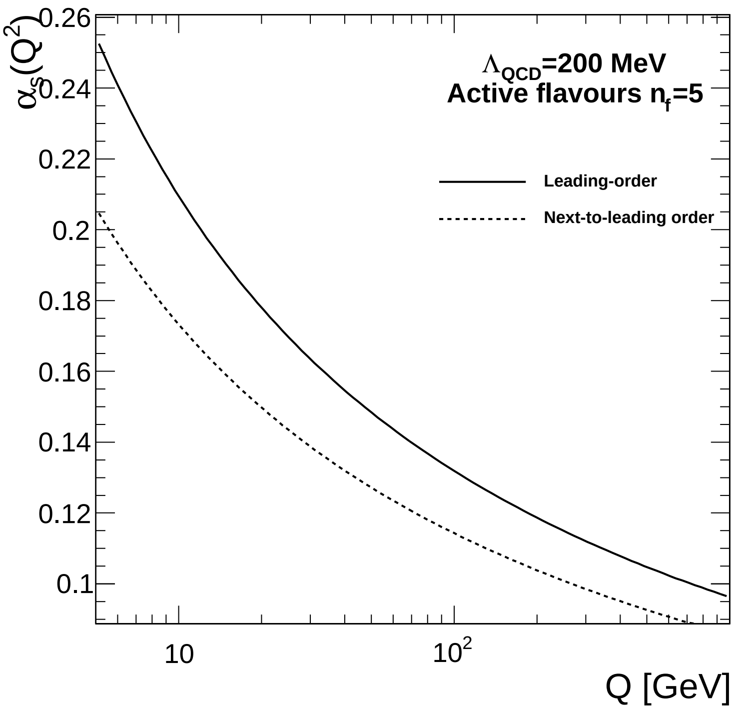


Figure 1.3: Strong coupling constant α_s as function of the energy scale Q

act as free particles. At low energy, however, the coupling strength becomes strong and ensures the so-called *confinement* of quarks into colourless hadrons.

1.5 Physics beyond the Standard Model

A great achievement in high-energy physics is the formulation of the so-called *Standard Model*, which is the current theory of fundamental particles and their interactions. It includes both the electro-weak theory, for which theorists S. Glashow, A. Salam and S. Weinberg were awarded, in 1979, the Nobel Prize in Physics, and the Quantum Chromodynamic which is based on the Quark Model, validated to a large extent in 1974 with the discovery of a fourth quark, the charm quark. In 1976, the Nobel Prize in Physics was awarded to B. Richter and S. C. C. Ting for their contribution to this discovery. Since its original formulation, the Standard Model have been successfully tested by the experiments. In 1983, the detection of the predicted W and Z particles by experimentalists C. Rubbia and S. van der Meer celebrated the advent of the Standard Model as a predictive theory. In 1984, these experimentalists were awarded the Nobel Prize in Physics for this work. Among other successes, the 1988 Nobel Prize in Physics was awarded jointly to L. M. Lederman, M.

Schwartz and J. Steinberger who demonstrated the $SU(2)$ doublet structure of the leptons through the discovery of the muon neutrino. Later, in 1990, J. I. Friedman, H. W. Kendall and R. E. Taylor were awarded the Nobel Prize for their contribution to deep inelastic scattering experiments, which provided evidences for the existence of quarks and in 1995, the Prize was awarded to M. L. Perl for the discovery of the tau lepton and to F. Reines for the detection of neutrinos.

1.5.1 Shortcomings of the Standard Model

Paradoxically, the experimental assessment of the predictive power the Standard Model also led to a certain number of questions indicating that that despite its ability to accurately describe the particles and their interactions at our current energy scale, it cannot be considered as the ultimate theory in particle physics and must be extended. In this section, we shall review some of these shortcomings of the Standard Model.

Gravity : at the current collision energy, the gravitational interaction between particles is negligible. However, this is no longer true as the energy scale of the interactions increases and ultimately, a theory aiming to describe fundamental interactions between particles at any energy scale must include the gravitational interaction. Unfortunately, there is presently no completely satisfactory quantum field theory of gravitation.

Hierarchy and naturalness problem : although the energy scale of the electro-weak symmetry breaking is of the order of $v \sim 246$ GeV, the related Higgs boson mass is a free parameter. However, indirect constraints from experimental observations or related to internal theoretical consistency allow to put limits on this mass. One of them is the requirement of partial wave unitary applied to WW scattering, which ensures that the probability of this process to occur does not exceed one when the collision energy increases. This leads to an upper limit of the Higgs boson mass of 1 TeV$/c^2$. However, the mass of the Higgs boson is very sensitive to radiative corrections such as those corresponding to the one-loop diagrams shown in Figure 1.4. It can be shown [19] that these corrections are proportional to the square of the cut-off scale, Λ_{UV}, used to regulate the ultra-violet divergences of the contribution associated to the loop diagrams. If the Standard Model is believed to describe Nature at any energy scale, the cut-off scale can be chosen as large as the Planck's scale $\Lambda_{Planck} \sim 10^{19}$GeV at which gravitational interactions cannot be neglected anymore. It becomes then necessary to tune of all the constants of the Standard Model to an

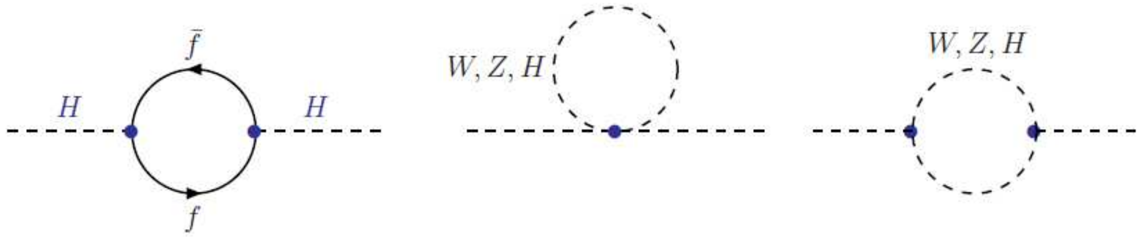


Figure 1.4: Feynman diagrams for one-loop radiative corrections to the SM Higgs boson mass [19].

extremely, and quite unnatural, degree of precision in order to keep the Higgs boson mass below 1 TeV$/c^2$.

Unification of the interactions : the unification of the weak and electro-magnetic interactions into a renormalizable gauge invariant theory, whose predictions were later confirmed by experimental observations, is indeed a remarkable theoretical achievement. But, despite its success, the electro-weak theory is built on a product of two disconnected groups $SU(2) \times U(1)$, involving two unrelated coupling constants. Therefore, it can be considered as a first step toward a more general theory, unifying these two disconnected groups into a larger one : $\mathcal{G} \supset SU(2) \times U(1)$. One step further would require to include the strong interaction as well : $\mathcal{G} \supset SU(3) \times SU(2) \times U(1)$. Theories based on such a gauge group, $\mathcal{G}$, are called *Grand Unification Theory*.

Dark matter : several experimental observations in Cosmology tend to indicate that the gravitational masses of some galaxies, inferred from their rotational motions, are several times larger than their visible mass inferred from the radiations they emit. Among several reasons, reviewed in [20], the concept of *dark matter* has subsequently been introduced to account for these observations, which postulates the existence of neutral weakly interacting particles, for which the Standard Model does not provide any candidate.

Neutrino mass : in the current formulation of the Standard Model, neutrinos are considered as massless. Nevertheless, recent experimental observations tend to show that neutrinos are massive [21, 22, 23, 24, 25] ; flavour transition (called *oscillation*) of neutrinos have a non-zero probability under the assumption of mass and can be detected, for example, as a deficit in the number of observed events involving a particular neutrino flavour.

1.5.2 Extensions of the Standard Model

In order to address the different aforementioned issues, several extensions of the Standard Model have been proposed. In what follows, two examples of such extensions have been highlighted.

Super-Symmetry : a popular extension is a theory which introduces a new symmetry relating fermions and bosons and is called Super-Symmetry (SUSY) (see [26] and references therein). It predicts the existence, for each SM particle, of a super-partner (a sparticle). This theory has the virtue to stabilize the convergence of the calculations of the Higgs boson mass when quantum corrections are added ; provided that these super-partners have a mass lower than few TeV$/c^2$, their corrections to the Higgs boson mass are of opposite sign compared to those from their SM partners. Therefore these corrections cancel each others and quadratic ultra-violet divergences are removed. Moreover, if the so called $\mathcal{R}$-parity, present in some super-symmetric models, is conserved in processes involving super-symmetric particles, the lightest super-symmetric particle (LSP) is always stable and could provide a candidate for the dark matter, if neutral. Finally, the running of the coupling constants of the fundamental interactions is altered by the introduction of the super-partner for each SM particle. This results into an improvement of the unification of these gauge coupling constants at the GUT scale. Super-partners are predicted to have the same mass as their SM partners but as there is no experimental evidence about the existence of such particles, it is most likely that the Super-Symmetry is broken at the energy scale of the Standard Model. Several breaking mechanisms have been proposed and therefore lead to several versions of Super-Symmetry.

Extra-dimensions : another possible way to solve the hierarchy problem consists in introducing additional spatial dimensions. Standard Model gauge and matter fields are assumed to be localized on a (3+1)-dimensional subspace while gravity extends to all dimensions as it is due to the space-time geometry. Therefore, gravity appears to be weak because it is somehow diluted over a larger space than the one accessible to the other fundamental interactions. The Planck's scale, now appearing as an effective scale, must be replaced by the true fundamental scale for gravity whose value is lower, allowing to solve the hierarchy problem. In 1998, N. Arkani-Hamed, S. Dimopoulos and G. Dvali proposed a phenomenological model [27] with large extra-dimensions, compared to the electro-weak symmetry breaking scale. In such model, extra-dimensions can be as large as 1 mm, depending on the number of extra-dimensions. Another model was proposed by L. Randall and R. Sundrum, involving only one additional spatial dimension [28]. But contrary to the previous model, this extra dimension is strongly curved (or *warped*).

2

Physics and simulation of high energy proton collisions

As discussed in Chapter 1, protons are not elementary particles ; they are made of quarks and gluons, whose dynamics are governed by the strong interaction and thus described by the Quantum Chromodynamic theory. In section 2.1, we shall see that a proton-proton collision at high energy can be modelled by treating separately the hard interaction, perturbatively calculable, between constituents, quark or gluon, of each proton and the soft interactions between constituents of a single proton, accounted for by parton distribution functions and obtained experimentally. Processes at work during the subsequent evolution of the interaction, leading to multiple hadrons in the final state, are reviewed in section 2.2. Indeed, accelerated quarks and gluons can radiate additional gluons or quarks as they carry a colour charge. These gluons or quarks, in turn, can radiate more gluons later on and gluons can fluctuate to a quark-antiquark pair. This process is called the parton shower. As newly created gluons and quarks recede progressively from each other, the perturbative treatment of their interactions becomes no longer valid and phenomenological models need to introduced, which account for the transformations of quarks and gluons into observable hadrons. Section 2.3 deals with the colour charged proton remnants from

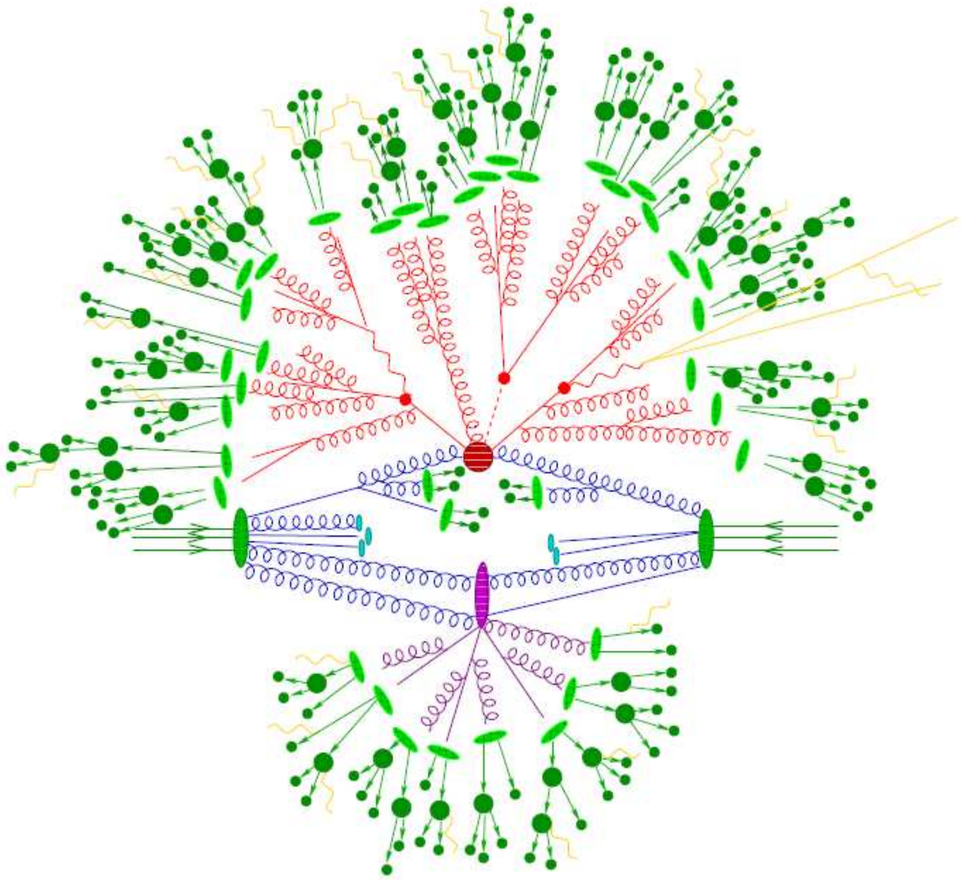


Figure 2.1: Pictorial representation of a proton-proton collisions [29].

the short-distance interactions which radiate also additional gluons and finally hadronize. This leads to the production of additional stable hadrons which represent the so-called *underlying event*. Although the various mechanisms at work during proton collisions are known (Fig. 2.1), a full description of the event remains complex. In addition, some of these mechanisms which occur during the evolution of the initial interaction are not calculable exactly. In practice, the description of the event final state of the interaction is provided by event generators which factorize and simulate the various stages of the interaction. In section 2.4, we shall review the event generators used in this thesis and illustrate the strategy applied to connect the simulation of the short-distance interaction to the simulation of the parton shower.

2.1 Strong interactions in proton collisions

As explained in chapter 1, the strong interaction is described by the Quantum Chromodynamic which accounts for both the weakness of strong interactions at small distances and their strength at short distances. This description is successfully achieved by introducing the concept of *asymptotic freedom* with the help of an energy-dependent coupling strength. This concept leads naturally to factorize proton interactions into two different mechanisms :

- short-distance interactions, also called hard scattering, between one or more partons of each incoming proton, for which the coupling strength is small, allowing a perturbative treatment of the calculations,
- long-distance interactions betweens partons inside each proton, which cannot be calculated analytically and are accounted for by the measured parton distribution functions.

2.1.1 From protons to partons : the QCD factorization theorem

The idea to extent the *Parton Model* formalism to hadron-hadron interactions was first introduced by Drell and Yan in 1971 [30, 31]. In this framework, the total cross section of two hadrons A and B, σ_{AB}, is obtained by multiplying the parton distribution functions, f_A and f_B, of the two incoming hadrons by the the partonic cross section, $\hat{\sigma}_{q_1,q_2}$, which corresponds to the interactions of the bare partons (q_1, q_2) :

$$\sigma_{A,B} = \sum_{q_1,q_2} \iint dx_1 dx_2 f_{q_1/A}(x_1, Q^2) f_{q_2/B}(x_2, Q^2) \hat{\sigma}_{q_1,q_2} \tag{2.1}$$

where $f_{q_1/A}(x_1, Q^2)$ represents the probability density for finding a parton q_1 inside the hadron A with the hadron momentum fraction x_1, when probed with a momentum transfer Q^2. In QCD, problems arise when calculating the quantum corrections to the cross section due to gluon emissions of the partons. These corrections exhibit singularities when the gluon momentum vanishes, called *infrared* singularities or when the gluon is emitted collinear to the parton, called *collinear* singularities. Fortunately, infrared singularities are cancelled when corrections including both real and virtual gluon emissions are added (Lee-Kinoshita-Nauenberg theorem [32, 33]). Thanks to the *factorization theorem* [34], it is possible to factorize out the collinear singularities, order-by-order in QCD corrections, and absorb them into the parton distribution functions. This mechanism is very similar to the renormalization of the ultra-violet divergences, mentioned in section 1.4 and needs the introduction of a so-called *factorization scale*, μ_F. This scale acts as a cut-off which regulates the singularities. Corrections that are not factorized into these renormalized parton distribution functions are now finite and are process-dependent.

2.1.2 Hard process and perturbative QCD

The partonic cross section can be calculated perturbatively as an expansion in the renormalized strong coupling constant, $\alpha_s(\mu_R)$:

$$\hat{\sigma}_{q_1,q_2} = \hat{\sigma}^{(0)}_{q_1,q_2} + \alpha_s(\mu_R^2)\hat{\sigma}^{(1)}_{q_1,q_2} + \alpha_s^2(\mu_R^2)\hat{\sigma}^{(2)}_{q_1,q_2} + \ldots \tag{2.2}$$

The terms $\hat{\sigma}^{(i)}_{q_1,q_2}$ are the coefficients of the strong-coupling constant expansion of the partonic cross-section. An exact calculation of the cross section requires in principle the calculation of the corrections at all orders in α_s. In addition, the invariance of the cross section under changes of the somewhat arbitrary chosen values for the renormalization and factorization scales, μ_R and μ_F, is only achieved when the corrections at all orders are added to the series expansion : the dependence in μ_R^2 and μ_F^2 of the series expansion coefficients $(\sigma^{(i)})$ being exactly compensated by the scale dependence of the parton distribution functions and the coupling constant. In practice, the complexity of such corrections grows so much at higher orders that cross sections are typically calculated at (next-to-)next-to-leading order in QCD corrections. It is therefore necessary to make a choice for the values of μ_R and μ_F. Typical values are chosen equal to the characteristic momentum scale of the hard interaction : for example, $\mu_R^2 = \mu_F^2 = \mu_0^2 = m_V^2, m_{top}^2$ for vector boson V or top quark production. Finally, a theoretical uncertainty is assigned to the cross section which is derived by varying both scales between $\mu_0^2/4$ and $4\mu_0^2$.

2.1.3 Parton distribution functions

Parton distribution functions (PDF's), accounting for the quark and gluon content of the hadron, cannot be obtained analytically. However, once their x-dependence is known for a given value of the momentum transfer Q_0^2, they can be calculated at any scale Q^2 with the

Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution equations [35, 36, 37] :

$$Q^2 \frac{\partial q_i(x,Q^2)}{\partial Q^2} = \frac{\alpha_s}{2\pi} \int_x^1 \frac{dz}{z} \left(P_{qq}(z,\alpha_s) q_i(\frac{x}{z},Q^2) + P_{qg}(z,\alpha_s) g(\frac{x}{z},Q^2) \right) \tag{2.3}$$

$$Q^2 \frac{\partial g(x,Q^2)}{\partial Q^2} = \frac{\alpha_s}{2\pi} \int_x^1 \frac{dz}{z} \left(P_{gq}(z,\alpha_s) \sum_i q_i(\frac{x}{z},Q^2) + P_{gg}(z,\alpha_s) g(\frac{x}{z},Q^2) \right) \tag{2.4}$$

where the functions, $P_{p'p}(z, \alpha_s)$, called either splitting functions or evolution kernels, represent the probability density for a parton of type p to turn into into a parton of type p' with a fraction z of the momentum of the parton p by emitting either a quark or a gluon.

The first equation (2.3) describes the Q^2-evolution of the quark distribution function, q due to gluon radiation (Fig. 2.2a) and quark pair production (Fig. 2.2b). The second equation (2.4) describes the Q^2-evolution of the gluon distribution function, g, due to radiation of gluons by quarks (Fig. 2.2a) or gluons (Fig. 2.2c).

They are thus three distinct processes, leading to three splitting functions, perturbatively calculable, whose expressions at leading order are the following [18] :

- Splitting function for gluon radiation (Fig. 2.2a) :

$$P_{qq}^{LO}(z) = C_F \left[\frac{1+z^2}{(1-z)} \right] \tag{2.5}$$

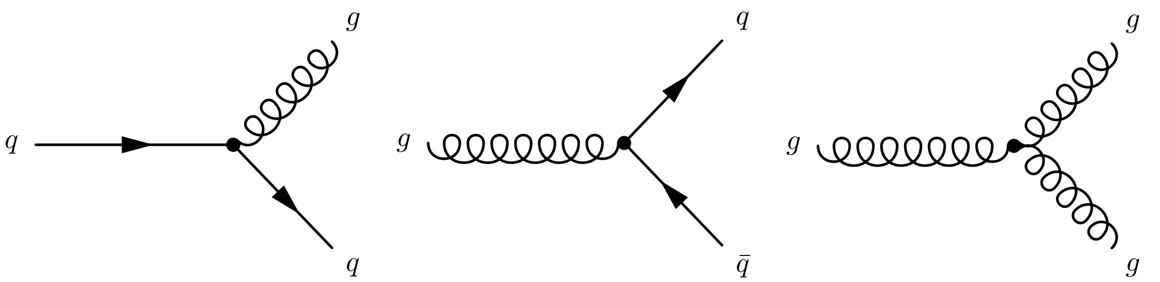
- Splitting function for quark pair production (Fig. 2.2b) :

$$P_{qg}^{LO}(z) = T_R \left(z^2 + (1-z)^2 \right) \tag{2.6}$$

- Splitting function for gluon splitting (Fig. 2.2c) :

$$P_{gg}^{LO}(z) = N_C \left[\frac{z}{(1-z)} + \frac{(1-z)}{z} + z(1-z) \right] \tag{2.7}$$

where N_C is the number of colours ($N_C = 3$), C_F, called the colour factor, is defined as $C_F = (N_C^2 - 1)/(2N_C) = 4/3$ and $T_R = n_f/2$ with n_f the number of possible quark flavours.



(a) **(b)** **(c)**

Figure 2.2: Feynman diagrams for processes generating parton interactions at leading order in QCD corrections : gluon radiation (a), quark pair production (b) and a gluon splitting (c).

The splitting functions can also be estimated at higher orders in QCD corrections using perturbative calculations. For consistency, it is required to calculate both the partonic cross section and the PDF at the same order in α_s.

The DGLAP evolution equations allow to extrapolate, for a given x value, the PDF's from a reference scale Q_0^2 to any desired scale Q^2. However, the x-dependence can not be calculated within the pQCD framework and, therefore, needs to be obtained from data.

As a consequence of the QCD factorization theorem, the PDF's do not depend on the hard scattering process and are thus assumed to be *universal*. In practice, they are derived from data obtained by different experiments. So far, the main data providers are the neutrino-nucleus fixed target experiments [38] and the deep-inelastic scattering experiments performed with the electron-proton collider, HERA [39, 40], at DESY. Tevatron high transverse momentum jet and Z boson rapidity distribution data have been recently incorporated, resulting in better constraints on the gluon content of protons at high x values [41, 42]. Finally, recent results for the LHC experiments have also been shown to be able to provide new constraints on PDF's [43].

Over the last decades, mainly two groups, the **M**artin-**S**tirling-**T**horne-**W**att (MSTW) [44, 41] (formerly **M**artin-**R**oberts-**S**tirling-**T**horne, (MRST) [45]) and the **C**oordinated **T**heoretical-**E**xperimental Project on **Q**CD (CTEQ) [46, 47, 48] collaborations, provided the high-energy physics community with regular updates on these PDF's as more and more data were collected. The set of PDF's of these two groups differ by technical details and their choice of x-dependence parametrization but basically, the methodology is the following :

- the PDF's are parametrized by functions of the form $f(x) \propto Ax^{\alpha}(1-x)^{\beta}$

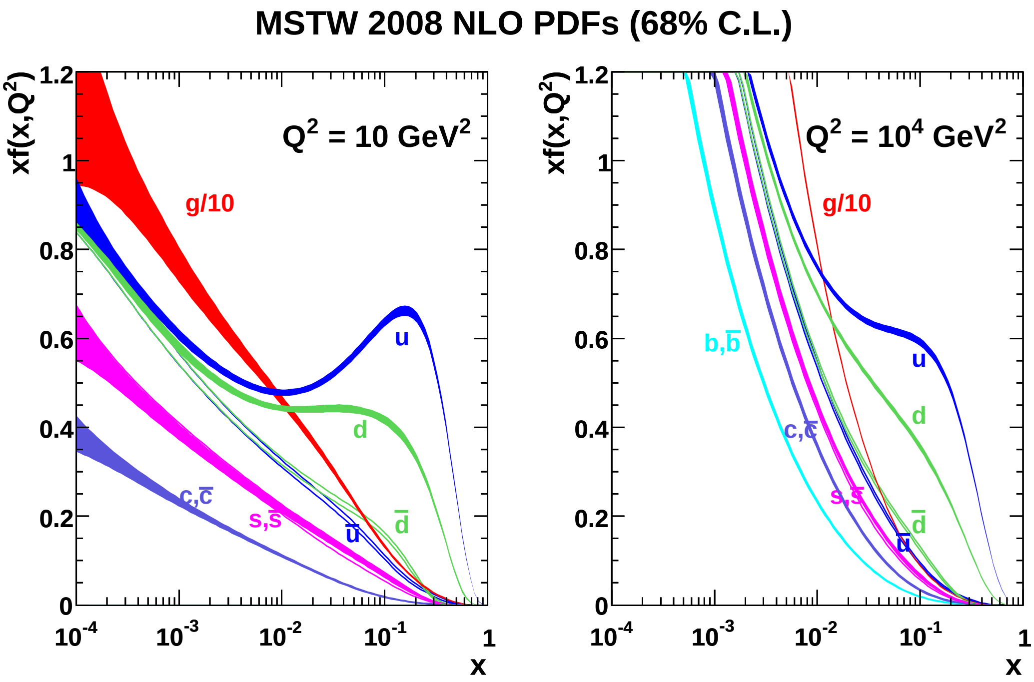


Figure 2.3: Next-to-leading order parton distribution functions and associated uncertainties as a function of the parton momentum fraction, x, obtained by the MSTW collaboration for two different momentum transfer Q^2 [44].

- the values of the constants A, α and β are derived from a global fit χ^2_{global} to all the data available

As an illustration, Figure 2.3 shows the next-to-leading order quark and gluon distribution functions for a proton, obtained by the MSTW collaboration. Additional constraints on the global fit may be derived from QCD sum rules [49], ensuring the conservation law for the constituents of the considered hadrons. More recently, new groups have joined this effort ; for instance, the NNPDF collaboration [50] whose parametrization approach is somewhat different. In order to avoid possible bias to a particular choice of functional form, this group chose a neural network-based parametrization.

Experimental uncertainties [51] accounts for experimental uncertainties on the data points which are fitted to extract the values of the different PDF's parameters. Attention needs to be paid in order to propagate correctly errors from various datasets. Theoretical uncertainties [52] includes, for instance, uncertainties from neglected higher-order contributions to the DGLAP evolution equations, QED corrections, choices of the functional forms, treatments of the heavy quark contributions, . . . These theoretical uncertainties are often difficult to quantify and are often deduced *a posteriori* when improved calculations are available. From the experimental point of view, what truly matters is to be able to assess correctly the impact of the PDF's uncertainties on the total cross section or on any desired observable. As an example, here is given a brief review of one of the most popular method used to quantify the uncertainties of PDF's : the *Hessian* method [53]. The idea is to propagate linearly the experimental uncertainties on the fitted data to any observable F via a Hessian matrix :

$$\Delta F = T\sqrt{\sum_i \sum_j \frac{\partial F}{\partial a_i} H^{-1}_{i,j} \frac{\partial F}{\partial a_j}} \tag{2.8}$$

where T is the so-called *tolerance* factor and the components, $H_{i,j}$, of the Hessian matrix are defined as :

$$H_{i,j} = \frac{1}{2}\frac{\partial^2 \chi^2_{global}}{\partial a_i \partial a_j}|_{min} \tag{2.9}$$

with $\{a_k\}_{k\in[1,n]}$, the PDF's parameters and χ^2_{global}, the global goodness-of-fit quantity, defined previously in this section, which is at its minimum when $\forall i \in [1,n]\ a_i = a_i^0$.

In practice, the error matrix, calculated by inverting the Hessian matrix, is diagonalized in order to obtain its eigenvectors. Then, the uncertainty on the observable F is calculated by summing quadratically the variations on F obtained by varying up and down the PDF's parameters along the direction of each eigenvector.

2.2 Towards a high-multiplicity hadronic final-state

2.2.1 Parton showering

Gluon radiations, called branching, may occur before or after the hard scattering at energies where the perturbative approach is valid. In principle, an exact computation of higher-order QCD corrections to the partonic cross sections, due to the emissions of both real and virtual particles, is required for an exact series expansions in the strong coupling constant, α_s. In practice, the complexity of such corrections grows exponentially with the parton multiplicity and corresponding calculations fall quickly beyond what is currently feasible. Unfortunately, there are some phase-space regions where these corrections are enhanced and can not be neglected. Therefore, instead of aiming for an exact calculation of both real and virtual corrections up to some fixed order, an alternative solution consists in making use of the parton splitting functions, introduced in section 2.1.3, to account for real emission corrections in phase-space regions where these corrections are enhanced. An arbitrary number of successive branching of one parton into two or more are applied, as long as the perturbative approach remains valid, generating more and more partons at each step and therefore accounting partially for corrections of higher and higher orders. This procedure allows to factorize a complex $2 \to n$ process with $n \gg 1$ into a calculable low multiplicity process which will afterwards be convoluted with a parton shower. It can be summarized as follow :

1. calculate the $2 \to 2$ process using Feynman rules ($gg \to t\bar{t}$ for instance)
2. make use of the parton splitting functions, for the branching of a gluon into two gluons ($g \to gg$) or a quark into a quark and a gluon ($q \to qg$) for instance, to account for the higher-order real corrections corresponding to the $2 \to 3$ process ($gg \to t\bar{t}g$ for instance)
3. reiterate the step 2 as long as the perturbative regime remains valid.

In order to study the evolution of the parton down to energy scales for which the perturbative approach breaks (Λ_{QCD}), it is convenient to introduce an evolution variable, t, similar to the one defined in section 1.4. This variable is defined as following :

$$t = ln(\frac{Q^2}{\Lambda_{QCD}^2}) \Rightarrow dt = dln(Q^2) = \frac{dQ^2}{Q^2} \tag{2.10}$$

with Q^2, the virtuality of the branching parton defined as $|m^2|$.

At each iteration of the parton shower, the differential probability, $d\mathcal{P}$, for a parton a to branch into two partons, b and c, with b taking a momentum fraction z is given by the corresponding splitting function introduced in section 2.1.3 :

$$d\mathcal{P}_a = \sum_b \frac{\alpha_s}{2\pi} P_{ba}(z) dt dz \tag{2.11}$$

Here the sum runs over all allowed splitting for b.

At a given n-th interaction, in the approximation of collinear emissions, the relation between the matrix-element with n partons, $\mathcal{M}_n$ and the matrix-element with n+1 partons, accounting for a higher order real correction in QCD is given by :

$$|\mathcal{M}_{n+1}|^2 d\Phi_{n+1} = |\mathcal{M}_n|^2 d\Phi_n dt \sum_b \frac{\alpha_s}{2\pi} P_{ba}(z) dz \tag{2.12}$$

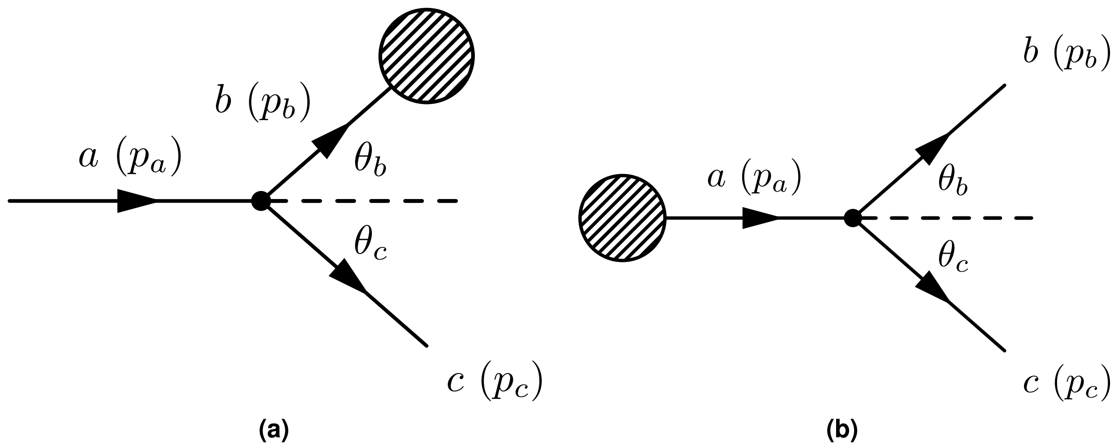


Figure 2.4: Space-like (a) and time-like (b) parton branching. The parton a represents the branching parton while the parton b represents the parton after branching. The parton c is the radiation. The shaded blob represents to the rest of the diagram, apart from the branching line, where the hard interaction of the process occurs.

where Φ_n is the n-body phase-space element.

In general, real emissions are divided into two categories : initial state radiations (Fig. 2.4a), which correspond to real emissions radiated the incoming colliding partons and final state radiations (Fig. 2.4b), which correspond to real emissions radiated by the outgoing partons. At a given value of t during the shower evolution, the probability for a specific branching $a \to bc$ to occur is defined as the integral of the corresponding splitting function over the allowed z values $[z_-, z_+]$:

$$\mathcal{I}_{a\to bc}(t) = \int_{z_-}^{z_+} dz \frac{\alpha_s}{2\pi} P_{ba}(z) \tag{2.13}$$

and the probability that a branching occur during a small range of t values, δt, is given by :

$$\mathcal{P}_{branching}(t, t+\delta t) = \sum_b \mathcal{I}_{a\to bc}(t)\delta t \tag{2.14}$$

and thus, the probability for no branching is :

$$\mathcal{P}_{no\ branching}(t, t+\delta t) = 1 - \mathcal{P}_{branching}(t, t+\delta t) \tag{2.15}$$

The probability that a parton has not yet branch when evolving from any arbitrary value t_0 to a given value t, with $t_0 < t$ is the product of the probability that it did not branch in any of the sub-intervals obtained by dividing $[t_0, t]$. The probability for no branching is multiplicative in the evolution variable t :

$$\mathcal{P}_{no\ branching}(t_0, t) = \mathcal{P}_{no\ branching}(t_0, t_i)\mathcal{P}_{no\ branching}(t_i, t) \tag{2.16}$$

Now, by subdividing further the interval $[t_0, t]$ with $t_i = t_0 + (i/n)(t - t_0)$, $i \in [0, n]$:

$$\begin{aligned}
\mathcal{P}_{no\ branching}(t_0, t) &= \lim_{n\to\infty} \prod_{i=0}^{n-1} \mathcal{P}_{no\ branching}(t_i, t_{i+1}) \\
&= \lim_{n\to\infty} \prod_{i=0}^{n-1} (1 - \mathcal{P}_{branching}(t_i, t_{i+1})) \\
&= exp\left(-\lim_{n\to\infty} \sum_{i=0}^{n-1} \mathcal{P}_{branching}(t_i, t_{i+1})\right) \\
&= exp\left(-\int_{t_0}^{t} \frac{d\mathcal{P}_{branching}(t')}{dt'} dt'\right) \\
&= exp\left(-\int_{t_0}^{t} \sum_{b,c} \mathcal{I}_{a\to bc}(t')dt'\right) = S_a(t)
\end{aligned} \tag{2.17}$$

The term S_a is often referred to as the *Sudakov form factor* [54]. Finally, the probability that the parton branching occurs at t is given by :

$$\frac{d\mathcal{P}_a}{dt} = -\frac{d\mathcal{P}_{no\ branching}(t_0, t)}{dt} = \left(\sum_{b,c} \mathcal{I}_{a\to bc}(t)\right) exp\left(-\int_{t_0}^{t} \sum_{b,c} \mathcal{I}_{a\to bc}(t')dt'\right) \tag{2.18}$$

This equation can be seen as a classic branching probability which is suppressed due to the Sudakov form factor in order to ensure the conservation of the total probability ; if a parton has already branched at $t' < t$, it can no longer branch at t.

2.2.2 Phenomenological hadronization models

The parton shower leads to a final state of multiple coloured partons. As explained before, due to colour confinement, these partons are not observable and evolve into colourless hadrons. When the evolution variable Q^2 reaches some infra-red cut-off value Q_0^2, the perturbative treatment is no longer valid ; as the value of the strong coupling constant increases when the energy scale decreases, at some point, the series expansion in α_s does not converge any more. Therefore, the parton shower approach has to be interrupted and is taken over by non-perturbative phenomenological models describing the fragmentation of

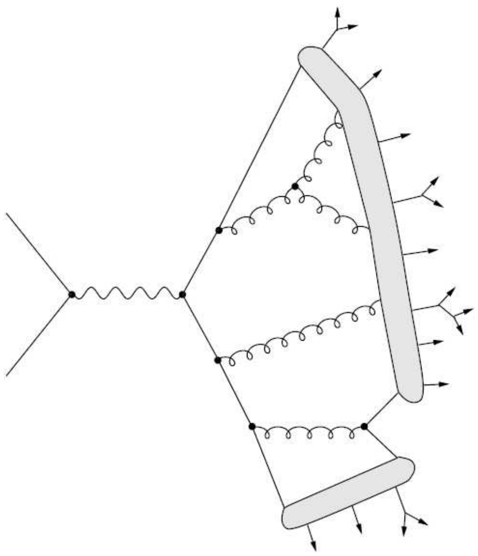


Figure 2.5: Schematic representation of the string hadronization model.

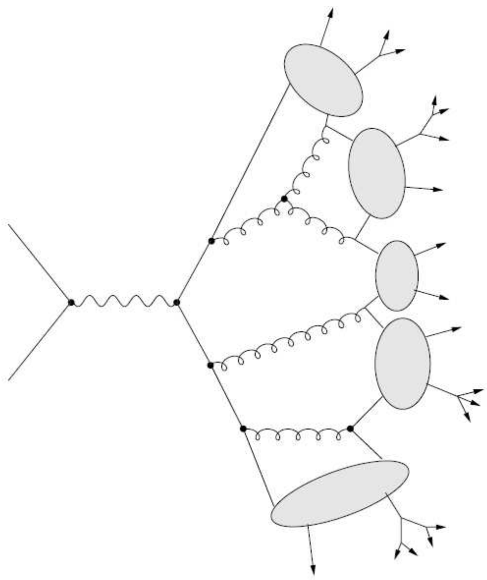


Figure 2.6: Schematic representation of the cluster hadronization model.

partons and confinement of these fragments into colourless hadrons ; this is the so-called hadronization process. In the present section are described the two main hadronization models : the *Lund* string model [55] and the low-mass cluster model [56].

The *Lund* string model (Fig. 2.5) : this model is based on the intuitive formulation of the quark confinement ; pairs of quark-antiquark ($q\bar{q}$) act as colour dipoles and the energy stored in the colour field between them increases as the quarks move apart from their common production vertex. This colour field is modelled by a one-dimensional string, stretched between the two quarks. The string tension, κ is assumed to be constant, corresponding to a linear quark confining potential. From hadron spectroscopy, the value is deduced to be $\kappa \simeq 1\ GeV/fm$. The string breaks up iteratively through the creation of new $q\bar{q}$ pairs. It is also assumed that the initial string has no transverse momentum, the transverse momentum of a quark being compensated by the antiquark transverse momentum. The last assumption concerns the possibility to simulate independently the mass, m, and the transverse momentum, p_T, of the created quark pairs, following a Gaussian probability distribution :

$$exp(\frac{-\pi m^2}{\kappa})exp(\frac{-\pi p_T^2}{\kappa}) \tag{2.19}$$

This formulation introduces the experimentally observed suppression of heavy quark production during the hadronization. Later on, the hadron transverse momentum is simply taken as the sum of the transverse momenta of its constituent quarks. The last step consists in simulating the remaining longitudinal momentum, p_Z, of the hadron and constrains at the same time its energy E :

$$E^2 - p_z^2 = (E - p_z)(E + p_z) = m^2 + p_T^2 = m_T^2 \tag{2.20}$$

Denoting by z, the fraction of $E + p_z$ taken by the hadron out the available $E + p_z$, the z-probability density function is given by the so-called *Lund symmetric fragmentation function*, f :

$$f(z) \ \propto z^{-1}(1 - z)^a exp(-b\frac{m_T}{z}) \tag{2.21}$$

where a and b are the *Lund a* and *Lund b* parameters. They cannot be calculated and need to be deduced from experiments.

The low-mass cluster model (Fig. 2.6) : the basic idea is to rely on the colour pre-confinement picture of the parton shower which states that partons produced during this perturbative evolution becomes organized in colour-singlet clusters with finite masses of the order of the infra-red cut-off scale which separates the perturbative and non-perturbative regime of the shower evolution. In the low-mass model, this cut-off value is chosen to be of the order of $1\ GeV$. Then, the clusters decay, if massive enough, isotropically into hadrons. The flavours of the decay products are chosen randomly ; for a cluster of flavour $f_1\bar{f}_2$, a flavour f is picked up at random among u,d,s (and the 6 possible di-quark flavour combinations : $u\bar{u}$, $u\bar{d}$, $u\bar{s}$, $d\bar{d}$, $d\bar{s}$ and $s\bar{s}$) and c. In this way, the flavour of the decay products is fully specified : $f_1\bar{f}$ and $f\bar{f}_2$. The final hadrons are selected again randomly from the list of all hadrons matching these flavour combinations.

2.2.3 Heavy quarks production and fragmentation

In both hadronization model described in section 2.2.2, heavy quark production is highly suppressed during the non-perturbative evolution. Nevertheless, heavy quarks can be produced during the parton shower via gluon splitting $g \rightarrow Q\bar{Q}$. And their fragmentation during the hadronization needs to be taken into account. A priori, nothing prevents us to use the same hadronization model for heavy and light quarks. But experimental data have shown that, in the case of the *Lund string model* for instance, the fragmentation function does not model accurately hadronization processes involving heavy quarks. A

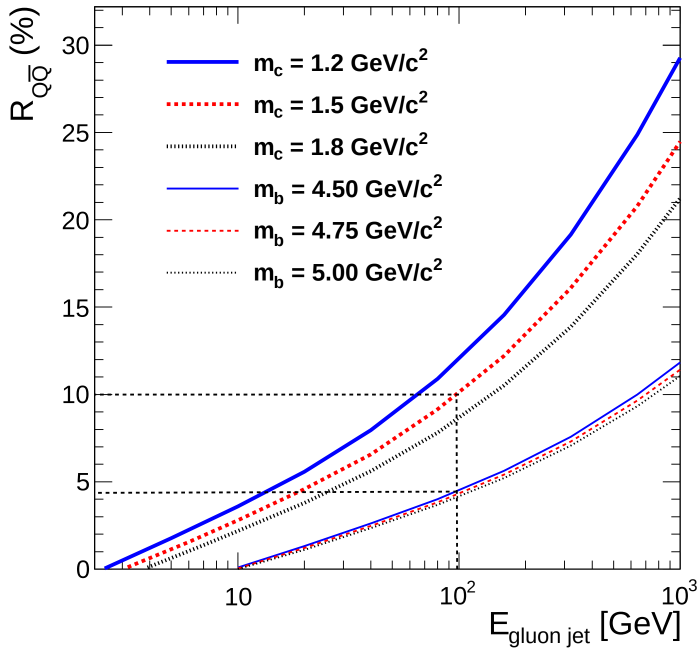


Figure 2.7: Probability to produce a heavy quark pair per jet as a function of the jet energy.

more satisfactory solution was provided by Peterson et al. in 1983 [57] :

$$f(z) \propto z^{-1}\left(1-\frac{1}{z}-\frac{\epsilon_q}{1-z}\right)^{-2} \tag{2.22}$$

where ϵ_q is a free parameter which is expected to be proportional to $1/m_Q^2$.

In order to produce a pair of heavy quarks $Q\bar{Q}$, the incoming gluon must have a virtuality $Q^2 > 4m_Q^2$. Therefore, for b quarks but also for c quarks, $Q^2 >> \Lambda_{QCD}^2$. Heavy quark production occurs then during the perturbatively calculable part of the shower evolution. It is interesting to determine the probability to produce a $Q\bar{Q}$ pairs per jet induced by a gluon of energy E, $R_{Q\bar{Q}(E)}$. This calculation can be achieved by considering the number of gluon with a given virtuality Q^2 per jet of energy E, $n_g(Q^2, E)$ and by using the gluon splitting function, P_{qg} defined in (2.6), generalized for heavy quarks [58, 59, 60] :

$$R_{Q\bar{Q}}(E) = \int_{4m_Q^2}^{E^2} \frac{dQ^2}{Q^2}\frac{\alpha_s(Q^2)}{2\pi} n_g(Q^2, E) \int_{z_-}^{z+} P_{qg}(z)dz \tag{2.23}$$

where the kinematically limited integration interval for the longitudinal momentum fraction z of the heavy quark is delimited by $z_\pm = (1 \pm \beta)/2$ with $\beta = \sqrt{1-(4m_Q^2/Q^2)}$. The probability to produce a $Q\bar{Q}$ pair increases as the gluon energy increases (Fig. 2.7) ; for a gluon with an energy of 30 GeV, the probability to produce a pair of c quarks with a mass $m_c = (1.5 \pm 0.3)$ GeV and b quarks with a mass $m_b = (4.75 \pm 0.25)$ GeV are $\sim 5 \pm 1$ % and 2.0 ± 0.1 % respectively. These probabilities increase up to 10 ± 2 % and to 4.3 ± 0.2 % for a gluon with an energy of 100 GeV. While the influence of the c quark mass uncertainty on this probability increases with the jet energy, the influence of the b quark mass uncertainty remains almost negligible up to an jet energy of 1000 GeV.

2.3 Multiple interactions and underlying events

In the previous sections, proton interactions have been modelled as the interaction between their constituents. But, so far, a unique interaction per proton interaction has been considered. However, it is perfectly possible that multiple partons from the same proton undergo separately an interaction ; this effect is often referred to as a multiple interaction (MI) and has to be taken into account to complete the picture of proton-proton interactions. Among the few models available to account for MI, the one used by the event generators used in this thesis has been proposed by T. Sjöstrand [61] ; it assumes that the pairwise interactions between partons are independent. The number of interactions follows therefore a Poisson distribution.Furthermore, as protons are extended objects, multiple interactions are more likely to happen for central collisions than for peripheral ones ; thus the number of interactions depends on the distance separating the centres of the colliding protons, measured in the transverse plane with respect to their direction. The mean number of interactions is finally obtained as the ratio of the $2 \rightarrow 2$ process cross section, which is supposed to include all the pairwise parton interactions, to the expected total proton-proton cross-section. As the $2 \rightarrow 2$ process cross-section diverges as the transverse momentum of the outgoing partons vanishes, a lower transverse momentum cut-off must be introduced as a free parameter. Nevertheless, for small value of transverse momentum, the aforementioned ratio may exceeds unity, accounting for multiple parton interactions.

Since the proton is a colourless object, the proton remnants, which do not participate to the hard interaction, carry now also a colour charge. Therefore, they can fragment and hadronize through the mechanisms described in the previous section, producing additional particles with a low transverse momentum with respect to the hard interaction. The initial transverse momentum of these remnants is such as to compensate for the transverse momentum of the partons participating to the hard interaction. This so-called primordial transverse momentum is usually assumed to be distributed according to a Gaussian distribution with a width value of the order of 1 GeV.

The set of particles arising both from multiple interactions and from the proton remnants represents the so-called *underlying event*. It is worth noticing that the underlying event cannot be unambiguously separated from initial and final state radiations.

2.4 Event simulations

From the hard scattering to the production of jets of colourless hadrons, it is impossible to calculate analytically all the processes occurring during proton collisions. Event generators are then the only way to access a detailed description of the final-state and compute any desired experimental observable. There are dedicated tools to compute the full kinematic properties of the hard scattering and the proton remnants. They are usually interfaced with programs taking care of the parton shower and of the hadronization. Then, interactions between the particles and the detector are simulated, as well as the detector response. Finally, the simulated events are usually processed within the same reconstruction framework as the one used for real events. In this way, simulated data can be used as if they were real data. As a matter of fact, experimentalists rely heavily on event generators and for us, there is a true interplay between real and simulated data :

- simulated data are needed in addition to real data to optimize the design of a detector with respect to the physics process of interest (signal), to study the observability of a given signal among other processes faking the same required final-state (background), to perform corrections to analysis strategies and estimate possible measurement biases, to test physics objects reconstruction strategies,...
- real data are helpful to tune the free parameters of the physics models used by the different event generators and therefore obtain a more accurate event description and by extension increase the predictive power of these event generators.

Most event generators are based on so-called Monte-Carlo techniques and are therefore referred to as Monte-Carlo generators [62]. Contrary to deterministic algorithms, Monte-Carlo methods rely on probabilistic algorithms. This approach is particularly well suited for calculations in quantum mechanics where only probabilities of the possible outcomes of an experiment is possible.

2.4.1 Monte-Carlo event generators

Different Monte-Carlo generators are available and a brief review of those used in this thesis is given below :

- Full event generators : these generators provide a full event description, including the hard interaction generation, the parton shower and hadronization.

  PYTHIA [63] : this generator is one of the most widely used general purpose generator for hadronic events in pp, e^+e^- and ep collisions. It contains tools to simulate hard-interactions as well as initial- and final-state radiations and hadronization and decay processes. It based on the Lund string model fro hadronization.

  SHERPA [29, 64] : this full event generator is based on a multi-purpose parton-level generator, AMEGIC++ [65], interfaced to a parton shower package, using the CKKW matching scheme, documented in the next section. The model used for hadronization, is based on the cluster model.

- Matrix-element generators : these generators provide, in an automated fashion, the calculations of the matrix elements for specific physics processes. They are meant to be passed to full event generator for the showering and hadronization.

  ALPGEN [66] : this generator calculates, at leading order in QCD and electro-weak corrections, the exact matrix-elements for a large set of pre-defined Standard Model processes in hadron collisions with emphasis on large parton multiplicity final state, including heavy quarks. A subsequent evolution of the final state via parton shower is possible by connecting the ALPGEN generator to a general purpose generator as PYTHIA , using the MLM matching scheme, documented in the next section.

  MADGRAPH /MADEVENT [67] : MADEVENT is a general purpose event generator relying on the matrix-element generator MADGRAPH . For any process specified by the user, MADGRAPH lists all the related Feynman diagrams at leading order in QCD and electro-weak interactions and calculates their amplitudes. An interface to PYTHIA for a parton shower evolution is available, using the MLM matching scheme. A web interface as well as the possibility to use or/and implement physics models beyond the Standard Model are also available.

Other generators, like HERWIG , MC@NLO , POWHEG , . . . are also common event generators but are not documented here as they are not used in this thesis.

2.4.2 Making the link between hard process and parton shower

A crucial aspect of event generation is the way to connect the different simulation steps. The definition of these steps is somewhat artificial and the resulting ambiguity leads to potential overlaps between the outputs of the different simulation steps. Indeed, from the formalism point of view, it is impossible to unambiguously separate the production of partons via the hard interaction, for which the exact matrix element calculation is possible, from the production of partons during the the parton shower. As a matter of fact, assuming that a hard parton leads to an observable jet in the final state, an event with n jets can be generated in two different ways :

Path 1 : from an event with n hard partons generated at the matrix-element level, undergoing a parton shower involving only soft and/or collinear radiations

Path 2 : from an event with only $n-1$ hard partons generated at the matrix-element level, emitting a hard radiation during the parton shower.

Nevertheless, such an ambiguity can be avoided during the Monte-Carlo generation by the use of a so-called *matching scheme*, also often called *merging scheme*. The goals of this scheme are the following :

- to ensure that a given event does not appear twice in the final set of generated events, generated by each of the two aforementioned paths,
- to ensure that each possible configuration of a n-jet event has been generated at least once, following one of the two paths and therefore that there is no uncovered phase-space region.

Basically, two different schemes are commonly used for matching leading-order matrix-element calculations and the parton shower development : the **M**ichelangelo **L**. **M**angano scheme (MLM) [68] and the **C**atani-**K**rauss-**K**uhn-**W**ebber scheme (CKKW) [69].

The MLM scheme : each event after parton shower is accepted only if each parton, generated at matrix-element level, leads exactly to one jet of partons after the parton shower evolution. This procedure consists in the following steps :

- Generate all possible parton configurations (called *event*) for all parton multiplicities n up to N with the chosen Monte-Carlo matrix-element generator. Partons are required to have a transverse momentum higher than a certain threshold, $p_T^{min.}$ and to be separated from each other by a minimal distance ΔR in the $\eta - \phi$

plane. Events are then sorted according to their parton multiplicity, creating N different exclusive samples.

- Evolve each exclusive sample via a parton shower program.
- Run a jet reconstruction algorithm[1] on each event, taking as input the list of partons produced after the parton showering. Each reconstructed jet is called a *cluster*. Clusters with a transverse momentum lower than a certain threshold are discarded.
- Try to associate the partons generated at the matrix-element level to the clusters : each parton, starting from the one with the highest transverse momentum, is associated with its closest cluster in the (η, ϕ)-space. If the distance between the parton and this cluster is smaller that a certain fixed parameter, R_{match}, the parton is flagged as *matched*. Then the cluster is removed from the list of clusters and the procedure is repeated with the next parton until all partons have been processed.
- Events with unmatched partons are discarded as each parton generated at matrix-element level is expected to lead to a cluster.
- If $n < N$, events with unmatched cluster are discarded as clusters have to arise exclusively from the showering of the parton generated at matrix-element level.
- If $n = N$, events are only accepted if all the unmatched clusters have a lower transverse momentum than the matched clusters. That is, no emission with a high transverse momentum has been produced during the shower evolution.

After this procedure, one is left with $N-1$ samples, each sample having an exact number of jets (from 1 to $N-1$) and an sample with N or more jets. These samples have then to be merged in order to obtain the fully inclusive description of the event.

The CKKW scheme : the idea is to divide the phase-space region of partonic emissions into a region of jet production based on matrix-element calculations and a region of jet evolution, described by the parton shower approach. This division is done using a variable, Q^2, which measure the distance between two jets (i, j) and consider them as separated at a given energy scale Q_0 if :

$$Q_{i,j}^2 = 2\, min\{m_T^i, m_T^j\}^2\, \frac{cosh(y^i - y^j) - cos(\phi^i - \phi^j)}{D^2} \geq Q_0^2 \tag{2.24}$$

[1] Jet reconstruction algorithms are documented in section 4.3.2 ; basically, these algorithms try to recover the kinematic properties of these underlying partons by clustering particles that are likely to come from the same parton into a single object. These particles could either be partons produced after the parton shower or hadrons produced after the hadronization.

where $m_T = \sqrt{m^2 + p_T^2}$ is the transverse mass of the jet, y, its rapidity and ϕ its azimuthal angle. D is an arbitrary parameter of the k_T-jet algorithm.

The CKKW procedure is the following :

- define a scale Q_{cut} to separate the two phase-space regions mentioned previously,
- rely on matrix-element calculations to generate n-parton momentum configurations, for $n \in [2, N]$, satisfying : $Q^2_{i,j} > Q^2_{cut}$ and $p^i_T > Q_{cut}$,
- reweight the obtained parton momentum configuration with Sudakov form factors, accounting for the probability of having no further radiation resolvable at Q_{cut} during the evolution from Q_0 down to Q_{cut} for external parton lines and down to q for internal parton lines, q being the nodal k_T value $Q_{i,j}$, defined for each parton emission by a backward clustering of the parton configuration with the k_T-jet algorithm. This clustering is performed iteratively until a $2 \to 2$ (di-jet) configuration is obtained, producing a pseudo shower history, whose evolution will be properly continued by the parton shower. Basically, for a n-parton configuration, these nodal values define the scale at which $2, 3, \ldots, n$ jets are resolved,
- reweight also the matrix-element parton configuration, to account for the running of the strong coupling constant, with a factor $\frac{\prod_i^n \alpha_s(q_i)}{\alpha_{s,0}^{n-2}}$, $\alpha_{s,0}$ being the value used to generate the initial parton configuration,
- apply a parton shower where any emission leading to a nodal value $Q_{i,j}$ larger than Q_{cut} is vetoed.

3

The CMS experiment at the Large Hadron Collider

In 1994, the Council of the European Organization for Nuclear Research (CERN) approved the construction of what is today the world largest particle accelerator, the *Large Hadron Collider* (LHC) [70]. Straddling the French-Swiss border, in the outskirts of Geneva, this circular accelerator has a circumference of 27 kilometres and is located on average 100 metres underground in the tunnel of the former *Large Electron-Positron* collider (LEP) which ran from 1989 to 2000. In October 1995, the LHC technical design report [71] was published, establishing the architecture of the future accelerator. In February 1996, two out the six future experiments carried out at the LHC were approved : the *Compact Muon Solenoid* (CMS) [72] and the *A Toroidal LHC ApparatuS* (ATLAS) [73]. Their physics programmes include the search for the Higgs boson as well as searches for evidences of new physics like a fourth quark generation, super-symmetric particles or extra dimensions. They also include searches for a particle candidate to the so-called dark matter which is believed to compose the major part of the matter in the Universe. Two other experiments, the *Large Hadron Collider beauty* (LHCb) [74] and *A Large Ion Collider Experiment* (ALICE) [75] were approved in February 1997 and September 1998. The main purpose of the LHCb experiment is to study the mechanism leading to the asymmetry between matter and anti-matter in the Universe in processes involving production of beauty quarks. The ALICE physics programme concerns the study of a state of matter known as quark-gluon plasma which is likely to be produced during heavy ion collisions. The two remaining smaller experiments are the *TOTal Elastic and diffractive cross-section Measurement* experiment (TOTEM) [76] and the *Large Hadron Collider Forward* experiment (LHCf) [77]. The TOTEM experiment aims, for example, to measure the proton size and will provide a precise monitoring of the LHC luminosity while the LHCf experiment uses particles produced during the LHC collisions to simulate how cosmic particles interact with the nuclei in the earth's atmosphere and produce particle cascades at ground level. This experiment aims to provide new informations for the calibration of large-scale cosmic-ray experiments.

3.1 The European Organisation for Nuclear Physics

The world's largest organization for particle physics just turned 50 years old in 2004. In what follows are reviewed the milestones of its foundation :

- December 1949 : the French Nobel laureate physicist, Louis de Bröglie made the first official proposal for the creation of a European laboratory at the European Culture Conference in Lausanne.
- June 1950 : the *United Nations Educational, Scientific and Cultural Organization* (UNESCO) is authorized to *assist and encourage the formation of regional research laboratories in order to increase international scientific cooperation*, during the fifth UNESCO General Conference in Florence. This decision followed a resolution of the American Nobel laureate physicist Isidor Rabi.
- December 1951 : at the intergovernmental meeting of the UNESCO in Paris, the first resolution for the establishment of a European Council for Nuclear Research was adopted. Two months later, an agreement was signed by eleven countries establishing the provisional Council. The acronym CERN (Conseil européen pour la Recherche Nucléaire) was born.
- October 1952 : Geneva was chosen as the site of the future laboratory during the third session of the provisional Council.
- July 1953 : the CERN convention was established and then progressively ratified by the twelve founding member States (Belgium, Denmark, France, the Federal Republic

of Germany, Greece, Italy, the Netherlands, Norway, Sweden, Switzerland, the United Kingdom and Yugoslavia).

- 29^{th} September 1954 : following a ratification by France and Germany, the provisional Council was dissolved and replaced by the European Organization for Nuclear Research. The acronym CERN remained though.

Nowadays, a total of twenty European countries are counted among the CERN Member States. Furthermore, CERN physics programmes involve thirty-five other Non-Member States. Consequently, besides its 2500 employees, scientific and technical staffs, CERN hosts 8000 visiting scientists from 580 different institutes, representing eighty-five different nationalities.

Over the last five decades, by bringing together these physicists from all around the world towards a common goal, experiments carried out at CERN have increased our understanding of the laws of Universe to an unprecedented state. Nevertheless, as the knowledge grows, grows equally, if not more, the conscious of what remains to be understood. With the new major CERN particle accelerator, the LHC, physicists hope to make a step further in their endless quest to reveal Nature's most intimate secrets.

3.2 The Large Hadron Collider at CERN

The Large Hadron Collider has been initially designed to collide two counter rotating beams of protons or heavy ions (Pb), at a centre-of-mass energy of 14 and 1150 teraelectronvolts (TeV) respectively. Using the existing CERN *accelerator complex* (Fig. 3.1), protons are first produced by removing electrons from hydrogen atoms. They are then injected from a linear accelerator (LINAC 2) into the PS booster and then into the Proton Synchrotron (PS). At that stage,the proton beam has an energy of 25 GeV. Finally, protons are accelerated up to an energy of 450 GeV by the Super Proton Synchrotron (SPS) before being injected into the LHC in both clockwise and anticlockwise directions, where they are further accelerated up to the collision energy. As a consequence of the radiofrequency acceleration scheme, protons are grouped into bunches. Under nominal operating conditions, each proton beam has 2808 bunches, containing $\sim 10^{11}$ protons per bunch with a bunch spacing of 25 ns. As they approach the collision points, the bunches are squeezed to about 15 µm in the transverse directions by 392 focusing quadripole magnets in order to increase the collision probability. The proton beams circulate in two ultra-high vacuum tubes (10^{-7} to 10^{-9} Pascal

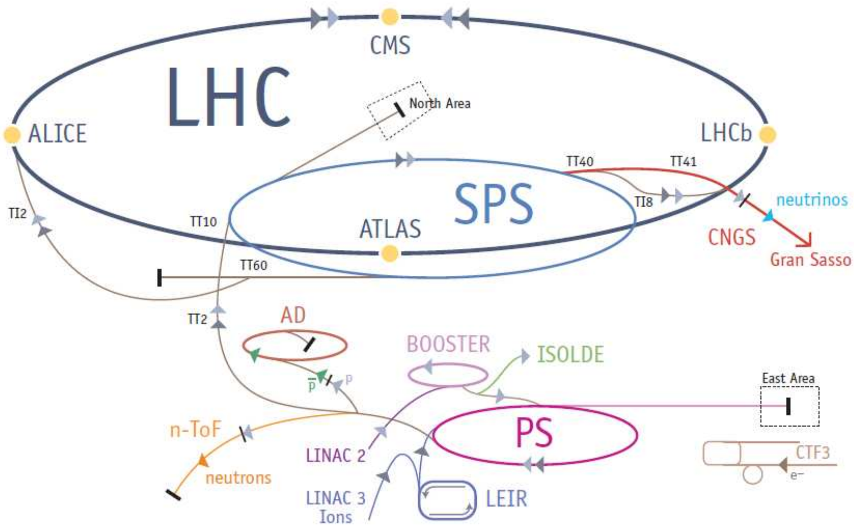


Figure 3.1: Overview of the CERN accelerator complex.

(Pa)[1]). They are kept on their orbit with 1232 dipole electromagnets of 15 meter length, delivering a 8.3 Tesla magnetic field. In order to avoid energy loss due to the 11850 A current flowing through the magnets, the LHC dipoles use niobium-titanium cables which become superconducting below a temperature of 10 Kelvin (K) (equivalent to -263.2℃). By using super-fluid helium, the cryogenic system allows the LHC to operate at a temperature of 1.9 K, needed to create this high magnetic field. The bending dipole magnets are connected in series in the eight arcs of the LHC. These arcs are linked by eight straight sections which host the different LHC detectors (ALICE, ATLAS, CMS, LHCb...) and the different injection, beam dumping or beam cleaning systems.

The collision rate (R) during an experiment can be factorized into a product of two independent terms ; the first one depends on the process of interest and corresponds to its cross-section, expressed in barns[2] (b). The second term is independent of the process under study and correspond to the luminosity ($\mathcal{L}$) achieved by the accelerator. For proton collisions :

$$R = \sigma_{pp} \times \mathcal{L} \tag{3.1}$$

[1]As a comparison, the Standard Atmospheric Pressure is equal to 100 kPa
[2] 1 barn is equal to 10^{-24} cm^{-2}

where σ_{pp} is the proton-proton cross-section. It can be shown that, for a Gaussian beam distribution, the luminosity can be written in first approximation as :

$$\mathcal{L} = \frac{f\ k_B\ n_p^2}{4\pi\sigma_x\sigma_y} \tag{3.2}$$

where f is the bunch collision frequency, k_B the number of colliding bunches in each beam, n_p the number of protons per bunch and σ_x, σ_y the gaussian sizes of the beam. The design luminosity, also often referred to as the high luminosity, of the LHC proton beams corresponds to a value of $\mathcal{L} = 10^{34}$ cm$^{-2}\cdot$ s^{-1}. During a given period, the number N of collisions, is given by :

$$N = \sigma_{pp} \times L \tag{3.3}$$

where L is the luminosity integrated over the period, called the integrated luminosity.

More than one proton collision may happen per bunch crossing, leading to overlapping collision events called *in-time pile-up events*. At high luminosity, a total of 22 in-time pile-up events per bunch crossing are expected on average.

On September, the 10^{th}, in 2008, the first proton beam in the LHC was successfully steered around the full 27 kilometres of the collider. Unfortunately, nine days after this successful start-up, a short circuit occurred in a magnet, leading to a leak in the cryogenic system and damaged the accelerator. After almost one year of repair, the LHC resumed its activity. A proton beam at SPS energy of 450 GeV circulated in the LHC the 20^{th} of November 2009. Three days later, proton beams were circulating routinely in both directions and a new world record was set on November, the 30^{th} with a beam energy of $1.18\ TeV$. On March the 30^{th}, the first LHC proton collisions at a centre-of-mass energy of $7\ TeV$ were recorded. Up to November 2010, the 3^{rd}, the proton beam luminosity has increased by more than a factor 10^5, from $\sim 10^{27}$ cm$^{-2}\cdot$s^{-1} to $\sim 10^{32}$ cm$^{-2}\cdot$s^{-1}. On November 2010, the 8^{th}, the CMS detector recorded its first lead-lead collisions at a centre-of-mass energy of $2.37\ TeV$ per nucleon pair and continued the data taking up to December, the 6^{th}. For the year 2010, the CMS detector recorded a total amount of 43.17 pb^{-1} of proton collision data and 8.7 µb^{-1} of heavy ion collisions. The LHC resumed its activity on March, the 14^{th} in 2011. Up to August, the 21^{st}, 2.35 fb^{-1} of integrated luminosity have been recorded by the CMS detector (Fig. 3.2). Over the last month of activity, the proton beam luminosity has reached peaks at $\sim 20 \times 10^{32}$ cm$^{-2}\cdot$ s^{-1}.

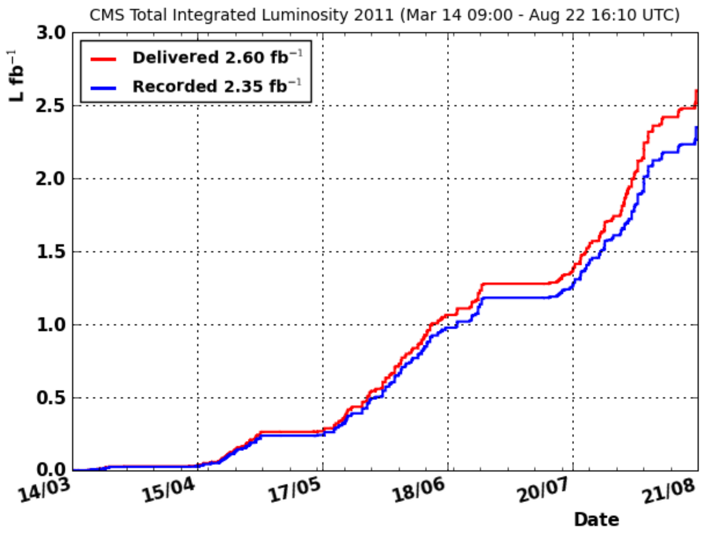


Figure 3.2: Integrated luminosity versus time delivered to (red), and recorded by CMS (blue) during stable beams at 7 TeV centre-of-mass energy.

3.3 The *Compact Muon Solenoid* detector

Surrounding the LHC beam interaction points, are located devices called detectors which interact and therefore provide informations about incident particles passing through. As the particles do not interact the same way with the matter, an efficient detection needs dedicated detectors for different kinds of particles. Furthermore, redundant detections usually ensure a better detection efficiency and a more accurate measurement of the particle kinematic properties.

The CMS detector [72] is a multi-purpose detector by opposition to detectors like the LHCb detector whose design optimizes the detection of a given physical process. The search for the Higgs boson has influenced the design of the CMS detector but it remains suitable to study a large class of processes. The CMS detector (Fig. 3.3) is 15 meters high, 22 meters long and weights 12500 tons. Its innermost part consists of silicon tracking devices which measure coordinates of impact points on charged particle trajectories and therefore allow their reconstruction. These devices are surrounded by calorimeters ; first an electromagnetic calorimeter which measures the energy of electrons and photons, then a hadron calorimeter which measures the energy of hadrons. These detectors are embedded into a solenoid which generates a uniform magnetic field of 3.8 Tesla, bending the charge particle trajectories and allowing the measurement of their momentum. Finally, surrounding

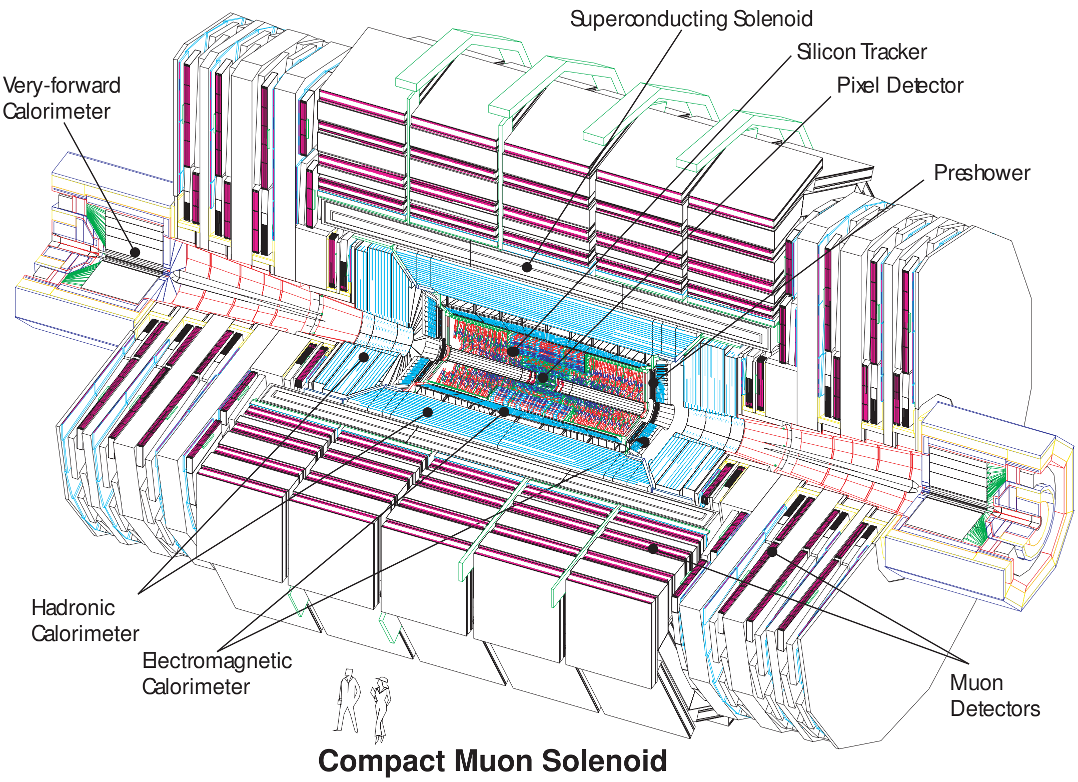


Figure 3.3: Overview of the CMS detector layout

the solenoid, there is the muon spectrometer in charge of measuring coordinates of additional impact points on muon trajectories.

The CMS coordinate system is a right-handed Cartesian coordinate system with the following conventions : the x-axis is horizontal, pointing to the centre of the LHC ring while the y-axis is vertical, pointing to the sky. The z-axis is taken parallel to the beam line. The origin corresponds to the centre of the CMS detector. The azimuthal angle, ϕ, is measured in the $x-y$ plane, starting from the x-axis. The polar angle, θ, represents the inclination angle with respect to the z-axis. However, as explained in the previous chapter, the proton momentum fraction carried by the interacting partons is unknown and the imbalance between the parton longitudinal momenta results in boosted final-state particles, mainly along the z-axis. Therefore, it is necessary to use Lorentz-invariant variables under boosts along the z-axis. That is the reason why the pseudo-rapidity, η, is often preferred to the polar angle. It is defined as :

$$\eta = -ln\left(tan\left(\frac{\theta}{2}\right)\right) \tag{3.4}$$

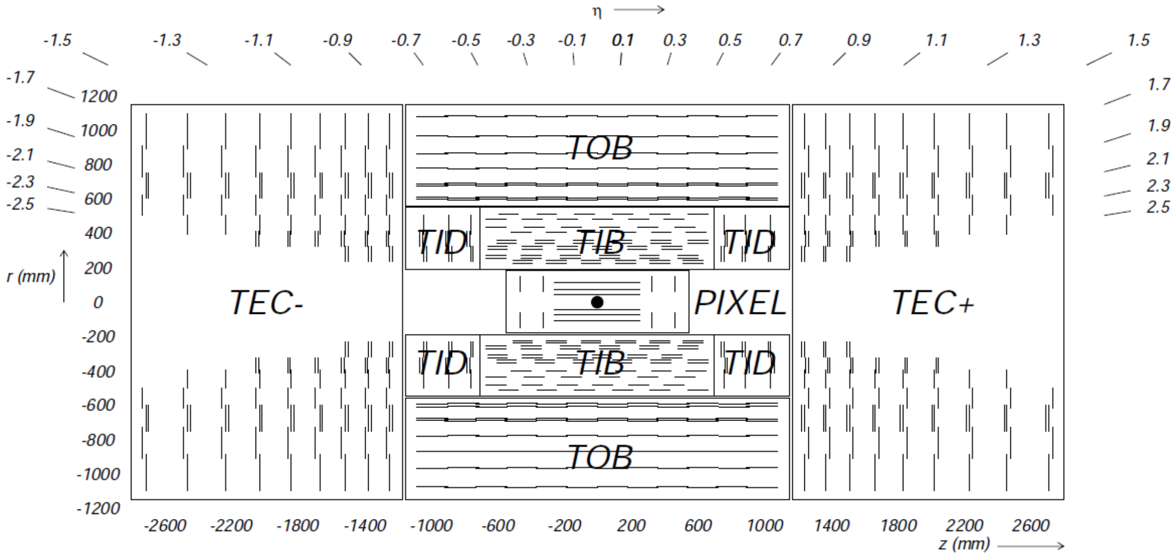


Figure 3.4: Schematic longitudinal overview of the CMS Tracker detector [72].

3.3.1 The silicon-strip and pixel tracking devices

The CMS tracking device (Tracker) [72, 78, 79] is the innermost part of the CMS detector (Fig. 3.4). It has a total length of 560 cm and a diameter of 232 cm. Its acceptance extends up to a pseudo-rapidity of 2.5. It consists of two main detectors : the *Pixel Vertex Detector* and the *Silicon Strip Tracker*. Both are located within the CMS solenoid. Their purpose is to measure the curvature of the charged particle trajectories in the magnetic field ; each time a charged particle crosses a detector unit, called module, this latter provides the spatial coordinates of the impact point, called a *hit*. Using these hits as inputs, particle trajectories are reconstructed later on with a track reconstruction algorithm. Finally, the particle momentum is calculated from the track curvature.

the Pixel Vertex Detector : this detector covers the inner part of the CMS Tracker (4 cm $< r < 15$ cm and $|z| < 49$ cm). It consists of three concentric cylindrical layers (barrel) and in two four-fan blade disks (end-cap), located on each side of the barrel. The three layers are located at mean radii of 4.4 cm, 7.3 cm and 10.2 cm respectively, and have an active length of 53 cm along the z direction. The end-cap disks have an inner and outer radius of 4.8 cm and 14.4 cm respectively and are located at mean longitudinal distance of 35.5 cm and 48.5 cm from the interaction point. The whole pixel detector contains 66 millions of active pixels arranged in 1400 modules, instrumenting a surface of 1 m^2. A pixel consists of a 150×100 µm^2 wide and 250 µm thick n^+ implant on a silicon substrate. The 3.8 Tesla magnetic field induces an azimuthal Lorentz drift of the electrons collected by the pixels in the barrel which enhances the charge sharing between the neighbouring implants and therefore increases the hit position resolution in the $r\phi$ direction. In the end-caps, the modules are tilted by 20° in order to benefit also from the Lorentz drift, resulting in a turbine-like geometry. The spatial resolution is measured to be around 15 µm in the z direction and 10 µm in the $r\phi$ direction. Signals produced by a pixel are read by a Pixel Unit Cell (PUC), integrated on a 52×80 pixel readout chip. For each PUC, a circuit turns the electrical signal into a positive logical one when it exceeds a tunable threshold.

the Silicon Strip Tracker : this detector surrounds the pixel detector, covering the outer part of the CMS Tracker up to 116 cm in radius and 560 cm in the z direction. Its geometry allows 8 to 14 measurements of the track impact points for trajectories with $|\eta| < 2.5$. It contains 9.3 millions of silicon microstrips, arranged into 15148 modules, instrumenting a surface of 198 m^2. Some modules, called double-sided modules, are mounted back-to-back with a stereo angle of 100 mrad between the microstrips in order to provide a measurement of a second coordinate. Each microstrip consists of a p^+ implant located on a silicon substrate whose thickness varies between 320 µm and 500 µm for the inner and outer part of the Silicon Strip Tracker respectively. Its layout is divided into three main structures :

1. the *Tracker Inner Barrel and Disks* consists of four concentric cylindrical layers for the barrel (TIB), completed by three disks on each end for the end-caps (TID). They extend in radius to 55 cm and cover up to ± 65 cm in the z direction. The silicon strips are oriented parallel to the beam axis for the barrel and radially for the disks. The distance between two strips (pitch) varies from 80 µm for the innermost barrel layer to 120 µm for the outermost one and from 100 µm to 141 µm for the disks. The first two layers and rings contain double-sided modules.
2. the *Tracker Outer Barrel* (TOB) consists of 6 layers which extend in radius up to 116 cm and cover up to ± 118 cm in the z direction. The microstrips have a pitch of 183 µm for the first four layers and 122 µm for the last two layers. The first two contain double-sided modules.
3. the *Tracker End-Caps* (TEC) : they consist of nine disks which instrument on either side the region with $|z|$ between 124 cm and 280 cm and have an inner radius which varies from 22 cm to 55 cm and an outer radius of 113.5 cm. The microstrip substrate thickness is of 320 µm for the four innermost rings and of 500 µm for the three outermost rings. Because of the trapezoidal shape of the modules in this part of the Tracker, the radial strip pitch varies from 97 µm to 184 µm. The first, second and fifth rings contain double-sided modules.

At the end of the silicon strips, the electrical signal is collected and pre-amplified in one of the 128 input channels of an APV readout chip (Analog Pipeline in Voltage mode). On a positive decision of the first level rigger, the signal is sent via optical fibers to the Front-End Driver (FED) where the signal is digitized by a Analog-To-Digital Converter (ADC). The data are then filtered to remove noise signals via a procedure called zero-suppression which allows to significantly reduce the data volume which is finally sent to the Data Acquisition system (DAQ).

3.3.2 The electro-magnetic and hadronic calorimeters

Surrounding the silicon track detector, the CMS calorimeter consists of two detectors : the electro-magnetic calorimeter (ECAL) and the hadron calorimeter (HCAL). The purpose of this calorimeter is twofold ; on one hand, it measures the energy of the incident particles. On the other hand, it allows to distinguish between particles for which electromagnetic interactions dominate, electrons and photons, from particles undergoing hadronic interactions like protons, kaons or pions. Besides, it allows to measure the energy of neutral particles which cannot be detected by the tracker. The calorimeter principle is based on a destructive process ; incident particles interact successively with the calorimeter material and thus produce a cascade of particles which also further interact. Energy deposits of primary and secondary particles are then measured so that it is possible to link this measured energy to the energy of the incident particle.

ECAL : it is the main component of the CMS detector to measure the energy and identify electrons and photons. It is a calorimeter based on lead tungstate crystal scintillators ($PbWO_4$), acting both as absorbers and detectors. The choice of the active material was mainly driven by the requirement of an energy resolution, precise enough to be sensitive to the decay of a potential Higgs boson into two photons. Additional constraints originated from the need of a radiation hard calorimeter, compact enough to be embedded within the CMS solenoid. The chosen type of crystals fulfils all these requirements ; its high density (8.28 g $\cdot$ cm^{-3}) and short radiation length ($\lambda_0 = 0.89$ cm) allowed to design a compact detector. Moreover, the small Molière radius[3] $R_M = 2.2$ cm of this type of crystal allows a position resolution below 1 mm for an incident electron energy above 20 GeV [80], by segmenting the detector into pieces with dimensions of the order of R_M. The energy of the incident particles are

[3]The Molière radius, R_M, is defined as the radius of a cylinder containing 90% of the shower's energy deposit.

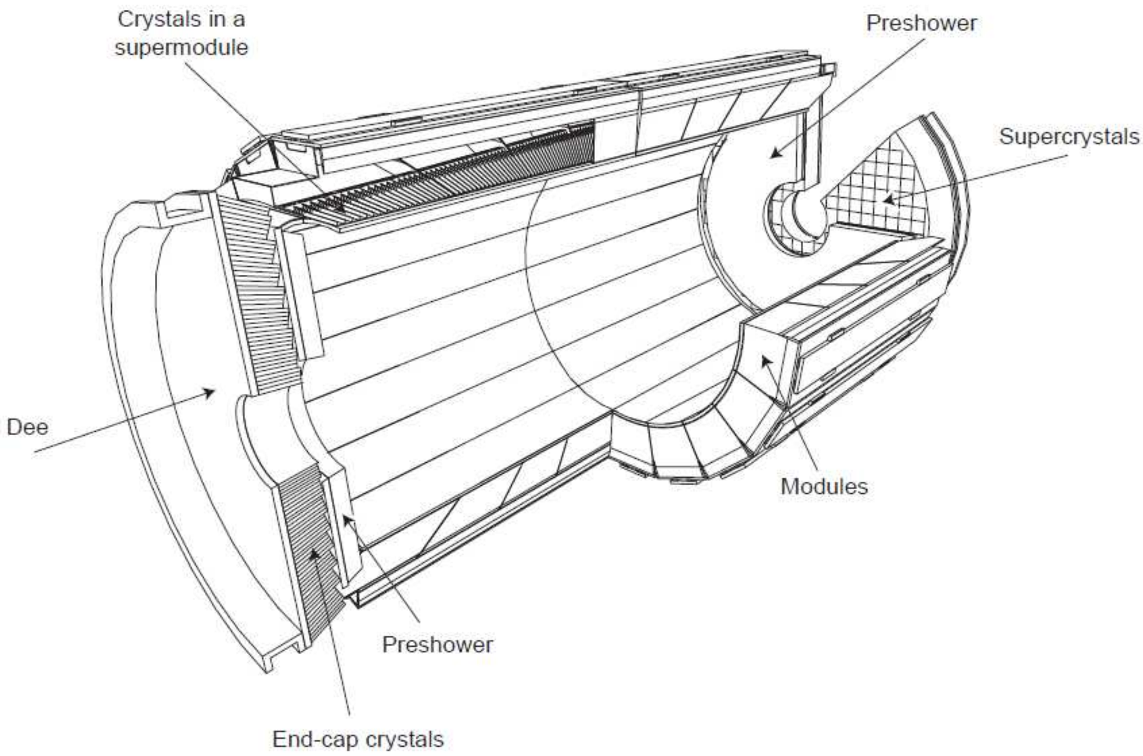


Figure 3.5: Layout of the CMS electro-magnetic calorimeter, showing the barrel supermodules, the two end-caps and the pre-shower detectors [72].

then reconstructed by summing 3×3 matrices of crystals, which contain 99% of the shower's energy deposit. Furthermore, the scintillation decay time, which is the time taken by the crystal to emit light consequently to the absorption of the incident particle energy, is of the same order of magnitude as the LHC bunch crossing time : 80% of the light is emitted within 25 ns. Although radiation resistant, crystals show a limited but rapid loss of optical transparency under radiations, which depends on the radiation dose. The monitoring of the crystal light transparency is ensured by laser probes. The ECAL (Fig. 3.5) is divided into three main parts : the central barrel divided into two parts, EB+ for $z > 0$ and EB- for $z < 0$ and, at each end, a disk, EE+ and EE-, in front of which a pre-shower detector, ES, is located.

- the ECAL barrel has an inner radius of 1.29 m and is segmented in 360 and 2×85 crystals in the ϕ and η directions respectively, for a total of 61200 crystals. Each of them has a tapered shape with a front section of 22×22 mm^2 (0.0174×0.0174 in the $\eta - \phi$ space) and a length of 230 mm, corresponding to $25.8\ \lambda_0$. The light produced by the electromagnetic shower initiated by the incident particles is collected by a pair of avalanche photodiodes (APD's) mounted on the rear face of each crystal. The crystals are in a quasi-projective direction ; they point slightly off by an angle of three degrees to the nominal interaction point in order to avoid dead spaces between crystals being aligned with the particle trajectories. Subsets of 400 to 500 crystals, depending on their position in η, are called modules and four modules are assembled in a supermodule which contains 1700 crystals. Each supermodule covers 20° degrees in ϕ, for a total of 36 supermodules.

- the ECAL end-caps cover regions with $1.479 < |\eta| < 3.0$ and are located at a longitudinal distance of 315.4 cm from the nominal interaction point. Each end-cap is divided into two halves, called *dees*. Each of these dees contains 31622 crystals grouped in units of 5×5 crystals (supercrystals). In this part of the detector, crystals have a front section of 28.6×28.6 mm^2 for a length of 220 mm and the light is collected by vacuum phototriodes (VPT's). In the end-cap, the crystals point at a focus of 1300 mm beyond the nominal interaction point, resulting in off-pointing angle values ranging from 2 to 8 degrees.
- the Preshower detector is a 20 cm thick detector located in front of each end-cap. Its purpose is to identify neutral pions in the end-caps within the region $1.653 < |\eta| < 2.6$ by improving the determination of the spatial coordinates for their decay products (electrons and photons). It is a sampling calorimeter which consists alternatively of layers of lead radiators, initiating electromagnetic showers, and of silicon strip sensors which measure a fraction of the deposited energy and the transverse shower profile.

The energy resolution of the electromagnetic calorimeter can be parametrized as the following :

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E}} \otimes \frac{b}{E} \otimes c \tag{3.5}$$

where E is the energy of the incident electron or photon, a is the stochastic term which accounts for event-by-event fluctuations in the lateral electromagnetic shower containment, photo-statistics and photo-detector gain. The noise term, b, is a function of the level of electric noise and c, the constant term, depends mainly on the inter-calibration constants between the ECAL crystals. In 2004, the energy resolution of the ECAL has been measured with electron beams with an energy between 20 GeV and 250 GeV with no magnetic field nor tracker material in front. Due to time constraints, only nine supermodules have been tested. The following values for the parametrization terms were measured [81] : $a = 0.028$ GeV$^{1/2}$, $b = 0.12$ GeV and $c = 0.003$. For energies higher than 90 GeV, the resolution starts to be dominated by the constant

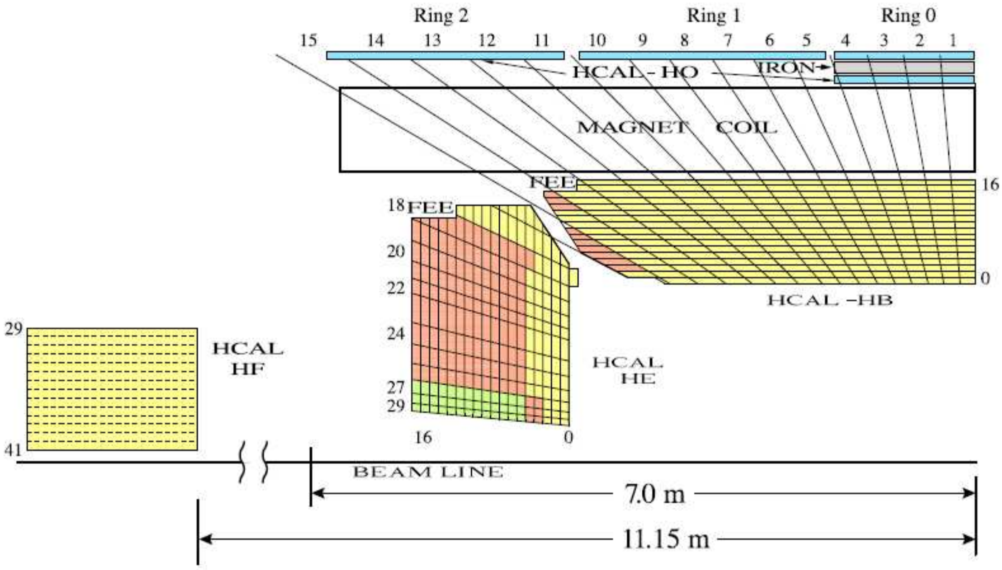


Figure 3.6: The CMS HCAL detector (quarter slice) in the (r, z) plane. FEE indicates the locations of the Front End Electronics for HB and HE. The signals of the tower segments with the same colour are added optically, to provide the HCAL longitudinal segmentation [72].

term but remains below the design resolution of 0.5%. In order to calibrate the remaining supermodules, in-situ inter-calibration methods have been developed in CMS in order to recover an energy resolution below 0.5% at an energy of 100 GeV. Using the first 123 nb^{-1} of data at a centre-of-mass energy of 7 TeV, collected by the CMS detector during 2010, a calibration precision of 1.2% was achieved in the barrel and is consistent with the Monte-Carlo simulations [82].

HCAL : As they interact differently with matter compared to electrons and photons, hadrons need a dedicated detector to identify them and measure their energies. Surrounding the ECAL, the hadron calorimeter (HCAL) is also mostly located within the solenoid (Fig. 3.6). It consists of the HCAL barrel (HB), closed at each end by the HCAL end-caps (HE) and cover regions up to $|\eta| = 3$. A forward hadron calorimeter (HF) is placed at 11.15 m from the interaction point, extending the detection coverage up to $|\eta| = 5.2$. Due to its limited volume, energetic particles produced during the showering can escape the HCAL and then lead to a mis-measurement of the incident energy. The barrel part of the HCAL has therefore been completed with an outer hadron calorimeter (HO), located outside the solenoid.

The CMS hadron calorimeter is a sampling calorimeter ; it is made of successive detection layers which interleave layers of non-instrumented absorber to initiate the shower and allow its containment within the limited volume of the calorimeter.

Detection layers measure the energy of the shower at different development stages and thus, only a fraction of the deposited energy is measured (sampling). In CMS, absorber layers are made of brass (70% Cu and 30% Zn) or steel and the detection layers are made of scintillators. The sampling fraction, defined as the energy deposited in the detection layers divided by the energy deposited in the absorber layers, is around 7%.

- HCAL Barrel : this cylindrical part of the HCAL detector is split into two halves, HB$+$ and HB$-$, with a radius between 1.77 m and 2.95 m, covering regions up to $|\eta| = 1.3$. Each half barrel consists of 18 identical azimuthal wedges, covering 20 degrees in ϕ. It is made of flat rectangular absorber plates, aligned parallel to the beam ; the innermost and outermost layers are made of steel to increase to structural strength, the others of brass. It results in a material thickness ranging from $5.82\ \lambda_I$[4] at $|\eta| = 0$ to $10.6\ \lambda_I$ at $|\eta| = 1.3$. The ECAL crystals in front of the HCAL barrel add roughly $1.1\ \lambda_I$ of material. The plastic scintillators are divided in 16 towers in the η direction, leading to a segmentation of 0.087×0.087 in the $\eta - \phi$ space.
- HCAL End-cap : both end-caps cover the region $1.3 < |\eta| < 3.0$ and are mounted on the muon end-cap yoke. They consist of brass disks, interleaved with trapezoidal shaped scintillators with a segmentation in the $\eta - \phi$ space of 0.087×0.087 for $|\eta| < 1.6$. The segmentation ranges from 0.09×0.175 to 0.35×0.175 for $|\eta| > 1.6$.
- HCAL Outer Barrel : it is composed of five rings, following the structure of the magnet return yoke and of the muon spectrometer. The locations of rings $0,\ 1$ and 2 are shown in figure 3.6. Each ring is divided in 12 identical sectors in ϕ. The main purpose of this detector part being to compensate for the lower absorber depth of the HB, $5.82\ \lambda_I$ at $\eta = 0$, the ring 0, covering the region $|\eta| < 0.35$, has two scintillator layers on either side of a 19.5 cm thick iron absorber layer. Rings ±1 and ±2 have only one layer of scintillators. These scintillators are segmented in tiles which roughly map the HB segmentation, leading to a calorimeter cell granularity of 0.087×0.087 in the $\eta - \phi$ space.
- HCAL Forward : its design consists in a cylindrical steel structure with an inner and outer radius of 12.5 cm and 130 cm respectively, housed in a hermetic radiation shielding made of layers of steel and concrete. This structure is divided into wedges of 20 degrees in ϕ. Each wedge contains quartz fibres, which collect

[4] The interaction length, λ_I, is defined as the mean free path of a particle between two successive nuclear interactions with the medium.

the Čerenkov light produced by charged shower particles passing through. These fibres are bundled in such a way that the HF segmentation reaches 0.175×0.175 and 0.175×0.35 in the $\eta - \phi$ space for $|\eta| < 4.7$ and for $|\eta| \geq 4.7$ respectively.

In the HB and HE, the scintillator layers of a given η segmentation are assembled in so-called calorimeter tower, leading to a total of 29 towers as shown in Figure 3.6. Each scintillator tile of a tower is read out by an embedded wave-length shifting fibre. The optical signals, detected by hybrid photodiodes (HPD's), of all tiles in a tower are added for towers 1 to 14. The expected degradation of the scintillators due to radiations is more important in the forward region. Therefore, for towers 15 to 29, the scintillators have different read-outs, resulting in a longitudinal segmentation of the towers. This segmentation allows to apply separately different corrections to the scintillator calibration coefficients in order to restore degraded energy resolution. In the HF, the quartz fibres used to collect the Čerenkov light are of two kind : long fibres (L) of 165 cm which measure the total signal coming from the full material length and short fibres (S) which read the signal at a depth of 22 cm from the front of the detector. The hadronic energy resolution of the combined electromagnetic and hadronic barrel calorimeters has been measured with electron and pion beams [83]:

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E}} \otimes b \tag{3.6}$$

with $a = 0.847 \pm 0.016$ GeV$^{1/2}$ and $b = 0.074 \pm 0.008$, achieving a resolution of $\sim 11\%$ at $E = 100$ GeV.

3.3.3 The superconductiong magnet

The inner part of the CMS superconducting magnet consists of a 220 ton cylindrical solenoid of 6 m diameter and 12.5 m length, surrounding the HCAL barrel. It is made of four layers of superconducting NbTi, maintained at a temperature of 4.5 K, producing a four tesla magnetic field in its bore when operating at the nominal current of 19.1 kA. The flux return is ensured by a 10000 ton iron yoke, interleaved with the muon system.

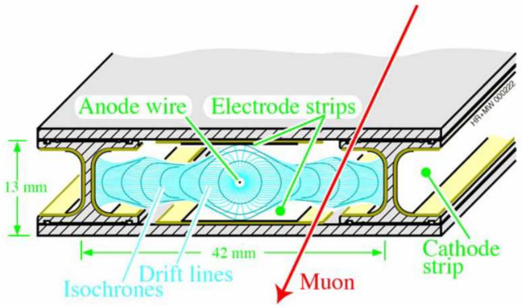


Figure 3.7: Sketch of a drift tube, showing the electrodes, the field lines and the isochrones [72].

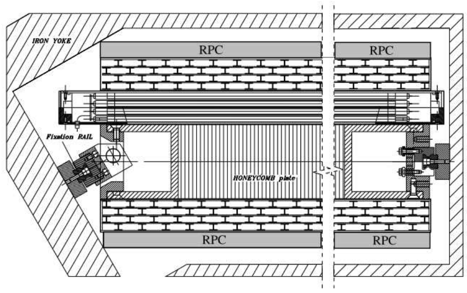


Figure 3.8: View in the $r\phi$ plane of MB chamber in position inside the iron yoke [72].

3.3.4 The muon spectrometer

The CMS muon system constitutes the outermost part of the CMS detector, surrounding the solenoid magnet. Its layout is then naturally divided into a cylindrical barrel section and a planar end-cap section on each side. In the barrel, the muon detection is ensured by drift tubes (DT) and by cathode strip chambers (CSC) in the two end-caps. The main purpose of this system is to provide the coordinates of impact points along the muon trajectories, in addition to the ones provided by the Tracker, and therefore allow a more precise determination of the muon momentum. Unlike most particles, muons are not stopped in any of the CMS's calorimeters. Thus, they are likely to be the only particles able to leave a signal in the outermost part of the CMS detector.

Although DT ans CSC subsystems are able to trigger on the muon transverse momentum, a complementary faster trigger system based on resistive plate chambers (RPC) has been added to both the barrel and the end-caps.

Muon Barrel (MB) and Drift Tubes (DT) : the Muon Barrel is assembled from drift tubes (Fig. 3.7). A drift tube is made of two rectangular aluminium plates, $42 \times 2400\ \mathrm{mm}^2$, separated by a I-shaped, 13 mm high, piece of aluminium. At the centre is located a 2.4 m long and 50 µm thick gold plated stainless steel anode wire ; the cathode, 11.5 mm wide, are fixed on the I shapes on both sides. The drift field is configured by 16 mm wide field electrodes fixed on the plate. The cathode and field electrodes are made of 50 µm thick aluminium tape insulated from the grounded aluminium walls. Figure 3.7 also shows the resulting field lines and the lines of equal drift times, called isochrones. Each cell is filled with a gas mixture of argon (85%) and carbon dioxide, CO_2 (15%) which gets ionized by charged particles passing through the cell.

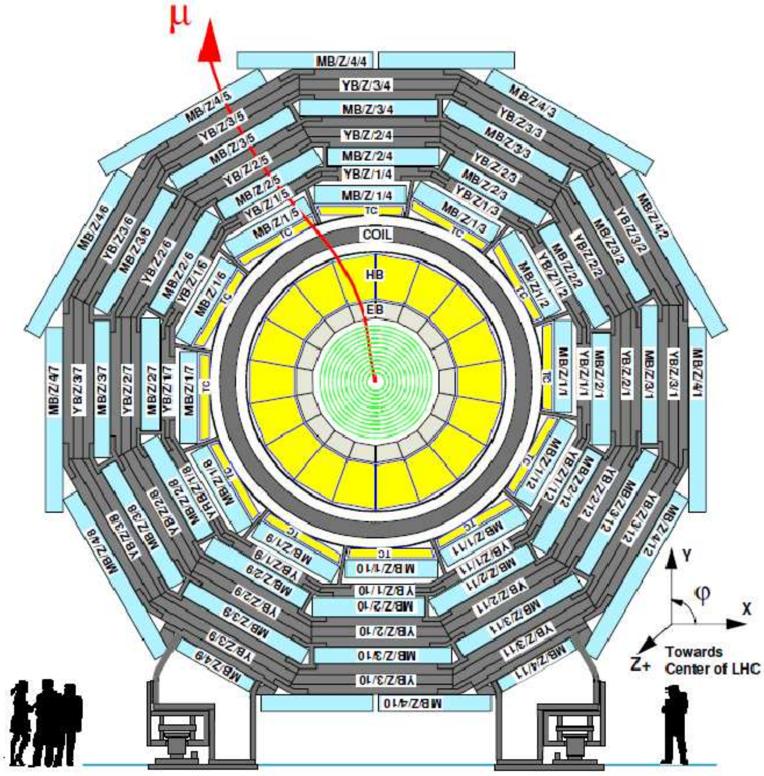


Figure 3.9: Layout of the CMS barrel muon DT chambers in one of the 5 wheels.

Subsequent electrons then drift to the anode, with a maximal drift time of 380 ns due to the cell dimensions and applied voltages. Finally, an charge-inducing avalanche develops near the anode wire. Drift tubes are assembled in so-called super-layers, each containing four layers of drift tubes staggered by half a tube to resolve the left-right ambiguity. A MB chamber is then made of two or three super-layers crossed at right angle, embedded in the iron yoke (Fig. 3.8). Wires of the two outermost super-layers are parallel to the beam line, providing coordinates of impact points in the $r - \phi$ plane with a resolution of ~ 100 µm. The inner super-layer contains wires orthogonal to the beam, measuring the z-position with a resolution of ~ 150 µm. The muon barrel detector consists of four concentric layers, called *stations*, interspersed among the layers of the magnet return yoke. Each station, is made of 60 DT for the three innermost ones and of 70 DT for the outermost one, for a total of 172000 wires, covering regions up to $|\eta| < 1.2$. In Figure 3.9, the MB chambers are labelled MB/Z/X/Y, where Z is the wheel number, X is the number of the station the DT belongs to (from 1 to 4) and Y is the MB chamber number (from 1 to 12).

Muon End-Caps (ME) and Cathode Strip Chambers (CSC) : CSC chambers are trapezoidal multi-wire proportional chambers (Fig. 3.10), made of six anode gold plated

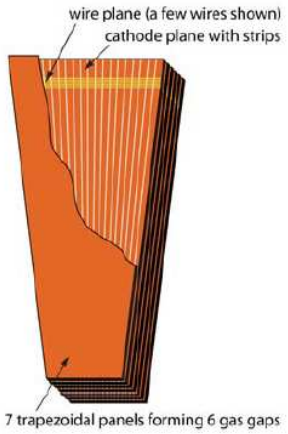


Figure 3.10: Layout of a CMS cathode strip chamber chamber. Only a few wires are shown to indicate their azimuthal direction. [72].

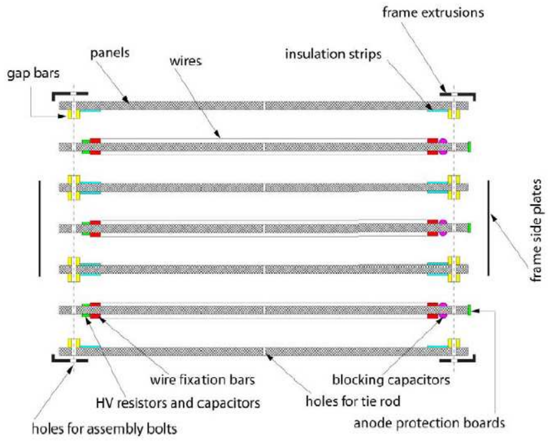


Figure 3.11: Mechanical design of the CMS cathode strip chambers [72].

tungsten wire planes, interleaved among seven cathode panels and filled with a gas mixture of argon (40%), carbon dioxide (50%) and tetra-fluoro carbon (10%) (Fig. 3.11). Wires run azimuthally and provide the r-coordinate of impact points. Cathode panels contain strips running along the ϕ-direction. In each end-cap, they are a total of 234 CSC chambers divided into four stations, mounted on the disks closing the magnet solenoid and arranged into two concentric rings. In Figure 3.12, the CSC are labelled MEX/Y, where X is the station number and Y, the ring number. The CSC chambers are designed to achieve a timing resolution of ~ 4 ns and an off-line spatial resolution in the $r - \phi$ plane of ~ 75 µm for ME1/1 and ME1/2 and of ~ 150 µm for all the other ones. Studies using cosmic muon data have shown that CSC chambers meet these requirements [72].

Resistive Plate Chambers : these double-gap gaseous plate detectors have a time resolution below 2 ns, much shorter than the 25 ns between two consecutive LHC bunch crossings, allowing these detectors to be used to unambiguously assign a muon track to its corresponding proton bunch crossing. In the barrel, an array of RPC is located on each side of the MB super-layers for the two innermost stations (see Figure 3.8). In the third and fourth stations, two RPC layers are located side-by-side on the inner side of the MB chambers. In total, there are 480 RPC chambers in the barrel, each one being 2455 mm long in the beam direction and with a width ranging from 1500 to 2080 mm. Each end-cap consists of three stations of trapezoidal RPC, interleaved in

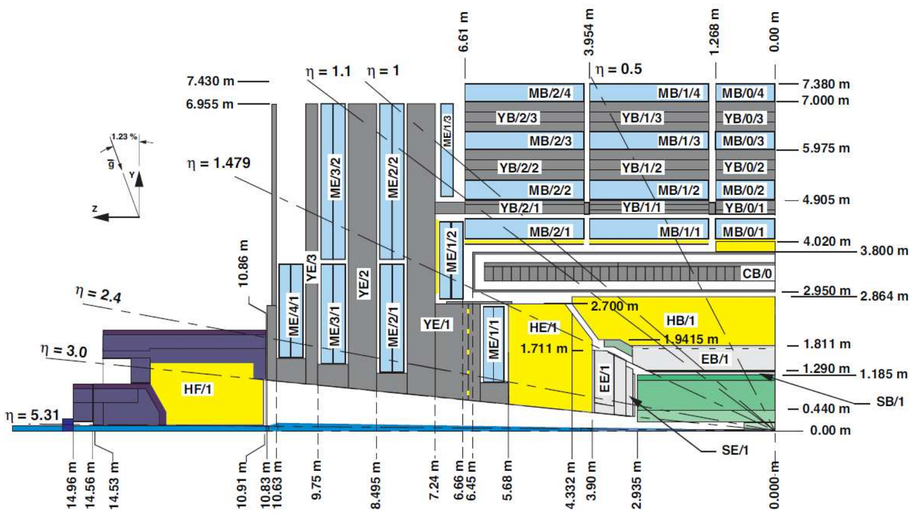


Figure 3.12: Quarter-view in the $r-z$ plane of the CMS detector. Cathode strip chambers of the End-Cap Muon system are highlighted in blue and labelled ME [72].

the four CSC stations and arranged in three concentric rings, for a total of 216 RPC chambers.

3.4 The CMS on-line event selection system

As mentioned in section 3.2, proton bunch spacing under nominal operating conditions is 25 ns with an average of 22 in-time pile-up collisions. This rate is unfortunately so high that the informations produced by the CMS detector cannot be recorded for each bunch crossing. Moreover, with an average size of 1.5 MB per recorded bunch crossing, the resulting amount of data to be stored would exceed the CMS storage limits. Therefore, CMS is equipped with an on-line event selection system, called trigger, in charge of recording collisions with high interest from the physics point of view. In most of the cases, proton-proton collisions lead to the production of hadrons with low transverse momentum, called minimum bias events. In comparison, the production and decay of heavy particles like a potential Higgs boson or a top quark lead to leptons or jets with high transverse momentum which are thus used by the CMS on-line event selection. However, usual lepton and jet reconstruction algorithms are typically too slow to cope with an acceptable event recording rate of the order of 100 Hz. Dedicated reconstruction algorithms have been developed and incorporated in the trigger system. These algorithms provide a more coarse description of the different physics objects than during the off-line reconstruction but are able to deal with the targeted event recording rate.

The CMS trigger system consists of two sub-systems :

- the level-1 trigger (L1) system, reducing the event rate to ~ 100 kHz, is based on hardware and has only access to the data produced by the calorimeter and the muon system. The tracker system, for which the reconstruction is time-consuming, can not be used at that level.
- the high-level trigger (HLT) system, further reducing the event rate to $\mathcal{O}(100)$ Hz, is based on softwares running on a farm of processors, allowing more sophisticated object reconstruction methods to be used, starting from data kept at L1.

At level-1, the CMS Data Acquisition System (DAQ) stores data in buffers with a limit of 128 different bunch crossings, allowing a maximal latency of the level-1 trigger decision of 128×25 ns $= 3.2$ µs. The level-1 trigger system includes 3 sub-systems : the calorimeter trigger, the muon system and the global trigger (Fig. 3.13). the first two sub-systems are further divided in independent systems based on the HF, HCAL and ECAL detectors for the calorimeter trigger and on DT, CSC and RPC detectors for the muon trigger. The role of these systems is to identify and reconstruct physics objects, called trigger objects at this stage. Based on the positions, the kinematic properties and the reconstruction quality of these objects, the Global Trigger actually decides to reject or accept the event for further evaluation by the HLT.

At high-level, the decision to store the event for off-line analyses is made. A software-based trigger offers the full flexibility of a programmable system, the possibility to run multiple triggers and multiple events in parallel. It is organized in sequential steps, using increasingly complex and time-consuming algorithms. This allows to reject as fast as possible uninteresting events.

As an illustration, we shall now review the relevant trigger for this thesis : the muon trigger [84].

The muon trigger consists of two successive levels : the level-1 hardware trigger and the high-level software trigger which is structured in two different sub-levels, the level-2 and the level-3. Along these levels, increasingly tighter constraints are applied on the minimal quality and kinematic properties of the reconstructed muon. The complexity and therefore the needed computing time also increase along these three steps. If the reconstructed muon fails to fulfil the requirements at a given step, the entire procedure is stopped. On a

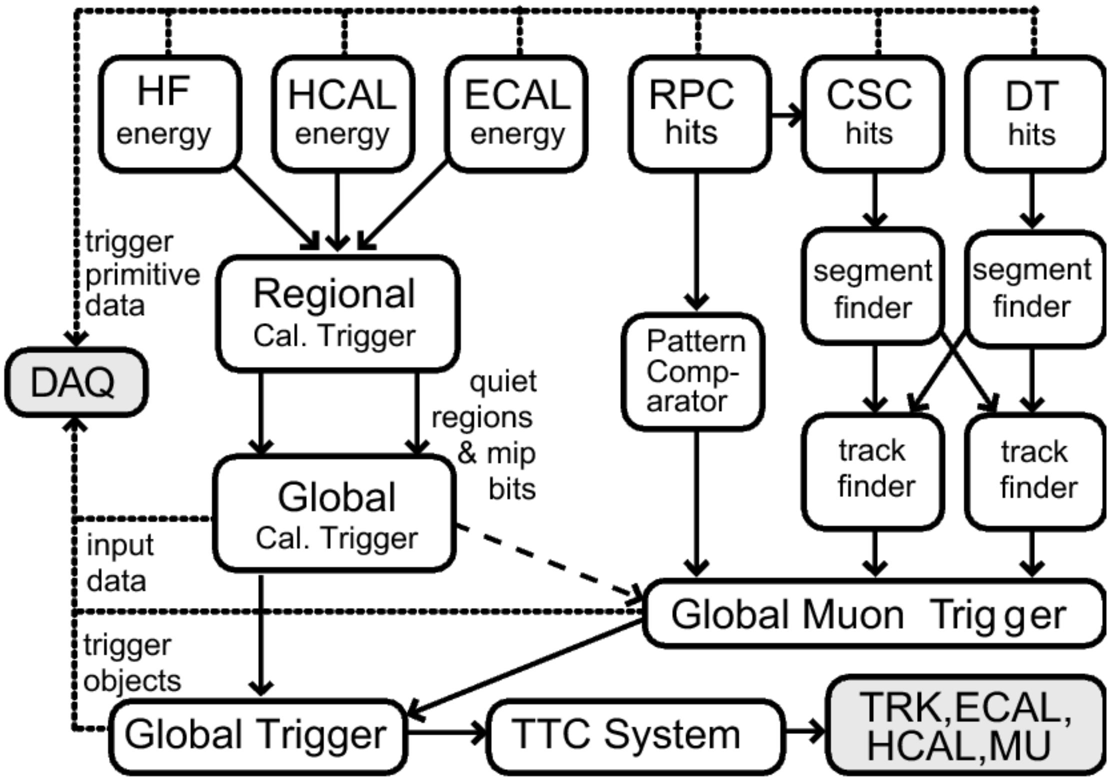


Figure 3.13: Schematic overview of the different Level-1 trigger sub-systems and of their relationships.

positive decision of the level 1 trigger, the HLT muon reconstruction chain starts ; at level 2, informations from the muon system only are used to reconstruct muon trajectories, the so-called stand-alone muon reconstruction. At level 3, tracks compatible with level 2 stand-alone muons are reconstructed in the inner tracker, using a regional pattern recognition and track fitting algorithm. Isolation informations from the inner tracker are also added in case of isolated muon based trigger. The on-line and off-line muon reconstruction algorithms overlap to a large extent. As the off-line muon reconstruction is described in details in section 4.2.1, we shall rather highlight in what follows the HLT-specific aspects of the muon reconstruction.

Level-1 muon trigger : at this level, the trigger decision is based on the informations provided by the three muon detection subsystems, DT, for the barrel region, CSC for the end-cap regions and RPC for both regions. These informations are combined by the Global Muon Trigger (GMT) and transmitted to the Level 1 Global Trigger (GT). Each subsystem has its own trigger logic :

- DT and CSC process the informations for each muon chamber locally ; muon track segments which consist of sets of aligned hits are first reconstructed. Then for each reconstructed muon segment and for each station, a so-called trigger primitive is delivered which consist of its position, its direction and a quality code. These trigger primitives are then combined by the Track Finder algorithm to reconstruct tracks and assign a transverse momentum to each reconstructed ones. The four muon candidates with the highest transverse momentum and quality from each subsystem are finally sent to the GMT.
- RPC hits from all the muon stations are collected by the Pattern Comparator which translate aligned hits along possible tracks into muon candidates with an estimated transverse momentum. These candidates are then ranked in decreasing order according to their quality code and their transverse momentum and finally up to the first four candidates from the barrel and the first four from the end-cap regions are sent to the GMT.

Among these muon candidates, the Global Muon Trigger searches for possible matched pairs between RPC and either DT or CSC candidates. The informations of the matching candidates are then merged and the remaining candidates are again ranked in decreasing order according to their reconstruction quality and transverse momentum. The four first candidates are then sent to the Global Trigger which applies transverse momentum and quality thresholds. Finally, for each of the successful muon candidates, the level 1 muon trigger returns the global position of the associated hit in the second muon station (MB2/ME2), a quality code and a number, according to the estimated muon transverse momentum, the highest one corresponding to $p_T^{\mu} \geq 140$ GeV$/c$.

High-level muon trigger : it reconstructs level-2 tracks in the muon system using a pattern recognition algorithm.

- level 2 muon trajectory seeding : for each level 1 muon candidate, a trajectory state, including muon momentum and direction, is created at the position of the level-1 hit at the second muon station surface with a transverse momentum equal to the level 1 muon candidate. The trajectory is then extrapolated with a Kalman filter [85] to the innermost compatible muon detector layer to evaluate the kinematic parameters of the muon candidate at this level.
- level 2 muon trajectory building : the same pattern recognition algorithm is used in the on-line and the off-line muon reconstruction, described later in section 4.2.1. Briefly, it consists of an inside-out, followed by an outside-in trajectory propagation based on a Kalman filter. The first propagation starts from the level-2 muon trajectory seed and uses DT or CSC track segments. The second filter starts from the last update of the track parameters obtained during the inside-out propagation and performs a trajectory propagation using individual hits instead of track segments. Then the track candidates are cleaned for tracks sharing hits or for badly reconstructed tracks based on their normalized χ^2 value. Finally, for each selected track, a copy is created whose parameters have been updated with a constraint to the beam-spot position.

The level 2 trigger returns for each muon candidate its position and transverse momentum which serve as a seed for the level 3 trigger muon reconstruction. The reconstruction of level-3 muons is identical to the muon off-line reconstruction, described in details in section 4.2.1 : it combines informations from both the muon system and the tracker system. First, track candidates are reconstructed in a limited region of the tracker in order to reduce the computing time and then, are matched to level-2 muon tracks found in the muon system.

4

Off-line object reconstruction with the CMS detector data

As explained in chapters 1 and 2, proton collisions are naturally described at theoretical level in terms of the most elementary particles, leptons and partons. At experimental level, only electrons and muons are observable in a detector. Tau leptons decay shortly after production and neutrinos interact so weakly that they escape detection, leading to missing energy. As quarks and gluons produced during hard interactions may radiate new quarks or gluons and evolve into mesons and baryons via the process of hadronization, their kinematic properties have to be inferred from the resulting flow of observable hadrons, called jets. In addition, at detector level, informations about detected particles are only accessible in terms of impact points, called hits, in the inner tracker and muon system and energy deposits in the calorimeter. Dedicated algorithms have been therefore developed in CMS to combine these informations and reconstruct the particles produced during the collision. As described in section 4.1, charged particle trajectories and vertices are first reconstructed from hits provided by the inner tracker. They are then combined either with energy deposits detected in the electromagnetic calorimeter in the case of electrons or with trajectories reconstructed in the muon system in the case of muons, as explained in section 4.2. Jets are more complex objects ; they are made of charged and neutral particles and lead to energy deposits both in the electromagnetic and hadronic calorimeters. Jet reconstruction algorithms used by the CMS collaboration are described in section 4.3, followed by the description of the algorithms used to identify the jets originating from heavy quarks. Algorithms used to compensate for detector and jet reconstruction algorithm inefficiencies and recover the kinematic properties of the original parton are described as well. The last section, 4.4, is dedicated to the reconstruction of the missing transverse energy due to neutral weakly interacting particles as neutrinos. The high-level physics objects described in this chapter serve as a basis for the analysis presented in the following chapters. For the description of other types of objects which have been omitted here, we refer to [86].

4.1 Track and primary vertex reconstruction with the CMS Tracker

4.1.1 Track reconstruction

As mentioned in section 3.3.1, the coordinates of the impact points of charged particles with the layers of the CMS tracker, called hits, are used as inputs to reconstruct their trajectory and determine their transverse momentum and emission direction. Due to the density of charged particles produced during proton collisions, it is however impossible to unambiguously assign the correct impact points to a charged particle. Pattern recognition algorithms aim to solve this issue. In CMS, both tasks are addressed by a single algorithm consisting of four successive steps [87] :

Seed Generation : hits are associated into small track segments, called *seeds*, which define the initial trajectory parameters. To properly define a seed, one needs at least three hits or two hits and an additional constraint, either from the position of the beam crossing region, called beam-spot[1], or from the position of a vertex. Four kinds of seeds are used as input to the second step, the Pattern Recognition. The first kind of seeds are formed from triplets of pixel hits, called Pixel Triplets, to obtain a first

[1]The beam-spot position is measured with the so-called $d_0 - \phi_0$ algorithm [88] which exploits the correlation between the beam-spot position and the shape of the distribution of the track impact parameter as a function of the track azimuthal angle. Using 1000 tracks, a precision of 2 µm for the transverse beam-spot position can be reached.

estimate of the vertex coordinates. Then, seeds are formed from pairs of tracker measurements compatible with one of the vertices. However, the vertex constraint is not used during the fit of the corresponding track. A similar kind of seeds using a looser beam-spot compatibility requirement instead of the vertices are also formed to recover tracks originating from long-lived particles which are thus not emitted at the primary vertex. The last kind of seeds are only formed with hits from the silicon strip detector. They ensure the track reconstruction of decay products of long-lived particles decaying outside the pixel detector.

Pattern Recognition : starting from a seed, the track is extrapolated outward using a Kalman filter method [85] which takes into account energy loss and multiple scattering ; a list of compatible tracker layers is made. Then for each layer, a list of compatible hits is made and, for each hit, a new track is formed by fitting the initial track hits with this additional hit, leading to updated track parameters. Finally, an additional track is created, for which no hit is used to account for possible detector inefficiency. This leads to the creation of a fake hit, called *invalid hits* by opposition to real hits called *valid hits*. As the number of tracks increases exponentially at each iteration, only a subset of tracks is kept based on quality requirements involving their number of valid and invalid hits as well as the normalized value of the χ^2 test of the fit. The procedure ends when the last detector layer is reached.

Final Track Fit : constraints applied during the Seed Generation and Pattern Recognition steps could bias the estimate of the track parameters obtained at the end of the Pattern Recognition step. In order to reduce this bias, the trajectory is finally re-estimated with a Kalman filter : a first fit is initialized with the innermost hit and proceeds through the whole list of available hits and a second fit is initialized with the final results of the first fit and runs backward through the list of hits towards the innermost one.

Track Selection : despite the quality requirements applied to the reconstructed tracks, the reconstruction procedure yields a significant amount of tracks which do not correspond to the trajectory of any of the produced charged particles, called *ghost* tracks . In order to reject these fake tracks, they are required to match several minimal quality criteria, based on their track parameters like $\eta,\ p_T$ and the number of crossed layers. Briefly, these cuts become tighter as the number of crossed layers decreases and that the track transverse momentum increases.

Six iterations of the four aforementioned steps are performed with the different seed collections as inputs and at each iteration, hits used to form a valid track after the Track Selection step are removed from the collection of available hits.

4.1.2 Primary vertex reconstruction

Subsequently to the track reconstruction, algorithms are used to find the vertex positions and their associated tracks. In CMS, the vertex reconstruction sequence consists of two steps :

1. Vertex finding [89] : tracks are clustered into primary vertex candidates, based on their separation distance extrapolated to the beam line, along the z direction with a maximal distance of 1 mm between two tracks belonging to the same vertex candidate. In order to reject tracks from secondary vertices, a pre-selection is applied, requiring a track transverse impact parameter significance less than 5. For each primary vertex candidate, an average vertex position is calculated from all the tracks associated to the candidate. The primary vertex is finally chosen as the candidate with the highest sum of the squared transverse momentum values of the associated tracks.
2. Vertex fitting : a dedicated vertex fitting algorithm has been developed in CMS, the Adaptive Vertex Fitter (AVF) [90], which consists in a traditional Kalman filter based vertex fitter, as described in [91], modified to prevent outlying tracks to significantly degrade the vertex fitter performances. The AVF consists in an iterative least-square fit which weights the tracks according to their distance to the vertex.

4.1.3 CMS Tracker performances

The performances of the CMS Tracker have been evaluated with the first data produced at a collision energy of 7 TeV [92]. Using 10nb^{-1} of integrated luminosity, the measured resolutions in x(y) and z of the primary vertex have been found in agreement with Monte-Carlo simulations. These resolutions are of the order of 20 µm and 25 µm respectively for vertices reconstructed with more than 30 tracks. The measured primary reconstruction efficiency is greater than 98.5 % if the vertex is reconstructed with more than two tracks with transverse momenta greater than 0.5 GeV$/c$. In addition, distributions of the track transverse momentum, pseudo-rapidity, transverse and longitudinal impact parameter have been compared to Monte-Carlo simulations. Measurements and Monte-Carlo predictions have been found to in good agreement. Finally, the track reconstruction efficiency and transverse momentum resolution have been measured with resonances decaying into muon pairs. For muons with a transverse momentum greater than 2 GeV$/c$, the measured efficiency is greater than 99.6 % [93]. The relative transverse momentum resolution, measured with muons from J/ψ decays, decrease from 3 % for $|\eta| = 2.4$ to 0.5 % for $|\eta| = 0$ [94].

4.2 Lepton reconstruction

In this section, the algorithms used to reconstruct leptons are reviewed with emphasis on the muon reconstruction as muons, unlike electrons, are part of the topology of interest in this thesis.

4.2.1 Muon reconstruction

The muons are first detected by the CMS inner tracker. As they have no strong interactions and few electro-magnetic interactions, they reach the muon detection system, the outermost component of the CMS detector. Local reconstruction of tracks is first performed independently in the inner tracker and in the muon system. Then, depending on whether the inner tracker informations are used or not and depending on how these tracker and muon system informations are combined, three different type of muons are reconstructed : the so-called *stand-alone*, *global* and *tracker* muons [95].

Stand-alone muon (STA) reconstruction : this type of muons are reconstructed as tracks in the muon system following a two-step procedure :

1. pre-filter : it consists in a coarse inside-out extrapolation of the muon track. A seed is first formed using segments of aligned hits from the same muon chamber. Then, for each seed, the muon trajectory parameters are extrapolated to the innermost compatible muon detector layer. Finally, a muon track is extrapolated in the inside-out direction using a Kalman filter technique, accounting for multiple scattering and energy loss effects as well as the magnetic field inhomogeneities.
2. filter : during this step, the final outside-in track reconstruction is performed, starting from the outermost muon detector layer. The seed consists of the final track parameters obtained with the pre-filter.

During the pre-filter, the extrapolation uses DT or CSC segments or RPC hits while it uses individual hits of all kind during the final filter. Inclusion or rejection of measurements is done on a χ^2 basis ; measurements improving the current track χ^2 are included and the trajectory parameters updated accordingly. If no compatible measurement is found, the trajectory is propagated to the next muon detector layer. A stand-alone muon track is accepted if it contains at least two measurements, one of them being either of DT or CSC type. Finally, the track is extrapolated to the interaction point and the track parameters are updated accordingly.

Global muon (GLB) reconstruction : in this reconstruction procedure, informations from the muon system and the inner tracker are combined. For a given STA muon, it searches for the most compatible tracker track in two successive steps :

1. track matching : this step consists in identifying tracker tracks compatible with a given STA muon track. In order to limit the number of tracker track candidates and therefore to speed-up the procedure, a so-called *region of interest* (ROI) is defined for each STA muon. This region is a rectangle in the $\eta - \phi$ space, centred around the primary vertex, pointing in the direction of the STA muon, with a size depending on the estimated uncertainties of the reconstructed STA muon direction. The subset of tracker tracks contained in this region of interest are then considered as compatible tracker tracks if the transverse momentum ratio between the track and the STA muon exceed a certain threshold. Finally, a comparison of the track parameters is made between the selected tracker tracks and the STA muon track, using the track position and momentum, propagated at a so-called *common matching surface*, which is chosen as to minimize at the same time the errors on the propagated parameters and the number of matches.
2. global fit : after the selection of matched tracker tracks for each STA muon, a global fit is attempted for each tracker track - STA muon track pair, using their hits. If several global tracks result from this fit, the one with the best χ^2 is chosen so that to a given STA muon corresponds at most one global muon.

Tracker muons (TRK) reconstruction : because of their bending in the magnetic field and of energy losses in the detector material, muons must have a minimal transverse momentum of 4.8 GeV$/c$, 3.6 GeV$/c$ and 1.2 GeV$/c$ in order to reach the first muon station for $|\eta| < 1.2$, $1.2 < |\eta| < 1.5$ and $1.5 < |\eta| < .2.4$ respectively. As a result, low transverse momentum muons may not leave enough hits in the muon system to be reconstructed as stand-alone muons and therefore as global muons. This is unfortunate because the best achievable momentum resolution is almost already obtained when the muon reaches the first muon station. In order to avoid this issue, an *inside-out* muon reconstruction procedure has been developed which starts with tracker tracks and then search for matching segments in the muon system. This procedure can be divided in two steps :

1. Propagation : each tracker track with a momentum above a given threshold is propagated outward into the calorimeters and then into the muon system, taking into account the magnetic field and energy losses. Possible multiple scattering are taken into account in the estimated uncertainties on the propagated trajectory parameters. Each segment contained in the muon chambers that are crossed by the extrapolated track is associated to the track.
2. Arbitration : one or more segments in the muon system can be associated to several tracker tracks if they are close enough to each other. The arbitration step is meant to solve this ambiguity ; currently two different arbitration algorithms are available, referred to as the DxArbitration and the DrArbitration algorithms. When a segment is associated to more than one track, these algorithms calculate, for each tracker track, the distance between the segment and the extrapolated tracker track in local X coordinate, ΔX and the distance in local X and Y coordinates, $\Delta R = \sqrt{\Delta X^2 + \Delta Y^2}$ inside the chamber. The segment is then assigned to the track having the minimal distance. Therefore, tracker muons are not unambiguously defined but depend on the arbitration algorithm used.

In order to be reconstructed as a tracker muon, a muon must have a minimal transverse momentum value of 1.5 GeV/c and have an extrapolated track matched with at least one segment with either $\Delta X < 4\sigma$ or $\Delta X < 3$ cm.

4.2.2 Electron reconstruction

An electron is identified by a track reconstructed in the tracker, matched to an electromagnetic shower developing in the electro-magnetic calorimeter and leaving energy deposits in a few neighbouring crystals. Bremsstrahlung radiations, due to interactions with the material in the tracker and in front of the electromagnetic calorimeter, leads to additional electron energy deposits which are spread in ϕ by the magnetic field. These energy deposits associated to a single electron are then grouped into a supercluster (SC). In CMS, two complementary approaches have been developed to reconstruct electrons [96] ; the first one makes use of SC to seed the search for compatible tracks in the inner tracker. This approach is particularly well-adapted to electrons with transverse momentum greater than $5 - 10$ GeV/c for which the bremsstrahlung radiations are collinear to the initial electron trajectory. At lower value of transverse momentum, radiations may create energy deposits which are well-separated from the electron energy cluster and therefore difficult to efficiently associate with the correct electron. Moreover, in the case of electrons contained in a jet of particles originating from quark or gluon hadronization, the SC may include additional energy from the other jet constituents, biasing the SC as an electron seed. In order to limit these effects, an alternative reconstruction approach has been considered which starts with track segments in the inner tracker and search for compatible SC in the ECAL. Starting from these seeds, the electron tracks are then built from tracker hits collected with a dedicated algorithm based on the combinatorial Kalman filter and taking into account bremsstrahlung energy losses. Electron trajectories are fitted with a so-called Gaussian Sum Filter (GSF) [97] algorithm which is a non-linear generalization of the Kalman filter, meant to deal with non-Gaussian fluctuations of the electron trajectory parameters due to bremsstrahlung energy loss. Finally, quality requirements on the track-SC matching are applied on electron candidates, formed by the association of a GSF track and its associated SC.

4.2.3 Lepton reconstruction performances

The muon and electron reconstruction performances have been evaluated with the data collected by the CMS detector in 2010. Lepton transverse momentum and pseudo-rapidity distributions have be found to be in agreement with Monte-Carlo predictions within $10\ \%$. Using events with a Z boson decaying into a pair of electron or muons, the lepton reconstruction efficiency has been measured ; in case of electrons, it reaches $(99.3 \pm 1.4)\ \%$ in the ECAL barrel and $(96.8 \pm 3.4)\ \%$ in the ECAL end-caps [98]. For muons, it exceeds $99.2\ \%$ over the entire detector acceptance range, $|\eta| < 2.4$ and the relative transverse momentum resolution remains below $2\ \%$ [99].

4.3 Jet reconstruction

As mentioned in section 2.2.2, quarks and gluon produced during hard interactions fragment and then hadronize, leading to collimated jets of hadrons. The purpose of jet reconstruction is to recover the kinematic properties of these partons by clustering hadrons that are likely to come from the same parton into a single object.

4.3.1 Jet reconstruction inputs

The CMS detector provides redundant informations about the particles produced during proton collisions. Based on which sub-detector parts are used and how these informations are combined, three different strategies have been developed in CMS to determine the kinematic variables of the various hadrons participating to the jet reconstruction, leading in the end to three different types of jets :

Inputs to calorimeter jets (CaloJets) [86] : only informations from the calorimeters are used to reconstruct this type of jets. The signals of the cells are combined into so-called *calorimeter towers*. These towers are treated as massless particles with the energy given by the tower energy and the direction by the interaction point and the centre of the tower. They consist of one or more HCAL cell and the geometrically corresponding ECAL crystals. In the barrel region of the CMS calorimeter ($|\eta| < 1.4$), a single HCAL cell matches a 5×5 matrix of ECAL crystals. In order to reject electronic noise, calorimeter energy deposits are only added to the corresponding tower energy if they exceed a given threshold and an energy cut is applied on each tower (Tab. 4.1). Finally, in order to reject towers created by the energy deposits of particles originating from pile-up events, only towers with a transverse energy greater than 0.3 GeV are considered for further clustering into jets. This leads to an offset in the reconstructed jet energy which can be corrected later on, as explained in section 4.3.4.

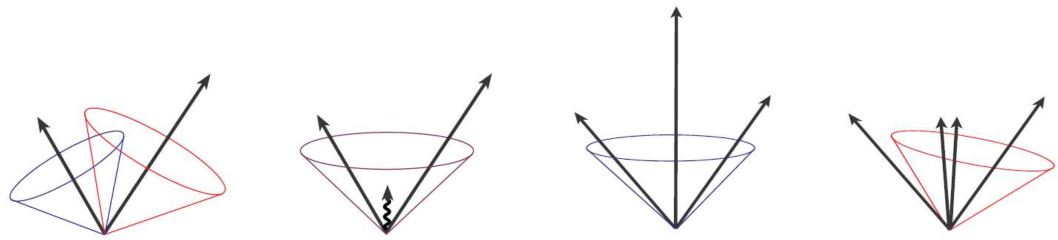
Calorimeter subsystem		Thresholds [GeV]
Electromagnetic calorimeter		
	Barrel	$0.07/0.2$ (per crystal/per tower)
	End-cap	$0.3/0.45$ (per crystal/per tower)
Hadron calorimeter		
	Inner Barrel	0.7
	Outer Barrel	$1.1/3.5$ (Ring 0/Ring $1,2$)
	End-cap	0.8
	Forward	$0.5/0.85$ (Short/Long fiber readout)

Table 4.1: Energy thresholds applied on calorimeter cells in a calorimeter tower [100].

Inputs to track-corrected calorimeter jets (Jet-Plus-Tracks) [101] : this jet reconstruction strategy is an extension of the calorimeter-based strategy which takes advantage of the high precision CMS Tracker to compensate for the non-linear response of the CMS hadron calorimeter. Charged particle tracks are associated to the reconstructed calorimeter jets with a jet algorithm-dependent scheme based on their distance in the $\eta - \phi$ plane. The tracks are then extrapolated onto the inner calorimeter surface and divided into so-called *in-cone* and *out-of-cone* tracks whether or not they are contained in the jet area. The momenta of both categories of tracks are added to the jet energy and the expected energy deposits of in-cone tracks are removed from the total jet energy.

Inputs to particle-flow jets (PFJet) [102] : the particle-flow event reconstruction approach aims at providing a coherent description of the full event by combining all the CMS sub-detector in order to reconstruct and identify individually all particles produced during the proton-proton collisions. These reconstructed particles, called particle-flow objects, are used as inputs for the jet reconstruction algorithms. Elements combining the informations of one or more subdetector, such as tracks and calorimeter clusters for example, are first reconstructed by specific algorithms. Then elements arising from the same particle are linked together into blocks of elements which are further used to identify every single particle in the event in a sequential way. Muons and electrons are first reconstructed and their associated blocks are removed from the list of the available blocks. Matching pairs of remaining tracks and calorimeter clusters are then used to identify charged hadrons. Unassociated calorimeter clusters are finally used to also identify neutral particles. In the case of neutral hadrons, they usually deposit energies in both the ECAL and the HCAL. Because of the non-linear response of the HCAL and because of the difference in ECAL response between photons and neutral hadrons, it is necessary to calibrate the calorimeter cluster energies although neutral hadrons account only for 10% of the measured event energy on average. Therefore, PFJets are expected to provide an accurate description of the kinematic properties of the original partons.

In addition to these three types of jets, another type of jets, called *generator jets* or *particle jets* (GenJet) is defined when using Monte-Carlo simulated data. These jets take as inputs the four-momentum of all the stable particles, except the neutrinos, generated during the final step of the event simulation.



(a) **(b)**

Figure 4.1: Sketch of jet reconstruction results in case of soft particle emission (a) and particle collinear splitting (b) when ignored or undetected (left) and when properly taken into account (right) [103].

4.3.2 Jet reconstruction algorithms

At theoretical level, jet reconstruction algorithms are required to be infra-red and collinear (IRC) safe. This means that any given jet configuration has to remain unchanged by emissions of additional soft particles or by the collinear splitting of particles, in order to preserve the cancellation of the infra-red divergences between real and virtual higher order QCD corrections. But IRC safe jet algorithms are also required to meet experimental constraints as detectors do not resolve energy deposits of collinear particles and energy thresholds suppress deposits from soft particles (Fig. 4.1). Furthermore, jet reconstruction deals with high multiplicity final states, several hundreds of particles on average result from proton collisions, and therefore has to rely on fast and robust algorithms. They also need to be efficient enough to provide similar results whether the inputs are particles or energy deposits in a detector calorimeter. Nowadays, there are two main classes of jet algorithms :

Cone jet algorithm : historically, the concept of cone jet was introduced by G. Sterman and S. Weinberg [104] in 1977. This concept is quite naturally inspired by the idea of conical energy flow resulting from successive parton branching occurring with smaller and smaller opening angle. Starting with the particle with the highest transverse momentum, particles present in a cone centred at this particle are merged and the resulting four-momentum is used to define the direction of a new trial cone. The procedure is iterated until the cone direction coincides with the sum of the particle four-momenta it contains. The particles associated to this stable cone are removed from the list of particles available for further jet clustering. This procedure is repeated until all particles have been clustered. Several versions of the cone jet algorithm exist, based on how particles in a jet are merged and/or overlapping jets are resolved.

Although it has been advocated for some years now that most of the cone jet algorithms in use were infrared and collinear unsafe, such algorithms are still used because of their simplicity and their speed.

Sequential recombination jet algorithm : the basic idea is to introduce a measure of the distance between the entities (particles, calorimeter towers, ...) and merge the closest entities iteratively until some stopping criteria is fulfilled. The main difference between the algorithms currently used within the CMS collaboration concerns the measurement of the distance two between entities, $d_{i,j}$ and between an entity and the beam direction, $d_{i,B}$:

$$d_{i,j} = min(k_{T,i}^{2p}, k_{T,j}^{2p})\frac{\Delta R_{i,j}^2}{R^2} \quad (4.1)$$

$$d_{i,B} = k_{T,i}^{2p} \quad (4.2)$$

where k_T is the transverse momentum of the entity i, R is a free parameter and $\Delta R_{i,j}$ is the distance in the rapidity (y)-azimuth (ϕ) plane between the entities i and j. The case $p = 1$ corresponds to the so-called k_T algorithm [105, 106]. Quite recently, a new algorithm, the anti-k_T algorithm [107], has been proposed for which $p = -1$ and is currently used within the CMS collaboration as one of the default jet algorithm.

In this thesis, both the k_T and the anti-k_T algorithms are used to reconstruct jets. Conceptually, theses algorithms can be described as following :

k_T/anti-k_T algorithm :

1. For each entity i and each couple of entities i and j, determine the values of d_i and of $d_{i,j}$ respectively.
2. Find the smallest of all the d_i and $d_{i,j}$. Label it d_{min}.
3. If $d_{min} = d_{i,j}$, merge the entities i and j by adding their four-momenta.
   If $d_{min} = d_i$, the entity is considered as a jet and removed from the list of entities.
4. Return to step 1.

The procedure is iterated until no entity is left.

4.3.3 Identification of jets from heavy quarks

During the hadronization process, the production of heavy quarks is strongly suppressed, as explained in section 2.2.2. As the fragmentation of b and c quarks produces B- and C-hadrons respectively, it is possible, to some extent, to identify jets originating from heavy

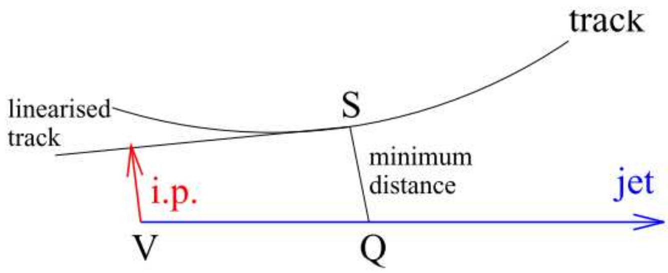


Figure 4.2: Schematic representation of the impact parameter (IP) of a track with respect the jet vertex (V), the decay length (VQ segment) and the track to jet distance (SQ segment) [109].

quarks thanks to the presence of these heavy long-lived hadrons ; in the case of B-hadrons decaying weakly, they have on average a mean life-time of 1.5×10^{-12} s for an energy of ~ 5 GeV [108] which allows them to travel on average 500 µm before they decay, producing a displaced secondary vertex. Tracks of the decay products are thus expected not to be compatible with the primary vertex of the jet. For a given track, this incompatibility is measured by a Lorentz invariant observable called the *impact parameter* (IP) which is defined as the minimal signed distance between the primary vertex and the track, linearised from the point of closest approach to the jet direction (Fig. 4.2). The sign is calculated as the sign of the scalar product between the jet direction and the IP segment, oriented from the vertex. Therefore, it is positive if the track originates from a decay downstream from the jet vertex. Tracks produced by the decay products of long-lived hadrons are expected to have a larger positive IP compared to other tracks. But as the uncertainty on this quantity vary for each track, the impact parameter significance (IPsig), defined as the IP divided by its error, is often used instead. In CMS, the IP can be measured either in the transverse plane or in three dimensions. Furthermore, B-hadrons decay semi-leptonically with a branching fraction of 11 %, which increases up to 20 % if the $b \to c$ cascade is included. Then, the resulting jets contain a charged lepton with a transverse momentum of a few GeV$/c$ with respect to the jet axis which can be efficiently detected and reconstructed, mainly in the case of muons. Exploiting these characteristics, dedicated algorithms have been developed to identify jets originating from b quark hadronization, called b-jets. Such algorithms are called *b-tagging algorithms*. They associate each jet with a discriminator value. If this value is above a certain threshold, to be defined by the user, the jet is *b-tagged*. However, these algorithms are not fully efficient and it happens that b-jets are not identified as such or that non b-jets are mis-identified as coming from b quarks. Therefore, for a given threshold value on the b-tag discriminator, the performances of a b-tagging algorithm is usually evaluated in terms of :

- the b-tagging efficiency : the probability for a jet to be b-tagged, given that it is indeed a b-jet
- the mis-tagging efficiency : the probability for a jet to be b-tagged, given that it is **not** a b-jet

Hadrons originating from the hadronization of a c quark tend, due to their mass, to mimic the behaviour of B-hadrons, resulting in a higher mis-tagging efficiency for c-jets compared to light jets. That is why the mis-tagging efficiency is often quoted separately for jets originating from c quarks and from light (u,d,s) quarks. Finally, the mis-tagging efficiency for gluon-induced jets is comparable to the one for light quarks and both efficiencies are usually merged. For each algorithm, standard thresholds on the b-tag discriminator values,called *working points*, are established. The loose (L), medium (M) and tight (T) working points correspond to a light jet mis-tagging efficiency of $0.1\ \%$, $1\ \%$ and $10\ \%$ respectively.

Inputs for B-tagging algorithms

As mentioned previously, the b-jet identification heavily relies on the impact parameter of displaced tracks and the presence of secondary vertices. Therefore, to ensure a sufficient precision on their determination, additional requirements on tracks and secondary vertices are applied compared to the usual quality criteria required by the standard track and vertex reconstruction algorithms. Selected tracks must have [109] :

- at least eight tracker hits (Pixel+Strip),
- at least two pixel hits to allow a precise extrapolation close to the primary vertex,
- a transverse momentum greater than 1 GeV$/c$,
- a normalized χ^2 value of the track fit smaller than 5, in order to reject badly reconstructed tracks.

Furthermore, tracks associated to a jet are required to have :

- a transverse impact parameter, d_0, smaller than 0.2 cm
- a longitudinal impact parameter[2], d_z, smaller than 17 cm
- a distance, ΔR, to the jet axis in the $\eta - \phi$ space smaller than 0.5

[2]The longitudinal impact parameter is the component along the beam-line of the distance of closest approach.

Then, for the tracks associated to a jet, the reconstruction of secondary vertices from B hadron decays is performed using the adaptive vertex fitter described in section 4.1.2. Any reconstructed vertex that shares more than 65% of the tracks associated to the primary vertex is rejected. Finally, the compatibility of the reconstructed vertices with the hypothesis that they correspond to a B hadron decay is checked. Selected vertices must have :

- a distance to the beam-spot in the transverse plane smaller than 2.5 cm
- a distance to the jet axis in the $\eta - \phi$ space smaller than 0.5
- a flight distance significance in the transverse plane with respect to the primary vertex smaller than 3.0 and a maximal flight distance of 0.1 mm
- a mass[3] incompatible with the mass of a K^0_S within a window of 50 MeV$/c^2$.

If more than one secondary vertex is found, the one with the smallest error on the flight distance is chosen as the best vertex. Secondary vertices obtained from a successful fit are called *RecoVertex*. If no secondary vertex is found but at least two tracks with a transverse impact parameter significance greater than 2.0 are present, a fallback *PseudoVertex* is formed without any explicit fit.

B-tagging algorithms

Several b-tagging algorithms are available in CMS and the ones used in this thesis are documented in what follows :

- Track Counting algorithm (TC) : this algorithm returns as a discriminator value the impact parameter significance of the N^{th} track, ordered in decreasing significance. Two versions of this algorithm are available : the *Track Counting High Efficiency* (TCHE) algorithm for which $N = 2$ and the *Track Counting High Purity* (TCHP) for which $N = 3$.
- Jet Probability (JP) : this algorithm is based on the probability for the jet tracks to be compatible with the jet vertex. This probability, P_{tr}, is obtained from the negative part of the signed impact parameter significance, IP_σ, distribution which is mainly due to tracks compatible with the jet vertex or fake tracks. For these tracks, the distribution is symmetric around zero. Hence, it is possible to derive from data an unsigned IP

[3]The vertex mass is defined as the invariant mass of the system of tracks associated to the vertex.

significance distribution R. The signed track probability is then defined as :

$$P_{tr} = \frac{|IP_\sigma|}{IP_\sigma} \int_{|IP_\sigma|}^{+\infty} R(x)dx \tag{4.3}$$

Finally the jet probability, P_{jet}, for all tracks within a jet is given by :

$$P_{jet} = \Pi \sum_{j=0}^{N-1} \frac{(-ln(\Pi))^j}{j!} \tag{4.4}$$

where $\Pi = \prod_{i=1}^{N} \tilde{P}_{tr}(i)$ and $\tilde{P}_{tr}$ is the redefined track probability : $\tilde{P}_{tr} = P_{tr}/2$ for $P_{tr} > 0$ and $\tilde{P}_{tr} = 1 + P_{tr}/2$ for $P_{tr} < 0$. In order to prevent the combined probability to be driven low by badly reconstructed tracks with extremely small probabilities, a lower cut-off of 0.05 is applied to the track probability. Unfortunately, this cut-off introduces saturation peaks in the distribution of final discriminator value D defined as :

$$D = -ln(P_{jet})/4 \tag{4.5}$$

Another version of this algorithm, called *Jet B Probability*, makes a special use of the four most displaced tracks, as on average the charged track multiplicity of B-hadrons is around 5. In this algorithm, D is defined as :

$$D = -\left(ln(P_{jet}^{all}) - ln(P_{jet}^{4\ tracks})\right)/4 \tag{4.6}$$

where $P_{jet}^{4\ tracks}$ is computed with the first four tracks having the lowest probability and a positive IP significance. If less than four such tracks are available, all tracks with positive IP significance are used.

- Simple Secondary Vertex (SSV) : the discriminator value returned by this algorithm is a function of the flight distance in three dimensions D_{3D} between the jet vertex and the secondary vertex. If no secondary vertex is found, it does not return any value, limiting its efficiency to the efficiency of finding such a vertex. The discriminator value D is defined as :

$$D = ln\left(1 + \frac{D_{3D}}{\sigma_{D_{3D}}}\right) \tag{4.7}$$

Two versions of this algorithm exist, based on the number of tracks, N_{trk}, associated to the secondary vertex ; $N_{trk} \geq 2$ for the high efficiency version (SSVE) and $N_{trk} \geq 3$

for the high purity version (SSVP). At the moment of the writing, only the SSVE was available and therefore used in this thesis.

The performances of these algorithms have been studied with the first data collected in 2010 by the CMS detector during proton-proton collisions at a centre-of-mass energy of 7 TeV. The discriminator distributions, obtained with the data, have been found to be in good agreement with the simulated distributions for the various b-tagging algorithms (Fig. 4.3). This agreement indicates the readiness of the b-jet identification with the CMS detector.

4.3.4 Jet energy correction scheme in CMS

Once a jet has been reconstructed, it is still not straightforward to deduce the parton energy from the measured jet energy for several reasons. A first one is related to the concept of jets itself ; large angle gluon emissions lead to separate jets that can be reconstructed outside the jet produced by the parton after emission. Another source of effects is related to the experimental conditions of the proton collisions and to detector inefficiencies ; underlying and pile-up events produce energy deposits that may be added to the jet energy although they come from particles that are not part of the hard event of interest. Furthermore, low-momentum particles are deviated from the jet direction by the magnetic field and could therefore not be clustered into the appropriate jet. Detector related effects arise mainly from electronic noise which artificially increases the jet energy. Finally, additional effects stem from the non-linear response of the hadron calorimeter. To correct for these effects and thus calibrate the jet energy scale (JES), dedicated correction factors, derived either from Monte-Carlo simulations or in a data-driven way, are successively applied (Fig. 4.4). This approach allows to calibrate the jets either at the particle level or directly at the parton level. Within the CMS collaboration, JES correction factors are produced for the different jet reconstruction algorithms [110].

Level 1 (Offset) : thanks to calorimeter cell and tower thresholds (see Tab. 4.1), the jet energy offset due to the electronic noise is found to be negligible and each pile-up event is expected to yield a transverse energy offset of 0.1 GeV to 0.3 GeV. This offset is estimated from data with zero-bias and minimum-bias events. This offset correction is subtracted from the reconstructed jet energy.

Level 2 (Rel. η) : due to the non-uniform response of the CMS calorimeters, the jet response, defined as the ratio of the reconstructed jet transverse momentum to the generated one, varies with the pseudo-rapidity, η. For particle jets with a low

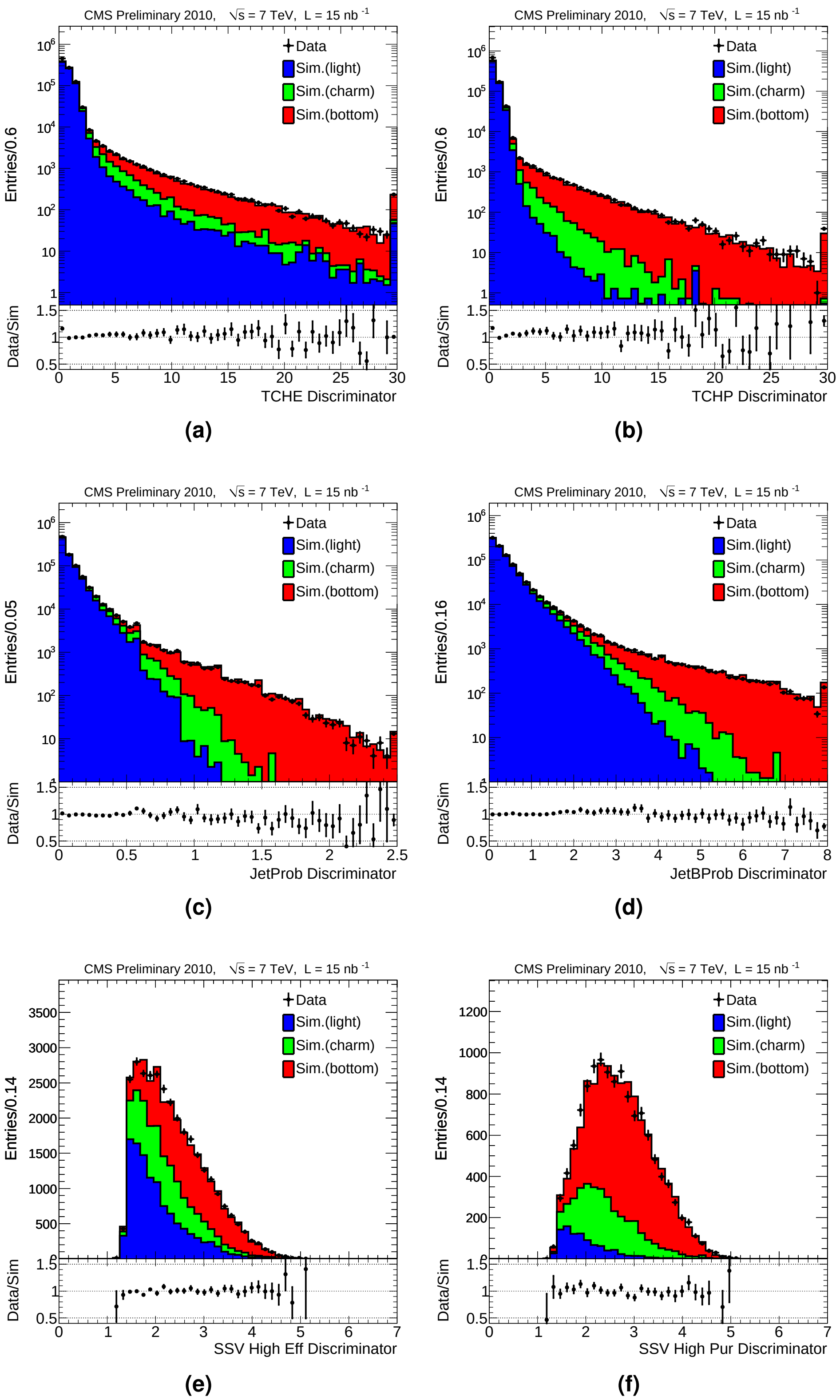
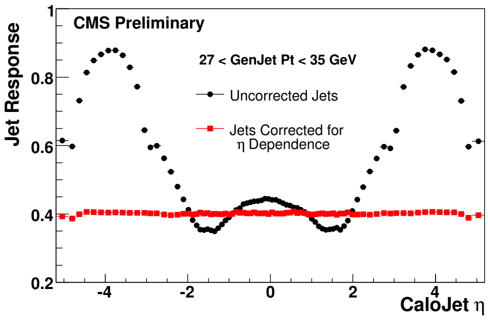
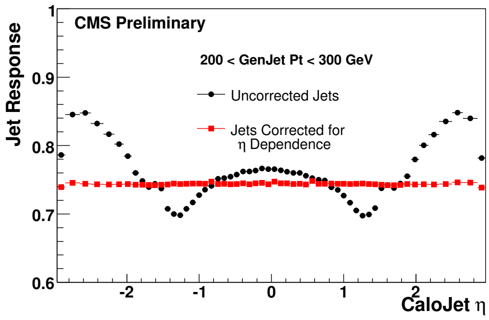


Figure 4.3: Distributions observed in data, superimposed to distributions obtained with Monte-Carlo simulations for several b-tagging algorithms : Track Counting (high efficiency : a, high purity : b), Jet Probability (c), Jet B-probability (d) and Simple Secondary Vertex (high efficiency : e, high purity : f) [109].

Figure 4.4: Schematic picture of the succession of the mandatory jet correction levels, shown in solid boxes and of the optional ones, shown in dashed boxes [110].



(a)



(b)

Figure 4.5: Jet response as a function the reconstructed calorimeter jet pseudo-rapidity (CaloJet η) derived from Monte-Carlo simulations for particle jets (GenJets) with $27 < p_T < 35$ GeV/c (a) and $200 < p_T < 300$ GeV/c (b) both before and after level 2 corrections [110].

transverse momentum, the jet response, derived from Monte-Carlo simulations, range from 0.6 to 0.9 in the forward regions ($3 < |\eta| < 5$) and decreases down to $0.35 - 0.45$ in the barrel regions ($|\eta| < 1.4$) (Fig. 4.5a). For particle jets with a high transverse momentum ($200 < p_T < 300$ GeV/c), the response varies between 0.7 and 0.85 (Fig. 4.5b). Corrections for this η-dependence are either derived from Monte-Carlo simulations or from collision data with the so-called *di-jet transverse momentum balance* technique which proved to be successful at the Tevatron experiments [111]. This technique consists in using the expected transverse momentum balance in back-to-back di-jet events with one of the jets in barrel region being the reference and the other jet at any arbitrary η being the probe. This leads to a constant corrected response corresponding to the average response in the barrel region.

Level 3 (Abs. p_T) : the CMS hadron calorimeter is a non-linear sampling calorimeter and, therefore, the jet response is smaller than unity and varies as a function of the jet transverse momentum (Fig 4.6a). The correction factors accounting for these effects, obtained from Monte-Carlo simulations, range from 1.9 down to 1.5 for a reconstructed jet transverse momentum varying between 30 GeV/c and 100 GeV/c (Fig 4.6b). Correction factors can also be derived from the transverse momentum balance expected in $\gamma + jets$ and $(Z \rightarrow \mu^+\mu^-) + jets$ events, making therefore use of two independent

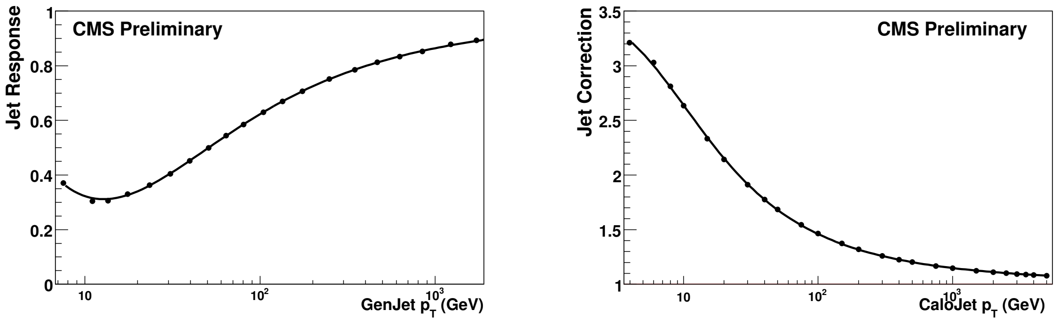


Figure 4.6: Monte-Carlo simulated jet response as a function of the transverse momentum of the particle jet, p_T^{GenJet} (a) and simulated correction factors as a function of the reconstructed calorimeter jet (b), CaloJet p_T [110].

sub-detector parts, the calorimeter and the muon system respectively. This technique has been successfully used at the Tevatron experiments [112].

Level 4 (EMF) : these corrections account for the differences in jet response because of the non-compensating feature of the CMS hadron calorimeter. They depend on the fraction of jet energy measured in the electromagnetic calorimeter. It has been shown that the use of these corrections would improve the resolution up to 10%.

Level 5 (Flavour dependence) : flavour-dependent variations in the jet energy response arise from differences in the jet fragmentation between light quarks and gluons. Jets from heavy quarks contain also on average more charged hadrons than jets from light quarks. Furthermore, b- and c-flavoured hadrons undergo semi-leptonic decays leading to the production of neutrinos whose energies cannot be measured. These correction factors have been derived from Monte-Carlo simulations so far but it has been shown that they could be derived from data using events where top quark pairs are produced [113].

Level 6 (Underlying events) : corrections intended to remove the contributions from underlying events to the reconstructed jet energy are not yet available.

Level 7 (Parton correction) : these corrections are meant to be applied on jets already corrected at particle level. They provide the needed additional corrections to go from the particle level to the parton level. Therefore, they correct for effects from the parton shower and jet hadronization processes. Although applied on already flavour-corrected jets, they have been made flavour specific, accounting for the residual corrections needed for heavy-flavour quarks. These corrections are derived from Monte-Carlo simulations [114].

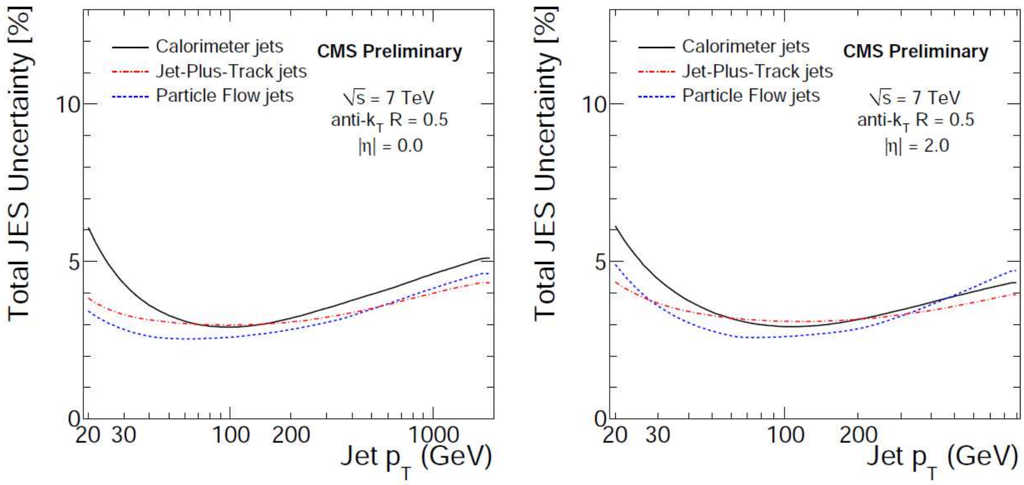


Figure 4.7: Total jet energy scale uncertainty measured from data, as a function of the jet transverse momentum for different $|\eta|$ values and for different types of jets [115].

The first three levels are mandatory while the others remain optional and are analysis-dependent. Level 1 to 3 corrections allow to recover the jet kinematic properties to the particle level. These corrections are applied as following :

$$E_{jet}^{L1L2L3corr} = (E_{jet}^{uncorr} - E_{L1}) \times C_{L2}(\eta_{jet}^{uncorr}, p_T^{L1corr}) \times C_{L3}(p_T^{L1L2corr}) \tag{4.8}$$

where E_{L1} is level 1 offset energy and C_{LX} is the level X correction factor which has to be evaluated for level $X-1$ corrected jets.

4.3.5 Jet reconstruction performances

The total uncertainty on the jet energy scale, combining the relative scale (L2) and relative scale (L3) uncertainties, has measured by the CMS experiment using data collected during proton collisions at a centre-of-mass energy of 7 TeV. Using 3 pb^{-1} of integrated luminosity, data-driven calibration techniques, mentioned in the previous section, have been performed [115]. As shown in Figure 4.7, the total jet energy scale uncertainty ranges between 4 % to 6 % for reconstructed calorimeter jets with a transverse momentum greater than 20 GeV$/c$.

4.4 Missing transverse energy

Weakly interacting particles like neutrinos escape detection. By assuming that the colliding partons during proton collisions do no carry any transverse momentum, it becomes possible to infer the presence of such particles from the imbalance of the event total transverse momentum. This imbalance is the so-called *transverse missing momentum*, whose magnitude is referred to as the missing transverse energy (MET), also denoted $\not{E}_T$ or E_T^{miss}. As explained in the previous section, the reconstructed jet energy needs to be corrected. These corrections being dependent on the type of reconstructed jet, three different algorithms have been developed to reconstruct the missing transverse energy, one for each jet type available in CMS : CaloJet, JPT or PFJets.

Calorimeter MET (CaloMET) [116] : the calorimeter based MET is derived from the vector sum of the energy deposits in the calorimeter towers projected in the transverse plane as following :

$$\vec{\not{E}}_T = -\sum_i^{towers} \vec{E}_{T,i} = -\sum_i^{towers} (E_i sin(\theta_i) cos(\phi_i)\vec{n}_x + E_i sin(\theta_i) sin(\phi_i)\vec{n}_y) \quad (4.9)$$

where E_i is the energy deposited in the calorimeter tower i with polar and azimuthal coordinates (θ_i, ϕ_i). The unit vectors $\vec{n}_x$ and $\vec{n}_y$ correspond to the x and y directions perpendicular the beam direction (z). In CMS, the MET corrections take into account the jet energy scale factors and the energy deposited in the calorimeter by the muons :

$$\vec{\not{E}}_T = -\sum_i^{towers} \vec{E}_{T,i} - \sum_j^{jets} \left[\vec{p}_{T,j}^{\,corr.} - \vec{p}_{T,j}^{\,uncorr.}\right] - \sum_k^{muons} \left[\vec{p}_{T,k}^{\,\mu} - \vec{E}_{T,k}^{\,deposit}\right] \quad (4.10)$$

where $\vec{p}_T^{\,corr.}$ and $\vec{p}_T^{\,uncorr.}$ are respectively the corrected and uncorrected jet transverse momentum vector; $\vec{p}_T^{\,\mu}$ and $\vec{E}_T^{\,deposit}$ are respectively the muon transverse momentum vector and it associated calorimeter energy deposit, which is usually of the order of a few GeV.

Track-corrected MET (tcMET) [117] : in this approach, additional corrections are derived from the tracks reconstructed in the tracker. Starting from the calorimeter MET, only corrected for muons, the expected energy deposit in the calorimeter of each tracker track, assumed to be pions, is replaced by the track momentum at the vertex. Only tracks that do not match any reconstructed electron or muon and pass loose quality requirements are used.

Particle-flow MET (PfMET) [102] : in the particle-flow approach, the full event is reconstructed in terms of muons, electrons and photons, as well as in terms of charged and neutral hadrons. Therefore, the PfMET is simply derived from a sum of transverse momentum vector over all particle-flow particles reconstructed in the event.

5

Physics in the top quark sector

The discovery in 1995 of the top quark [118, 119] by the Tevatron experiments, CDF and DØ at Fermilab, within the mass range of the expected weak isospin partner of the bottom quark achieved to complete the quark sector of the Standard Model. It was at the same time another proof of the predictive power of the Standard Model, as well as a proof of its internal consistency. At that stage, the top quark might just appear as any other quark. But the Standard Model does not predict the mass of the particles and from the point of view of its mass, the top quark is really a special quark ; being roughly 40, 120, 1700 and 35000 times heavier than the b, c, s and (u, d) quarks respectively, the top quark is as heavy as a gold atom and is therefore by far the heaviest elementary particle to date. This particularity provides it with unique features compared to the other quarks. It decays before it has the time to form bound-states like any other quark. This gives the unique opportunity to study the spin polarization of a quark at production, which is otherwise depolarized by strong interactions within the bound-states. Furthermore, a precise measurement of its mass would provide stringent constraints on the mass of the so far elusive Higgs boson. Its large mass might also suggest a special role in the electro-weak symmetry breaking mechanism : alternatives to the Higgs mechanism, like dynamical electro-weak symmetry breaking mechanisms often involve the top quark : TopColour [120], TopColour-assisted

Technicolour [121, 122]. Finally, the top quark properties are interesting by their own, as their precise measurements would pursue the validation of the Standard Model as a predictive theory in the top quark sector. But measurements could also reveal some discrepancies with the expectations and shed some light on some potential new physics beyond the Standard Model. If one thinks about the Standard Model as a low-energy approximation of a more general theory, it is reasonable to expect that deviations from the Standard Model predictions are more likely visible in the top quark sector, due to the large top quark mass. In addition, in the context of direct searches for evidences of new physics, the interest for the top quark production process is twofold : additional production channels, like the decay of a new heavy gauge boson to a top quark pair [123], are predicted by several theoretical models for new physics and in general, the top quark production is one of the major background processes for signals of new physics.

In section 5.1, a review of some of the measurements made by the Tevatron and, for the first time, the LHC experiments is given. Section 5.2 presents an overview of the most relevant background processes for top quark studies. Then, the list of the Monte-Carlo generated samples used in this thesis to simulate the top quark pair production as well the background processes is given in section 5.3. Finally, section 5.4 comprises a detailed description of the strategy used to select events where top quark pairs have been produced and reject background events.

5.1 Top quark physics

In this section, we shall review some of the top quark properties and then explicit its production and decay modes within the Standard Model.

5.1.1 Top quark properties

Mass

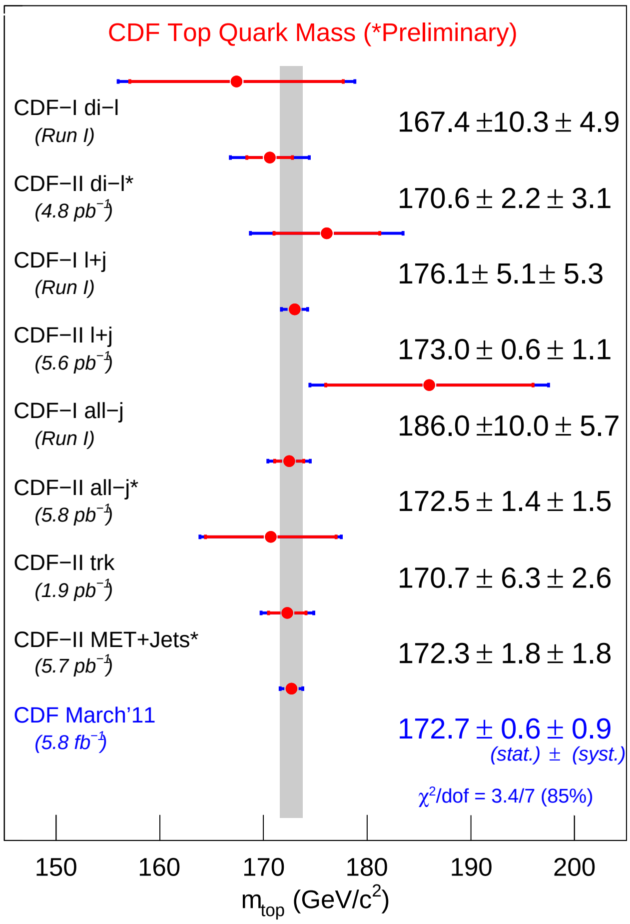
Since the Standard Model does not predict the mass of the quarks, this property has to be determined experimentally. Since the top quark discovery, its mass has been measured by the CDF (Fig. 5.1a) and DØ (Fig. 5.1b) experiments. The latest CDF preliminary result [124] yields a mass of 172.7 ± 0.6 (stat.) $\pm\, 0.9$ (syst.) GeV$/c^2 = 172.7 \pm 1.1$ GeV$/c^2$, corresponding to a precision of 0.6 %. This measurement has been performed using data corresponding to an integrated luminosity of 5.8 fb^{-1}, decreasing the statistical uncertainty (stat.) below 1 GeV and below the systematic uncertainties (syst.). The latest measurements made by the DØ experiment yields a slightly higher top quark mass : 174.9 ± 0.8 (stat.) $\pm$ 1.2 (syst.) GeV$/c^2$ $= 174.9 \pm 1.5$ GeV$/c^2$ [125] using events with a lepton and multple jets corresponding to 3.6 fb^{-1} of integrated luminosity and 174.0 ± 1.8 (stat.) $\pm$ 2.4 (syst.) GeV$/c^2$ $= 174.0 \pm 3.0$ GeV$/c^2$ [126] using events with two leptons corresponding to 3.6 fb^{-1} of integrated luminosity. But the Tevatron accelerator has already delivered and recorded on tape more than 8 fb^{-1} and the current plan is to run the machine until the end of the year 2011, hopefully reaching 10 fb^{-1}. Therefore, most efforts from the Tevatron experiments are presently devoted to the understanding of the systematic uncertainties. For the first time, the top quark mass has also been measured by the LHC experiments, ATLAS and CMS . Using proton-proton collision data collected during the year 2010, representing $\sim$ 36 pb^{-1} of integrated luminosity, these measurements yield a mass of 169.3 ± 4.0 (stat.) $\pm$ 4.9 (syst.) GeV$/c^2$ $= 169.3 \pm 6.3$ GeV$/c^2$ for the ATLAS collaboration [129] and 173.1 ± 2.1 (stat.) ± 2.8 (syst.) GeV$/c^2$ $= 173.1 \pm 3.5$ GeV$/c^2$ for the CMS collaboration [130], using events with a lepton and multiple jets. These measurements correspond to a precision of ~ 3.7 % and 2.0 % respectively. These results are in agreement with the top quark measured by the CMS collaboration using events with two leptons : 175.5 ± 4.6 (stat.) $\pm$ 4.6 (syst.) GeV$/c^2$ $= 175.5 \pm 6.5$ GeV$/c^2$ [131].

Decay width and life-time

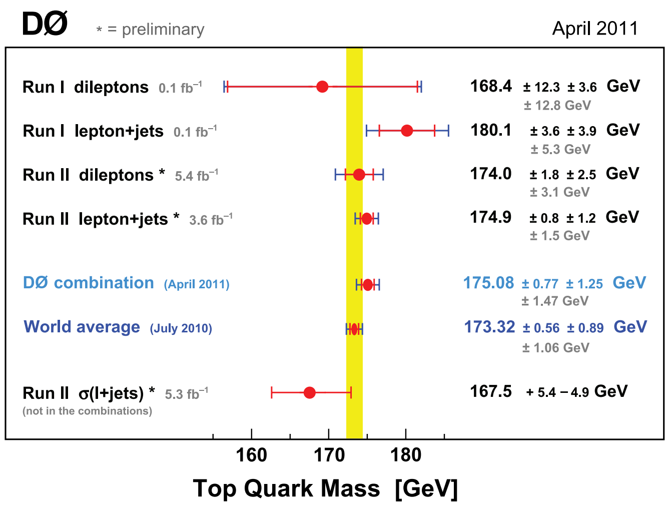
Within the Standard Model, the top quark can decay via the electro-weak force, into one of the $SU(2)_L$ doublet partner with a negative weak isospin 3^{rd} component : $t \to Wq$. Nevertheless, final state including a d or a s quark are highly suppressed due to the smallness of the square of the CKM matrix-element values $|V_{td}|$ and $|V_{ts}|$. Therefore, top quarks decay predominantly into a b quark and a W boson. This fact has been experimentally verified at Tevatron where the fraction R of top quarks decaying into a b quark and a W boson has been measured. If one assumes only three quark generations $\left(\binom{u}{d}, \binom{c}{s}, \binom{t}{b}\right)$, this ratio can be expressed as the following :

$$R = \frac{\mathcal{B}(t \to Wb)}{\mathcal{B}(t \to Wq)} = \frac{|V_{tb}|^2}{|V_{td}|^2 + |V_{ts}|^2 + |V_{tb}|^2} \tag{5.1}$$

Both Tevatron experiments, CDF and DØ have found values of R compatible with the Standard Model expectations [132, 133] : $R = 1.12^{+0.21}_{-0.19}$(stat.)$^{+0.17}_{-0.13}$(syst.) for CDF , using



(a)



(b)

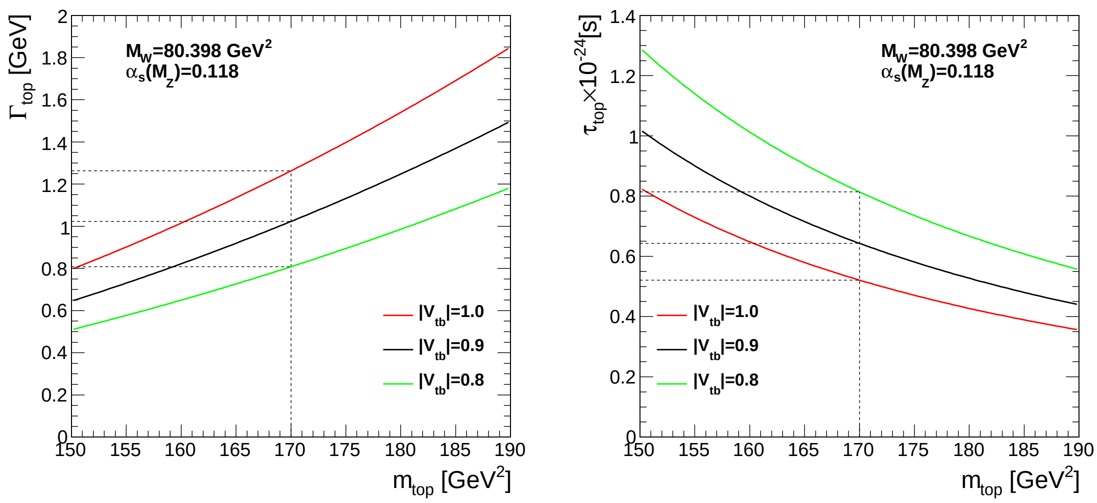
Figure 5.1: Summary of the top quark mass measurements performed by the Tevatron experiments CDF [127] and DØ [128]

data corresponding to an integrated luminosity of 162 pb^{-1} and $R = 0.97^{+0.09}_{-0.08}$(stat.+syst.) for DØ , using data corresponding to an integrated luminosity of 0.9 fb^{-1}. Furthermore, the unitarity of the CKM matrix ensures that $|V_{td}|^2 + |V_{ts}|^2 + |V_{tb}|^2 = 1$ and the ratio R can also be interpreted in terms of $|V_{tb}|^2$. However, it has been shown that in absence of direct constraint on the number of quark generations, it is possible to extend the three-generation structure of the Standard Model to include new quarks and thus, to lower the $|V_{tb}|$ value without modifying the ratio R [134]. Recently, the production of single top quark via the electro-weak force has been observed and measured. Measuring the single top production cross-section is currently the only experimental way to directly constraint the CKM matrix-element $|V_{tb}|$ without any assumption on the number of generations. Combined result of the Tevatron experiments on the single top quark cross-section measurement translates into the following constraint on the CKM matrix-element : $|V_{tb}| = 0.88 \pm 0.07$ [135].

Assuming no fourth quark generation and including first order QCD corrections, the expression of the decay width is given by [136, 137] :

$$\Gamma_{top} = \frac{G_F m_{top}^3}{8\pi\sqrt{2}}|V_{tb}|^2 \left(1 - \frac{M_W^2}{m_{top}^2}\right)^2 \left(1 + 2\frac{M_W^2}{m_{top}^2}\right) \left[1 - \frac{2\alpha_s}{3\pi}\left(\frac{2\pi^2}{3} - \frac{5}{2}\right)\right] \tag{5.2}$$

neglecting terms of order α_s^2, $\alpha_s M_W^2/m_{top}^2$ and m_b^2/m_{top}^2. In this formula, m_{top} refers to the top quark pole mass. The m_{top}-dependency of the decay width and the corresponding lifetime $\tau_{top} = 1/\Gamma_{top}$ are shown in Figures 5.2a and 5.2b respectively. For a top quark pole mass of 170 GeV$/c^2$ and a CKM matrix-element $|V_{tb}| = 1.0$, one finds a decay width of 1.26 GeV and a lifetime of 0.52×10^{-24} s. For a CKM matrix-element $|V_{tb}| = 0.8$, the decay width decreases to 0.80 GeV and the lifetime increases to 0.82×10^{-24} s. Measuring the decay width of the top quark could provide an additional constraint on V_{tb}. Unfortunately, as shown in Figure 5.2a, for top quark pole masses comprised between 150 GeV$/c^2$ and 190 GeV$/c^2$, the value of the decay width remains of the order of 1 to 2 GeV, values below current detector resolutions. Nevertheless, several attempts have been made to measure this quantity ; the latest value from the CDF experiment is compatible with the Standard Model expectation, 0.4 GeV $< \Gamma_{top} <$ 4.4 GeV (68% $C.$ $L.$) [138]. Concerning the top quark lifetime, even for top quark pole mass as low as 150 GeV$/c^2$, the values shown in Figure 5.2b are roughly one order of magnitude lower than the typical time it takes for top-flavoured hadrons or toponium bound-states to form [139]. As a result, we will assume in this thesis that 100 $\%$ of the top quarks decay and decay exclusively into a b quark and a W boson.



(a) (b)

Figure 5.2: Decay width (Γ_{top}) and lifetime (τ_{top}) of the top quark as a function of its pole mass m_{top}.

Spin correlation

Another unique feature of the top quark derives from its short life-time ; it decays before the strong interaction depolarizes its spin. Therefore, it is possible to measure the correlation between the orientations of the top quark spins in top quark pair production as it is reflected in the angular distribution of its decay products. This measurement is appealing because it can be related to the value of the CKM matrix element V_{tb} without any assumption on the number of quark generations and it is sensitive to potential new physics effects, predicting new or different top quark production and decay dynamics.

5.1.2 Production and decay channels

In high-energy proton collisions, the top quark is mainly pair produced via the strong interaction of the proton constituents. This production occurs via two different processes : quark annihilation (5.3a) and gluon fusion (5.3b,5.3c), the latter being the dominant process at the LHC. Another source of top quarks is the production of single top quark. But as it an electroweak process, it is less probable due to lower value of the electroweak coupling constant with respect to the value of the strong coupling constant. In this thesis, the single top quark production is treated as a background to the observation of top quark pair and

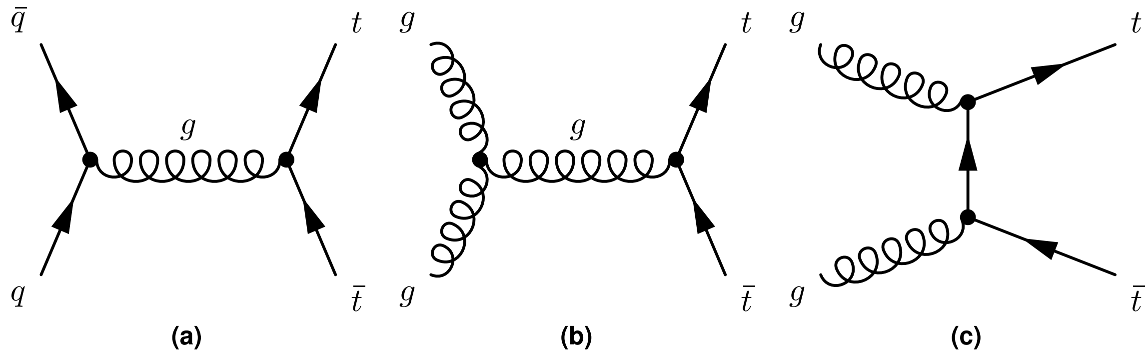


Figure 5.3: Example of Feynman diagrams at leading order in QCD corrections for top quark pair production via (a) quark annihilation and (b,c) gluon fusion.

therefore, shall be reviewed in section 5.2.1 where top quark pair background processes are reviewed.

Top quark pair production

The effective centre-of-mass energy, $\sqrt{\hat{s}}$, of the partonic interaction during a collision at a centre-of-mass $\sqrt{s}$ is $E = \sqrt{\hat{s}} = \sqrt{x_1 x_2 s}$ where x_1 and x_2 are the momentum fractions carried by the two incoming partons respectively. Assuming that, on average $x_1 = x_2 = x$, one deduces $\sqrt{\hat{s}} = x\sqrt{s}$. In order to produce a top quark pair, the available partonic energy should exceed the mass of the top quark pair : $\sqrt{\hat{s}} \geq 2m_t \Rightarrow x \geq \dfrac{2m_t}{\sqrt{s}}$. When probed with a momentum transfer $Q^2 = m_{t\bar{t}}^2 = (2m_t)^2$, assuming a top quark mass of 173.3 GeV$/c^2$, the gluon probability function for protons increases rapidly as the parton momentum fraction x decreases (Fig. 5.4a) and dominates the quark probability functions for $x < 0.1$. The minimal parton momentum fraction needed at the present LHC energy, $\sqrt{s} = 7$ TeV, to create a top quark pair is $x = 2m_{top}/\sqrt{s} \simeq 0.05$. The fraction of top quark pairs produced via gluon fusion can be found by summing over the different parton density functions integrated over the dynamically accessible range of momentum fractions. Contrary to the top quark pair production at Tevatron, at the LHC, this process is mainly gluon-induced for top quark pairs with low invariant mass, making the cross-section value sensitive to the gluon content of the protons. It is also worth to notice that the quark probability functions become dominant for high x-values, typically $x > 0.1$. This indicates that the higher the invariant mass of the top quark pair, the higher the rate of quark-initiated processes. The uncertainty on the gluon probability function (Fig. 5.4b) is rather low (2%) and stable for small values of x (typically, for $10^{-4} < x < 10^{-1}$) and grows exponentially at high values of x. Fortunately, the typical x-value for top quark pair production at the LHC ($x_{1,2}^{t\bar{t}}$) is located

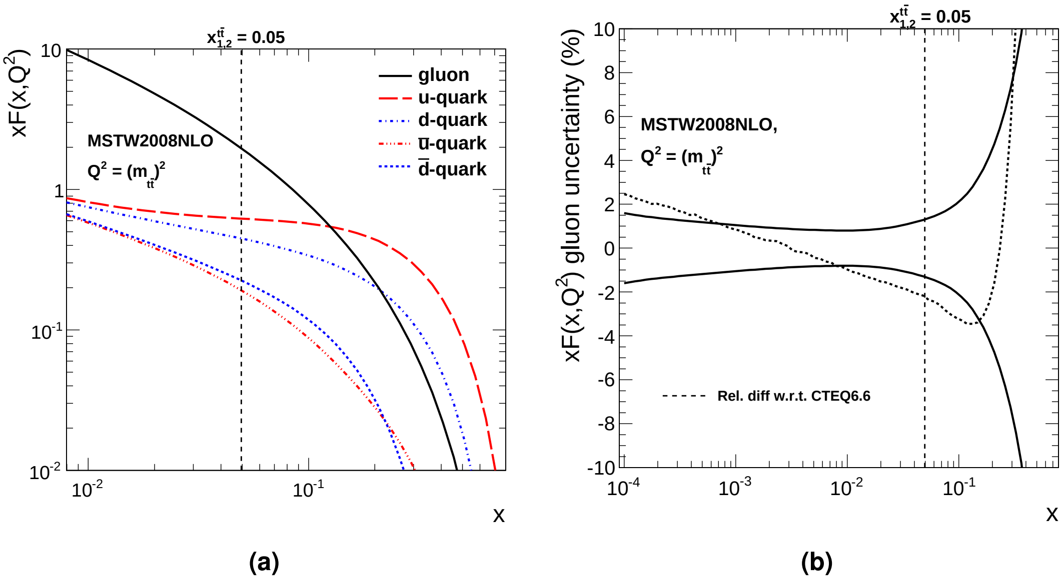


Figure 5.4: MSTW 2008 NLO parton probability functions (a) and relative uncertainties on the gluon probability function (b) for protons as a function of the parton momentum fraction, at a momentum transfer scale $Q^2 = m^2_{t\bar{t}}$. The relative difference between the MSTW 2008 NLO and CTEQ6.6 gluon probability functions is also shown (b). Figures have been made with data obtained from [140].

in region where this uncertainty is minimum. Finally, the relative difference between the MSTW 2008 NLO and CTEQ6.6 gluon probability functions (Fig. 5.4b) ranges from $+2\%$ to -4% for $10^{-4} < x < 10^{-1}$ and increases steeply for higher x-values. For $x = x^{t\bar{t}}_{1,2}$, this relative difference amounts to -2%, meaning that CTEQ6.6 predicts a slightly lower gluon probability value.

Production cross-section

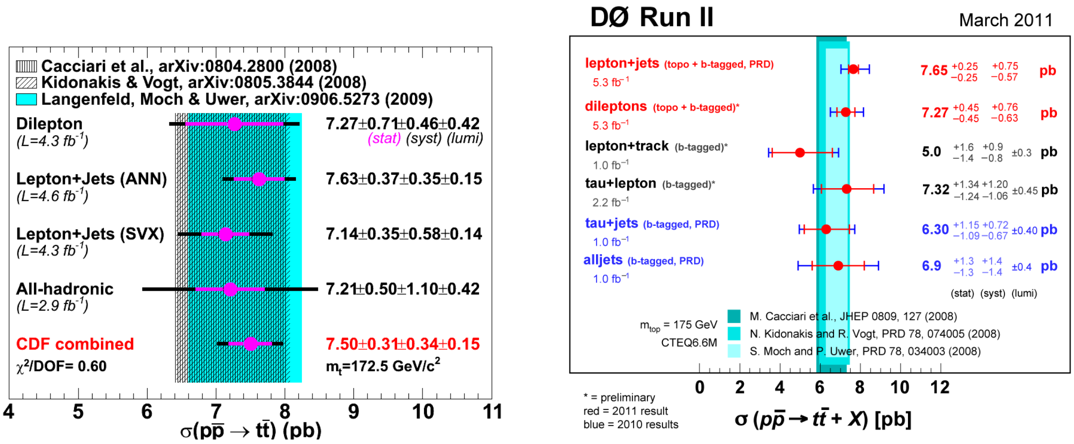
The state-of-the-art concerning theoretical predictions for the top quark pair production cross-section in proton collisions includes next-to-leading order QCD corrections as well as next-to-next-to-leading order soft gluon corrections [141]. These values are summarized in Table 5.1, showing the top quark pair production cross-section assuming two different top quark mass values, 172 GeV$/c^2$ and 173 GeV$/c^2$, for proton-antiproton collisions at $\sqrt{s} = 1.96$ TeV at Tevatron and for proton-proton collisions at the LHC at $\sqrt{s} = 7,\ 10\ 14$ TeV, using the MSTW 2008 NNLO pdf set.

	m_{top}	Tevatron	LHC		
		1.96 TeV	7 TeV	10 TeV	14 TeV
$\sigma_{t\bar{t}}$ (pb)	172 GeV$/c^2$	$7.24^{+0.30}_{-0.34}$	170 ± 10	427^{+32}_{-19}	943^{+66}_{-42}
	173 GeV$/c^2$	$7.01^{+0.29}_{-0.33}$	165 ± 10	415^{+31}_{-19}	918^{+64}_{-41}

Table 5.1: Theoretical predictions for top quark pair production cross-section at approximate NNLO, using the MSTW 2008 NNLO pdf set. Values are quoted for proton-antiproton collisions at Tevatron and for proton-proton collisions at the LHC for different centre-of-mass energies. Uncertainties on the cross-section values include pdf uncertainties as well as uncertainties associated to the variation of the renormalization and factorization scales [141].

The combination of preliminary results using events with different final-state topologies (Fig. 5.5a), obtained by the CDF collaboration, with data samples representing up to 4.6fb^{-1} of integrated luminosity is in agreement with the theoretical predictions : $\sigma_{t\bar{t}} = (7.50 \pm 0.48$ (stat.+syst.)$)$ pb for a top quark mass of 172.5 GeV$/c^2$ [142]. Results obtained by the DØ collaboration using events with different final-state topologies (Fig. 5.5b) are also compatible with the theoretical predictions within the their uncertainties. However, the DØ collaboration does not provide a combination of all these measurements. Nevertheless, this work is in progress and the DØ collaboration already provided the combination of its two latest and most accurate results, using events with two leptons and events with a lepton and multiple jets, representing 5.4 fb^{-1} of integrated luminosity : $\sigma_{t\bar{t}} = (7.56^{+0.63}_{-0.56}$ (stat.+syst.)$)$ pb for a top quark mass of 172.5 GeV$/c^2$ [143].

The first top quark pair production cross-section in proton-proton collisions has been measured by the CMS collaboration, using events with two leptons, representing a data sample of 3.1 pb^{-1} of integrated luminosity : $\sigma_{t\bar{t}} = (194 \pm 72$ (stat.)± 24 (syst.)± 21 (lumi.)$)$ pb for a top quark mass of 172.5 GeV$/c^2$ [144]. Shortly after, this result was confirmed by the ATLAS collaboration [145]. Using data collected during the year 2010, representing ~ 36 pb^{-1} of integrated luminosity, more precise measurements have been performed by both collaborations (Fig. 5.6). The combination of the results obtained by the CMS collaboration yields : $\sigma_{t\bar{t}} = (158 \pm 10$ (stat.) ± 15 (syst.) ± 6 (lumi.)$)$ pb for a top quark mass of 172.5 GeV$/c^2$ [146]. The combined measurement obtained by the ATLAS collaboration, $\sigma_{t\bar{t}} = (180 \pm 9$ (stat.)± 15 (syst.)± 6 (lumi.)$)$ pb [147], is somewhat higher but yet compatible with the result produced by CMS . Both combinations are in agreement with theoretical predictions.



(a) (b)

Figure 5.5: Summary of the top quark pair cross-section measurements made by the CDF [127] (a) and the DØ [128] (b) collaborations. Statistical uncertainties are shown by the light colour line, which is superimposed on total uncertainties including systematic and luminosity uncertainties as shown by the black line. Theoretical predictions are represented by the hatched or shaded areas.

Decay channels

As discussed in section 5.1.1, the top quark decays almost exclusively into a b quark and a W boson : $t \to bW$. Therefore, the final-state topology of the top quark pair production is determined by the subsequent decay of the two W bosons, either hadronically ($W \to q\bar{q}'$) or leptonically ($W \to l\nu_l$). Consequently, three main topologies can be distinguished :

1. the fully hadronic topology (Fig. 5.7a) : both W bosons decay hadronically, leading to six quarks in the final state : four light quarks coming from the W decays and two b quarks : $pp \to t\bar{t} \to b\bar{b}q\bar{q}'q''\bar{q}'''$.
2. the di-leptonic topology (Fig. 5.7b) : both W bosons decay leptonically, leading to two b quarks, two charged leptons and missing transverse energy due to the two neutrinos escaping the detection : $pp \to t\bar{t} \to b\bar{b}l\bar{\nu}l'\bar{\nu}'$.
3. the single-leptonic topology (Fig. 5.7c) : one of the W bosons decays hadronically while the other one decays leptonically, leading to a charged lepton, missing transverse energy due to the neutrino and four quarks, among which two are b quarks : $pp \to t\bar{t} \to b\bar{b}l\bar{\nu}q\bar{q}'$.

These topologies are also sometimes referred to as *all jets*, *di-lepton* and *single lepton plus jets* respectively. Strictly speaking, with these definitions of the different topologies, di-leptonic and single-leptonic topologies should include tau lepton production. However,

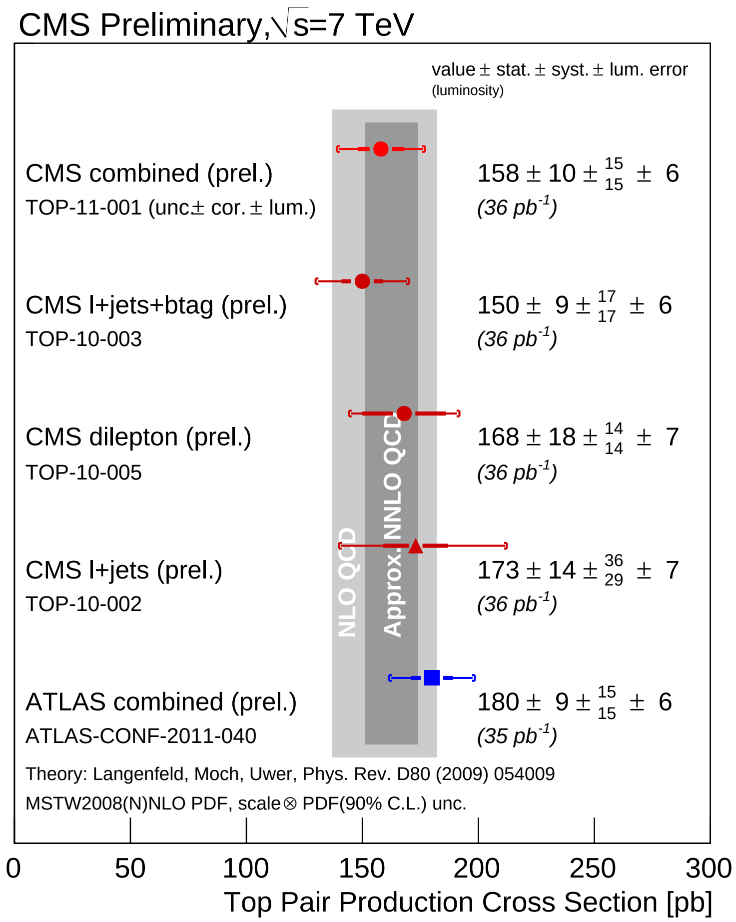


Figure 5.6: Summary of various inclusive top pair production cross section measurements made in 7 TeV proton-proton collisions by CMS and ATLAS [148]. The inner error bars of the data points correspond to the statistical uncertainty, while the outer (thinner) error bars correspond to the quadratic sum of statistical and systematic uncertainties. The outermost brackets correspond to the total error, including a luminosity uncertainty which is also added in quadrature. Theory predictions at NLO and approximate NNLO, represented by the grey-shaded area, are obtained using the HATHOR generator [149].

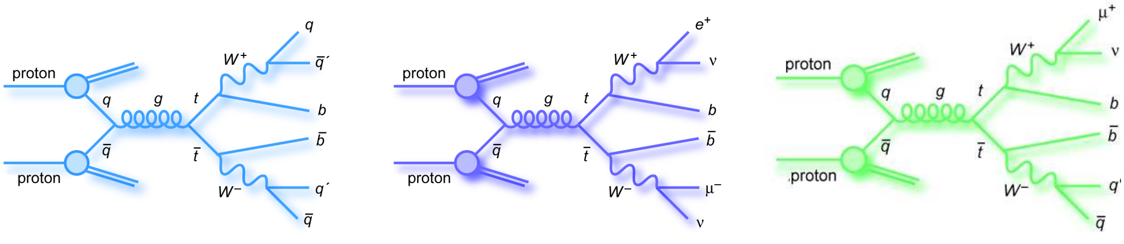


Figure 5.7: Sketches representing the top quark pair final state topology for the fully hadronic (a), di-leptonic (b) and semi-leptonic (c) decay channels. Sketches adapted from [128].

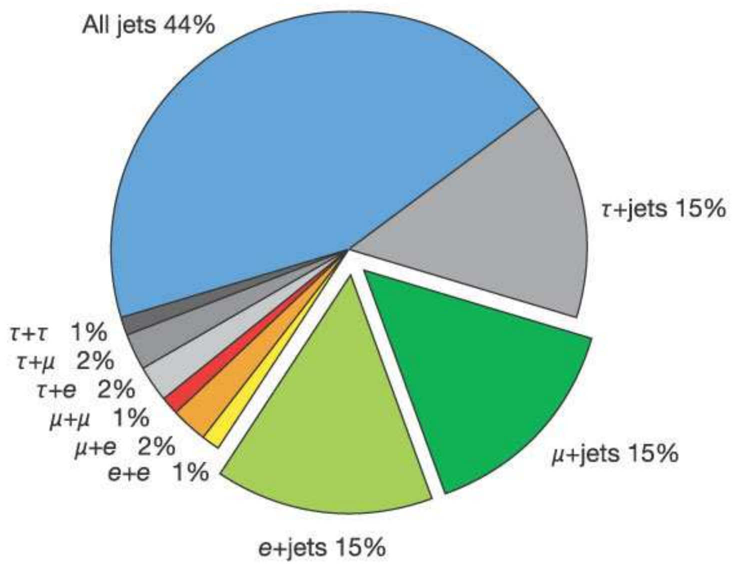


Figure 5.8: Branching fractions of the different final-state topologies for top quark pair production.

due to the shorter tau mean lifetime with respect to the muon one, these topologies are often regarded as an extra category. The branching fraction of the fully hadronic topology represents $\sim 44\%$ of the total, the semi-leptonic (muon and electron) topology $\sim 30\%$ and the di-leptonic topology $\sim 4\%$. All together, topologies involving tau leptons amount to 21% (Fig. 5.8).

5.1.3 Search for new physics with top quark pairs

Many extensions of the Standard Model predict the existence of gauge interactions with enhanced couplings to the third quark generation, resulting in new particles, such as a top-colour Z' boson [121] or a Kaluza-Klein excitation of the gluon, g_{KK}, in models with warped extra-dimensions [150]. These particles could show up as resonances in the $pp \to X(= Z', g_{KK}) \to t\bar{t}$ production channel, leading to distortions of the top quark pair invariant mass distribution, as illustrated in [151] . Such particles have been searched for by Tevatron experiments, allowing to put lower limits on their masses. Recently, similar searches have been performed by the LHC experiments, ATLAS and CMS. After reconstruction, the top quark pair invariant mass obtained with the selected events, is compared with the Standard Model predictions, using Monte-Carlo simulations. An illustration is shown in Figure 5.9. No evidence for the existence of a leptophobic top-colour Z' nor of a KK excitation of the gluon have been found by the CMS and ATLAS experiment. Using 36 pb^{-1} and 200 pb^{-1}

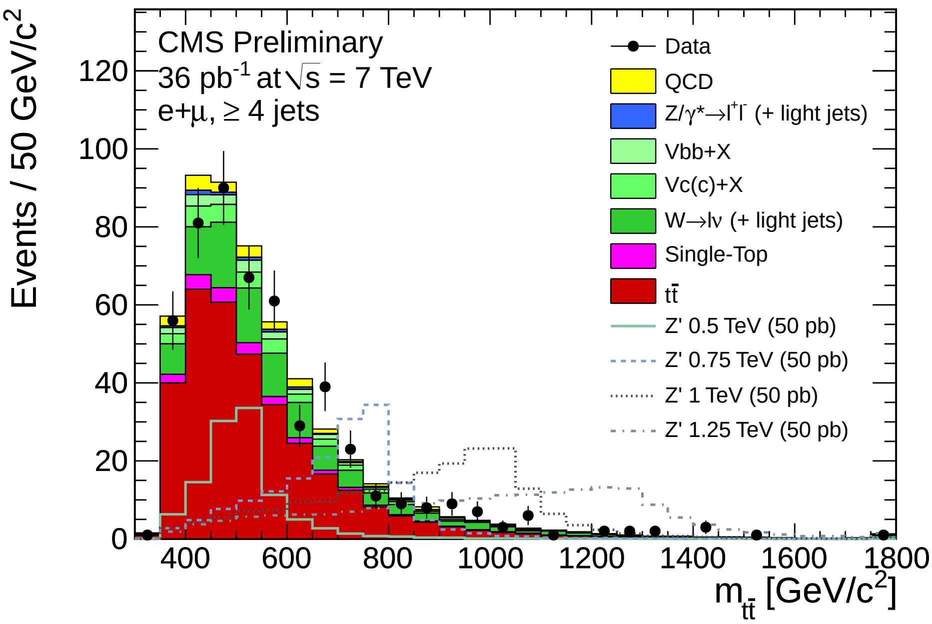
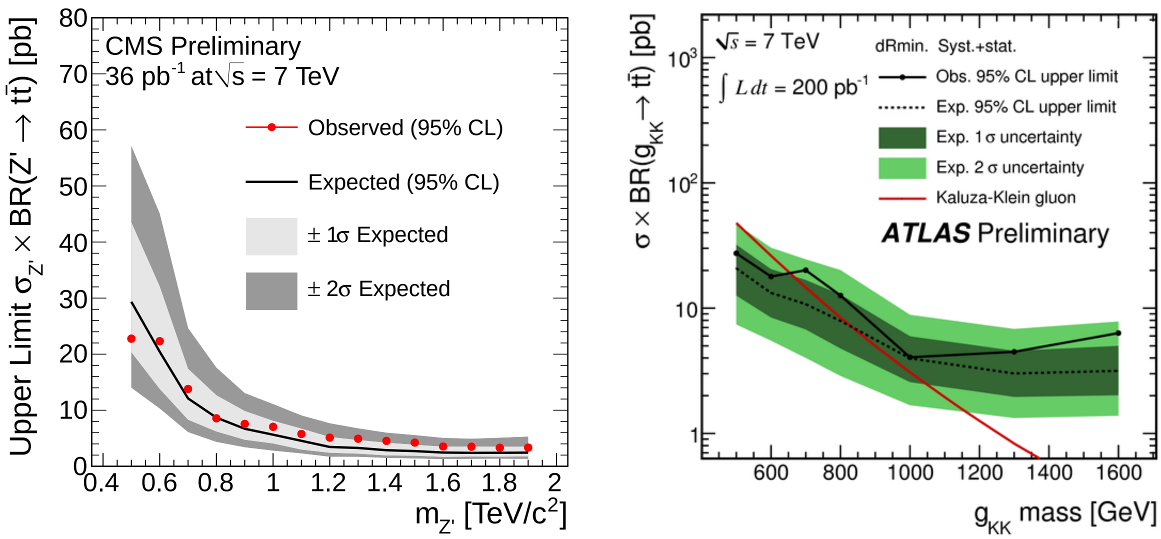


Figure 5.9: Reconstructed top quark pair invariant mass with events with at least four jets in the final state, in the semi-leptonic top quark pair decay channel (electron and muon) [152]

of integrated luminosity respectively, both experiments have therefore deduced lower limits on the production cross-section or on the mass of these particles (Fig. 5.10). A limit of the order of 7 pb for invariant masses in the region $m_{Z'} = 1$ TeV$/c^2$ has been set by the CMS collaboration [152] and a limit of the order of 4 pb for invariant masses in the region $m_{g_{kk}} = 1$ TeV$/c^2$ has been set by the ATLAS collaboration, excluding g_{KK} resonances with masses below 650 GeV$/c^2$ at 95 % C. L.



(a) (b)

Figure 5.10: Expected (dashed line) and observed (black points connected by a line) upper limits on $\sigma(pp \to Z') \times BR(Z' \to t\bar{t})$ as obtained by the CMS collaboration (a) and $\sigma(pp \to g_{KK}) \times BR(g_{KK} \to t\bar{t})$, as obtained by the ATLAS collaboration (b).

5.2 Top quark background processes

As described in the previous section, top quark pairs decay via three different channels, leading to different final-state topologies. In this thesis, we are interest in the semi-muonic decay channel[1] whose final state contains an isolated energetic muon, multiple energetic jets and some missing transverse energy due to the neutrino, escaping the detection. The di-leptonic decay channel suffers from a small branching fraction and the presence of two neutrinos lead to an ambiguity during the event reconstruction. With respect to the fully hadronic decay channel, the semi-leptonic decay channel suffers less from overwhelming processes leading to a final state, which is either identical or mis-identified as such. These processes are referred to as background processes and can be of two kind :

- irreducible background : although different, this kind of process leads the very same final-state particles as the ones produced via top quark pair decays.
- instrumental background : the final state is identified as identical to the top quark pair one because of detector effects and/or particle mis-identification.

In this section, the most relevant background processes for top quark pair studies considered in this thesis are reviewed, emphasizing their theoretical uncertainties, as well as the difficulties to simulate reliably such processes.

5.2.1 Single top production

As mentioned in section 5.1.2, single top quarks can be produced via weak interactions of two quarks, either in the s-channel (Fig. 5.11a) leading to a single top quark and a b quark in the final state or in the t-channel (Fig. 5.11b), leading to a single top quark and a light flavour quark. Single top quarks can also be produced in association with a W boson via b quark excitation (Fig 5.11c).

In the context of studies of top quark pairs in the semi-muonic decay channel, the single top quark production is a source of irreducible background ; in the s-channel, with the subsequent top quark decay, the final state presents two b quarks and a W boson which may decay to a muon and a neutrino. When associated with initial or final state radiations (ISR/FSR) of quarks or gluons, this final state is identical to the final state of top quark pairs decaying via the semi-muonic channel. At next-to-leading order, the incoming b quark in

[1] In principle, all the methods developed and applied in this thesis could use the semi-electronic decay channel, given some modifications of the event selection, presented in the next section.

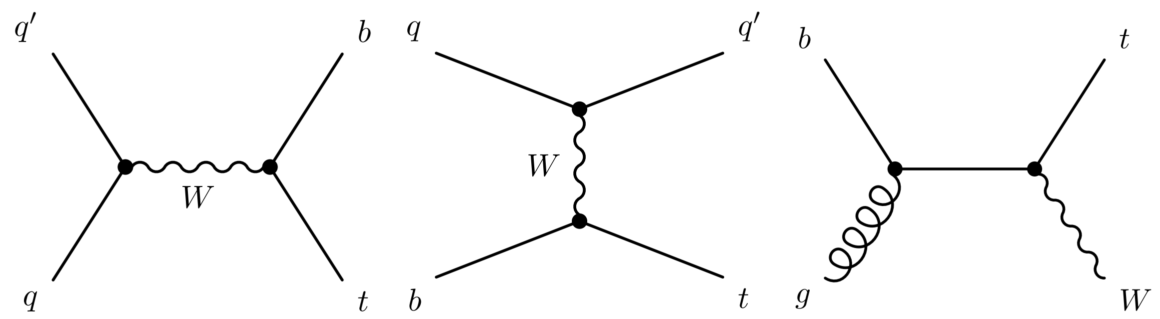


Figure 5.11: Feynman diagrams at leading order in QCD and EWK corrections for weak production of single top quarks, via the s-channel (a), the t-channel (b) and the tW-channel (c).

Process		$t\bar{t} + jets$	Single top quark		
			s-channel	t-channel	tW-channel
Tevatron	($\sqrt{s} = 1.96$ TeV)	$7.12^{+0.29}_{-0.33}$	1.06 ± 0.05	2.12 ± 0.15	$.22 \pm 0.06$
LHC	($\sqrt{s} = 7$ TeV)	$157.5^{+23.2}_{-24.4}$	$4.21^{+0.19}_{-0.18}$	$64.6^{+3.4}_{-3.2}$	10.6 ± 0.6
$\sigma^{LHC}/\sigma^{Tevatron}$		~ 22	~ 4	~ 30	~ 48

Table 5.2: Top quark pair and single top quark production cross-sections calculated at approximate next-to-next-to leading order (NNLO) and next-to-next-to-next-to leading order (NNNLO) respectively, for Tevatron [141] and calculated at the next-to leading order (NLO) at LHC with MCFM [153, 154]. The single top cross-section has been split according to its different production channels. Ratios between the cross-sections at Tevatron and at the LHC have also been reported.

the t- and tW-channels comes from a gluon-splitting, yielding then an additional so-called spectator b-quark. This leads, with the subsequent top quark decay and association with ISR/FSR, to the same final state as the s-channel.

At Tevatron, the single top quark production rate represents $\sim 40\%$ of the production rate of top quark pairs (Table 5.2). However, its observation is extremely difficult because of an overwhelming background and the single top quark production has been observed only recently by the Tevatron experiments [155, 156], thanks to the development of advanced detection techniques like neural network or boosted decision trees. The most accurate measurement performed by the CDF experiment, combines production cross-section for single top quark in the s- and t-channel : $\sigma_t^{s+t-ch.} = (2.3^{+0.6}_{-0.5}$ (stat.+syst.)) pb, using a data sample representing 3.2 fb^{-1} of integrated luminosity [157]. More recently, the DØ experiment measured the single top quark production cross-section in the s- and t-channels

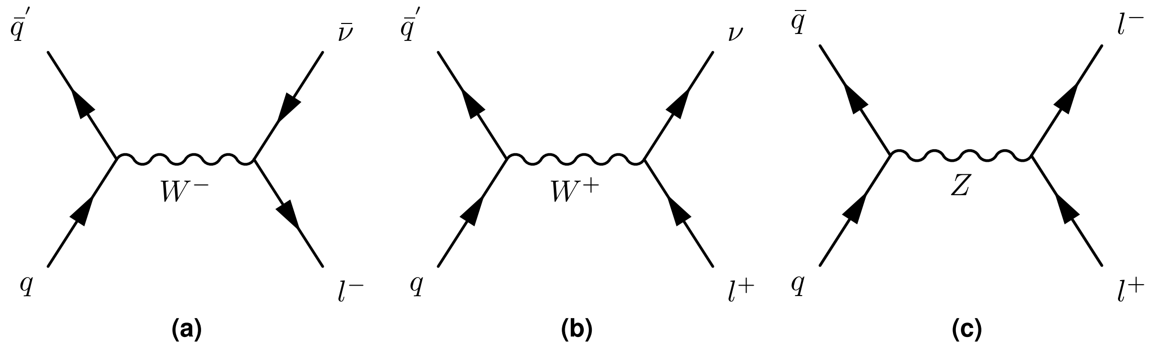


Figure 5.12: Feynman diagrams at leading order in QCD and EWK corrections for W boson (a,b) and Z boson (c) production.

separately, using 5.4 fb^{-1} of integrated luminosity [158] : $\sigma_t^{s-ch.} = (0.98 \pm 0.63$ (stat.+syst.)$)$ pb and $\sigma_t^{t-ch.} = (2.90 \pm 0.59$ (stat.+syst.)$)$ pb. Both measurements assume a top quark mass of 172.5 GeV$/c^2$ and are in agreement with theoretical calculations including next-to-next-to-next-to-leading order QCD corrections (Table 5.2). At the LHC, the expected single top quark production rate via weak interactions represents more than 50 % of the rate of top quark pairs. Recently, using proton-proton collision data representing ~ 36 pb^{-1} of integrated luminosity, both ATLAS and CMS measured the single top quark production cross-section in the t-channel : $\sigma_t^{t-ch.} = (53^{+26}_{-24}$ (stat.)$^{+38}_{-27}$ (syst.)$)$ pb for ATLAS [159] and $\sigma_t^{t-ch.} = (83.6 \pm 29.8$ (stat.+syst.) $\pm\, 3.3$ (lumi.)$)$ pb for CMS [160].

5.2.2 Vector boson production

In hadron collisions, the production of vector bosons, $V = W^{\pm}, Z/\gamma^*$, decaying leptonically has a clean signature (Fig. 5.12) : an isolated lepton with a high transverse momentum value and missing transverse energy due to the escaping neutrino in the case of the W boson and two isolated leptons of opposite charges with high transverse momentum values in the case of the Z boson. When such a vector boson is produced in association with multiple jets ($V + jets$), it may become a source of irreducible background events for studies of top quark pairs in the semi-leptonic channel if there is no requirement for two jets identified as b-jets in the final state. The leptonic decay of the W boson may lead to a muon and a neutrino while, in the case of the leptonic decay of a Z boson, one of the two leptons may be outside the detector acceptance, leading here also to a single lepton in the final state. In this case, there should be no missing transverse energy due to the escaping neutrino but fake missing transverse energy may arise from detector inefficiencies and/or jet energy mis-calibration.

Jet multiplcity	$\sigma(W + (n)jets) \times BR(W \to e\nu)$ (pb)		
($p_T > 25$ GeV/c)	Data	LO	NLO
$n = 1$	53.5 ± 5.6	$41.40^{+7.59}_{-5.94}$	$57.83^{+4.36}_{-4.00}$
$n = 2$	6.8 ± 1.1	$6.159^{+2.41}_{-1.58}$	$7.62^{+0.62}_{-0.86}$
$n = 3$	0.84 ± 0.24	$0.796^{+0.488}_{-0.276}$	$0.882^{+0.057}_{-0.138}$
$n = 4$	0.074 ± 0.053	N.A.	N.A.

Table 5.3: Cross-sections measured at Tevatron by the CDF collaboration [161] using 320 pb^{-1} of integrated luminosity and theoretical cross-sections calculated at leading and next-to leading order [162] for the production of $W \to e\nu$ boson in association with n jets. Theoretical calculations for n=4 are not available (N.A.). Uncertainties associated to the theoretical calculations are related to the renormalization and factorization scale dependence.

At Tevatron, the $V+$ n-jet production cross-sections have been measured by both the CDF and DØ experiments, up to $n^W = 4$ for W boson production [161, 163] and up to $n^Z = 3$ for the Z boson production [164, 165]. In table 5.3, results obtained by the CDF collaboration for W bosons are compared with theoretical cross-sections calculated at leading and next-to leading order. For Z bosons, these comparisons are made in Figure 5.13 for the CDF collaboration using 6.1 fb^{-1} of integrated luminosity and in Figure 5.14 for the DØ collaboration using 1.0 fb^{-1}. Compared to data, predictions obtained with leading order matrix-element generators (ALPGEN and MADGRAPH) for such processes are underestimated and corrections at least at next-to-leading order are needed to reproduced the observed total and differential $V + jets$ cross sections.

At the LHC, the $V+$ n-jet production cross-sections have also been measured by the ATLAS collaboration, up to $n^W = 4$ and $n^Z = 3$ for W [166] and Z bosons [167] respectively and by the CMS collaboration [168], up to $n^W = 6$ and $n^Z = 5$. Measurements performed by both collaborations use a data sample representing ~ 36 pb^{-1} of integrated luminosity. Good overall agreement has been found between measured differential cross-sections and predictions using leading-order matrix element generators, rescaled to match the next-to-next-leading order total vector boson cross-section. However, it is worth noticing that the statistical uncertainties on the measured cross-sections exceed 25 % and 40% for W and Z boson production associated with at least four jets respectively.

At the moment of the writing, next-to-leading order corrections are available for the $V+$ n-jet processes up to 3 jets [162, 169, 170, 171] and first results have been presented very recently for the $W + 4\ jet$ process [172]. As more and more data are accumulated

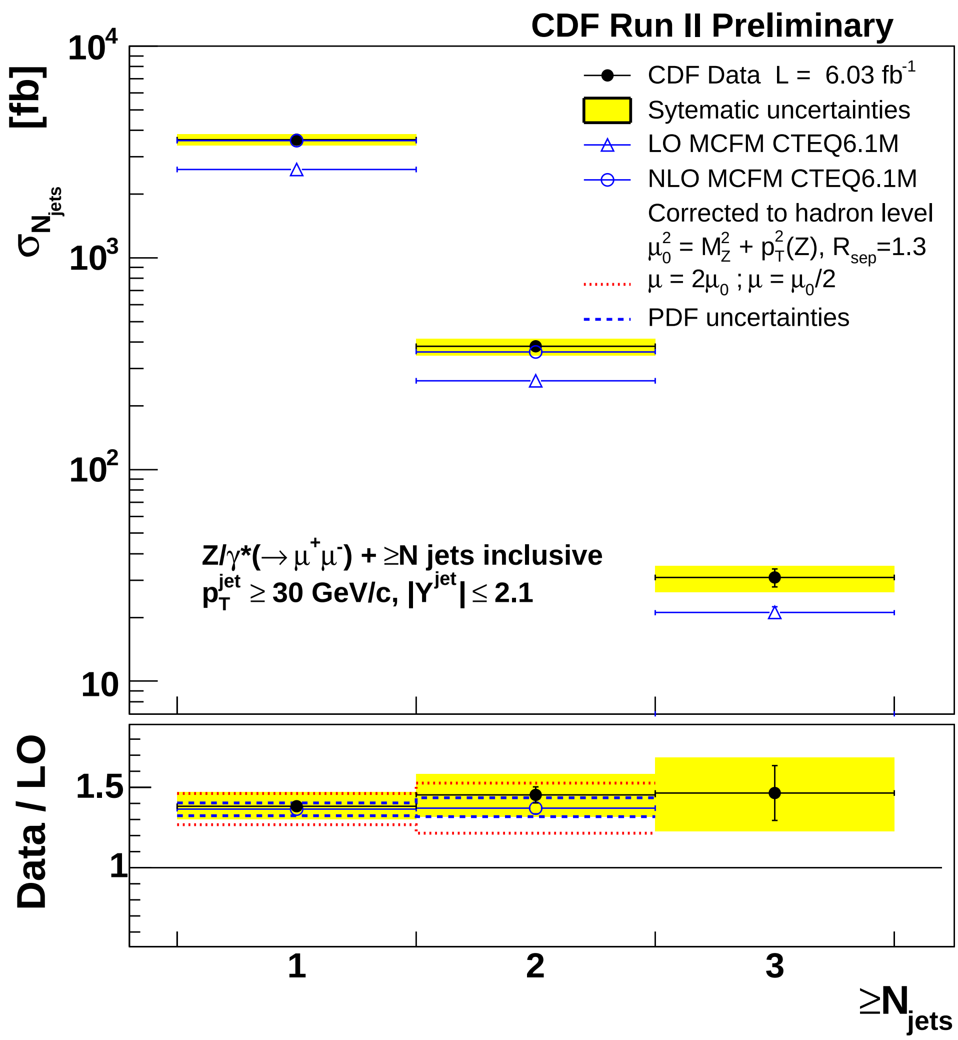


Figure 5.13: Upper plot : Cross sections of $Z(\to \mu\mu)$ boson production as a function of the inclusive jet multiplicity, measured at Tevatron by the CDF collaboration. Next-to leading order predictions as determined using `MCFM` are also shown.
Lower plot : ratios of measured cross-sections to leading order predictions as well as ratios of next-to-leading to leading order predictions. Systematic uncertainties on data and theoretical predictions are included [164].

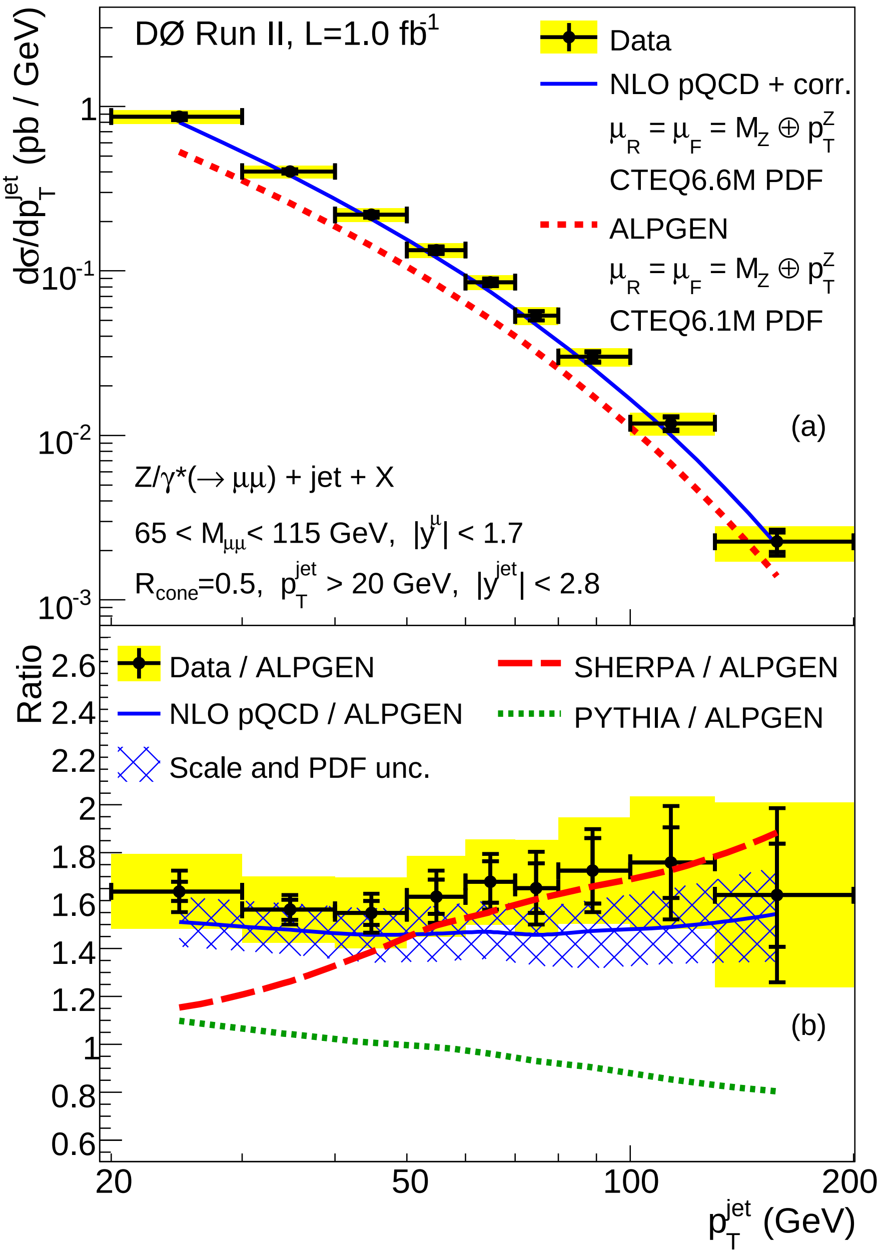


Figure 5.14: Upper plot : Differential cross sections of $Z(\rightarrow \mu\mu)$ boson production as a function of the leading jet transverse momentum, measured at Tevatron by the DØ collaboration. Leading order predictions using the ALPGEN generator and next-to leading order predictions calculated with MCFM are also shown.
Lower plot : ratios of measured cross-sections, next-to leading predictions calculated with MCFM and leading-order predictions obtained with the SHERPA and PYTHIA generators to leading order predictions obtained with the ALPGEN generator [165].

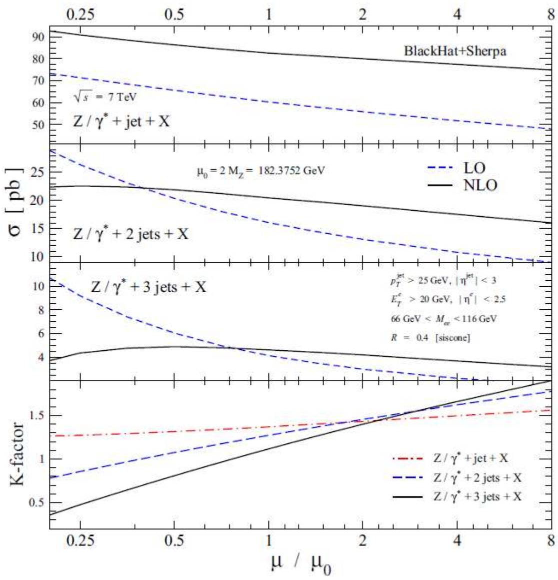
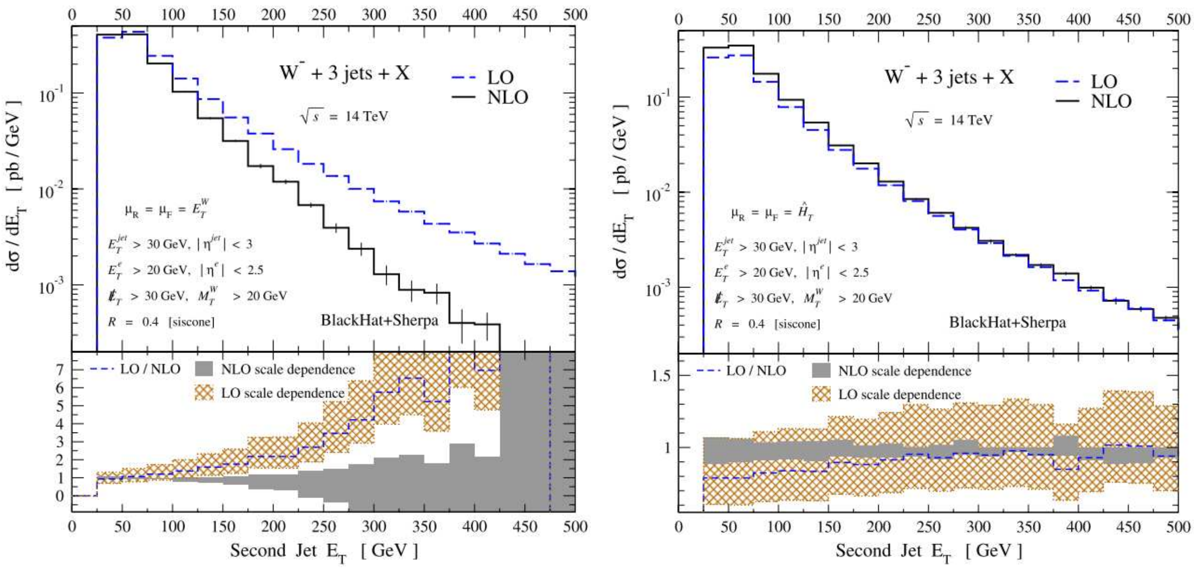


Figure 5.15: Theoretical predictions of leading-order (LO) and next-to-leading order (NLO) cross sections for Z boson production in association with up to 3 jets in proton collisions at the LHC at 7 TeV, as a function of the common renormalization and factorization scale μ, with respect to a reference scale μ_0, being equal to twice the Z boson mass. The bottom panel shows the ratio between the NLO and LO cross sections (K-factor), for the different jet multiplicities [173].

at the LHC, these corrections will become necessary for precise measurements as the ratio between cross-sections at the next-to leading order and at leading order in QCD corrections grows as a function of the jet multiplicity. In addition, as shown in Figure 5.15 for Z boson production, cross sections obtained at leading order strongly depend on the chosen values for the renormalization and factorization scales. This dependence, growing also with the jet multiplicity, is reduced when next-to-leading order corrections are applied .

Besides, $V + jets$ processes are difficult to simulate reliably. In most cases, it is impossible to reproduce the shape of differential cross section distributions at next-to-leading order by simply applying a constant K-factor to the shapes calculated with leading order matrix-element generators. In addition, differences in shapes between differential cross section distributions at leading and next-to-leading order corrections in QCD are sensitive to the choice of the renormalization and factorization scales. Figure 5.16 shows differential $W^- + 3$ jets cross-section as a function of the transverse energy of the second leading jet in the event, calculated with a common renormalization and factorization scale equal to the W boson transverse energy, E_T^W (Fig. 5.16a) and equal to the scalar sum of the transverse energy of all the final-state partons, $\hat{H}_T$ (Fig 5.16b). In the first case, the difference between cross-sections at leading order and next-to-leading order grows with



(a) **(b)**

Figure 5.16: Theoretical predictions at leading order (dashed lines) and next-to-leading order (solid line) for the differential $W^- + 3$-jet production corss-sections as a function of the second jet transverse energy at LHC, using a common renormalization and factorization scale, μ, equal to the W boson transverse energy, E_T^W (left plot) and equal to the scalar sum of the transverse energy of all the final-state partons, $\hat{H}_T$ (right plot). The lower plots show the ratio of the cross-sections calculated at leading and next-to leading order as well as their scale-dependence [162].

the jet transverse energy while the latter case, this ratio remains constant. Unfortunately, it remains unclear if such an appropriate scale exists and is unique for all the observables experimentalists are interested in. Finally, it has been found that these corrections might show a strong dependence on the kinematic cuts applied on the final state particles [174] in the case of QCD corrections. For electro-weak (EWK) corrections, their contributions to the total cross-section are of the order of the percent. Nevertheless, they can affect the transverse momentum of its decay product up to 10% [175, 176] or the transverse momentum of the associated jets [176]. In the latter case, the effects scale from 5% up to 30% for $p_T \in [100, 1000]$ (GeV$/c$) of the additional hard jet for the $W + 1\ jet$ process, which is the state-of-the-art for NLO EWK corrections.

Production in association with heavy-flavour jets

Typically, top quark pair selection criteria take advantage of the presence of two b-quarks, coming from the top quark decay, by requiring in the final state one or two jets identified as b-jets (see section 4.3.3). By doing so, the number of background events can be significantly reduced. Unfortunately, the production of vector bosons may also be associated with jets from heavy-flavour quarks and therefore perfectly mimic the top quark pair final state. Although, for a given jet multiplicity, this kind of processes should account only for a few percent of the total vector boson production, their topology make them very difficult to reject efficiently.

$V + c\text{-}jets$: the production of vector boson associated with a single c quark occur via gluon-quark scattering, mainly strange quark for W boson and charm quark for the Z boson and are therefore sensitive to the corresponding quark content inside a proton. In the case of the production of W boson, the production rate is also sensitive to the CKM matrix element $|V_{cs}|$. This kind of background events are particularly relevant for top quark physics because jets originating from the hadronization of charm quarks have the highest probability to be mis-identified as b-jets, compared to any other quark. Pair of c quarks may also be produced in association of a vector boson via gluon splitting, $g \to c\bar{c}$. Recently, the ratio of the cross-section of single charm quark production in association with a W boson to the total W boson production cross-section has been measured by the CMS collaboration, using 36 pb^{-1} of integrated luminosity : $\sigma(W{+}c)/\sigma(W{+}jets) = (0.143{\pm}0.015$ (stat.+syst.)) [177]. This result is in agreement with predictions calculated at the next-to-leading order in QCD corrections with `MCFM` .

$V + b\text{-}jets$: these processes occur mainly via the splitting of an initial state gluon radiation into a b quark pair. Theoretical predictions for the $Wb\bar{b}$ and $Zb\bar{b}$ cross-sections foresee large QCD corrections at the next-to leading order, as well as a strong dependence of the k-factor on kinematic variables, such as the transverse momentum of the vector boson or the leading b-jet, having thus a large impact on the trigger and event selection efficiencies. This makes the use of next-to leading order QCD corrections mandatory, not only for the calculation of the total production cross-section but also for the prediction of the differential distributions. But in contrast to the $V + lf - jets$ processes, for which the NLO QCD corrections stabilize the scale dependence of their cross-sections, this dependence might even be worse at NLO than LO for $V + bb$ processes [178], because of the presence of new production channels. The difficulty to predict and simulate accurately $V + bb$ processes is illustrated by the Tevatron measurement of their cross-sections. Despite the agreement of the measured inclusive $W + jets$ cross-section and the theoretical predictions at the next-to leading order in α_s, the measured $W + b\text{-}jets$ cross-section ($\sigma_{bjets} \times BR(W \to l\nu) = 2.74 \pm 0.27$ (stat.) $\pm\, 0.42$ (syst.) pb) [179] differs by more than a factor 2 from the theoretical expectation at next-to-leading order in QCD corrections.

Recently, the ratio of the cross-section of the associated production of a Z boson with at least one b quark to the total Z boson production cross-section has been measured by the CMS collaboration, using 36 pb^{-1} of integrated luminosity : $\sigma(Z+b)/\sigma(Z+jets) = (0.054 \pm 0.016$ (stat.+syst.)$)$ [180]. This result is higher than predictions calculated at the next-to-leading order in QCD corrections with `MCFM` (0.043 ± 0.005) but remains in agreement due to its large statistical and systematic uncertainties.

5.2.3 Weak boson pair production

The weak boson pair ($VV,\ V = W^{\pm}/Z$) production is also a source of background events for top quark pair studies, to a lesser extent than the background processes mentioned in the previous sections though. Produced in association with jets, these processes are sources of both instrumental (1) and irreducible (2) background :

1. $pp \to (Z \to l^+l^-)(Z \to q\bar{q}) + jets$, where one of the lepton is produced out of the detector acceptance region or rejected by the event selection cuts. Fake missing transverse energy may arise from mis-measured energies of the final state particles.
2. $pp \to (V \to qq)(W \to l\nu) + jets$.

Corrections for the quark-induced process, $qq \to VV$, are known at the next-to-leading order in α_s, including single-resonant contributions [181]. It seems that a constant k-factor of 1.5 can safely be applied on the cross-section as well as on the differential distributions. More recently, the contribution of the gluon-induced process to the total cross-section has been calculated. Although the gluon-induced process contributes at the next-to-next-to leading order in α_s, it is enhanced at the LHC, compared to Tevatron, because of a higher gluon flux. Moreover, it may be that its contribution is further enhanced by the event selection cuts, as it has been shown in the case of Higgs boson searches [182].

The weak boson pair production cross-section has only been measured recently by the CDF collaboration [183] : $\sigma_{WW+WZ} = 16.0 \pm 3.3$ (stat.+syst.) pb. This measurement appears to be consistent with the theoretical expectations. Using 36 pb^{-1} of integrated luminosity, the weak boson pair cross-sections have also been measured recently by the CMS experiment : $\sigma_{WW} = 55.3 \pm 3.3\ (\mathrm{stat.}) \pm 6.9(\mathrm{syst.}) \pm 3.3(\mathrm{lumi.})$ pb [184] and by the ATLAS experiment : $\sigma_{WW} = 41^{+20}_{-16}(\mathrm{stat.}) \pm 5(\mathrm{syst.}) \pm 1(\mathrm{lumi.})$ pb [185]. Both measurements are in agreement with predictions calculated at next-to-leading order in QCD corrections with `MCFM` .

5.2.4 Multi-jet process

A multi-jet process refers to $2 \to 2$ processes involving only quarks and gluons, leading to multiple jets in the final states due to initial or final state radiations. This kind of background becomes relevant for top quark studies at high jet multiplicity, when pion or kaon decays or semi-leptonic decays of c or b quarks inside a jet lead to a charged lepton in the final-state, in addition to jets. It is also possible that either a narrow jet is mis-reconstructed as a lepton or that an energetic jet is not completely absorbed by the hadronic calorimeter and reaches the muon system, leading to the reconstruction of a fake muon. The rejection rate of such a background is expected to increase as the knowledge of the detector and reconstruction algorithm performances increase. Nevertheless, small fluctuations in the selection efficiency lead to huge fluctuations in the number of selected multi-jet events because the huge production cross-section of such process, almost 2×10^6 times larger that the top quark pair production cross-section. Furthermore, in practice, is is difficult to simulate enough multi-jet events to estimate their selection efficiency with a sufficient precision. Finally, the multi-jet cross-section suffers from the same theoretical uncertainties at high jet multiplicity in the final state as these mentioned for the single vector production in associated with jets.

At Tevatron, most of the measurements concern the inclusive or differential di-jet production cross-sections and have been found in agreement with theoretical predictions with next-to-leading order QCD corrections. Recently, the ratio $R_{3/2}$ of the inclusive 3-jet cross-section to the inclusive 2-jet cross-section has been measured, as a function of the leading jet transverse momentum, by the DØ experiment [186], using 700 pb^{-1} of integrated luminosity. Results have been compared with predictions of several Monte-Carlo generators. While predictions from the SHERPA generator have been found in agreement with the data within ± 10 %, predictions from the PYTHIA generator have been found to be systematically higher than the measurements, up to 50 % and very sensitive to the parametrization of the parton shower process. A similar measurement has been recently performed by the CMS experiment [187], using 36 pb^{-1} of integrated luminosity ; the ratio $R_{3/2}$ has been measured as a function of the total jet transverse momentum, H_T, and compared with different Monte-Carlo generator predictions. The predictions of the MADGRAPH generator have been found in agreement with the measured ratio over the entire H_T range of the measurement. Predictions from other generators overestimate the ratio for low value of H_T. Finally, the multi-jet cross-section has been measured by the ATLAS experiment [188] using 2.43 pb^{-1} of integrated luminosity ; measurements both as a function of the jet multiplicity and as a function of the third leading jet transverse momentum for events with at least

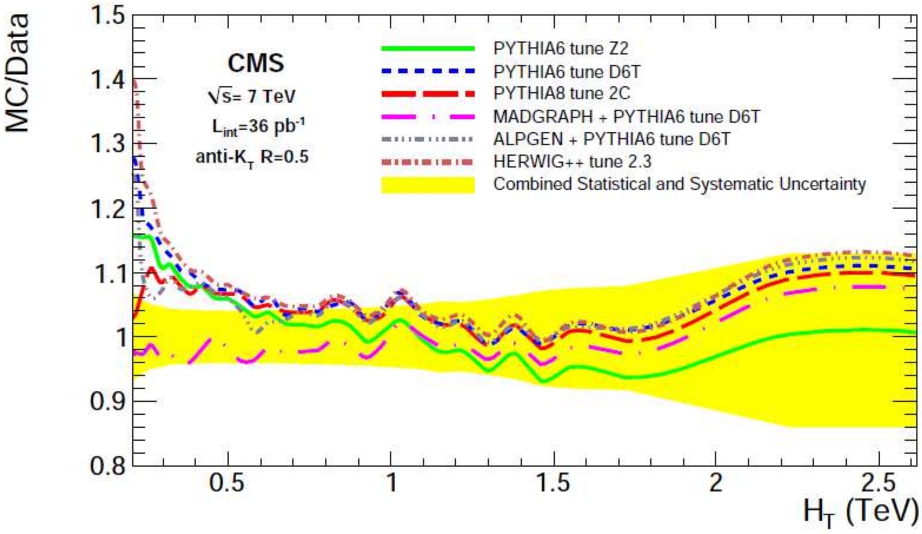


Figure 5.17: Ratios of the predicted R_{32} values from the PYTHIA , MADGRAPH , ALPGEN , and HERWIG Monte Carlo generators to the measured value, as a function of H_T. The shaded area indicates the size of the combined statistical and systematic uncertainty [187].

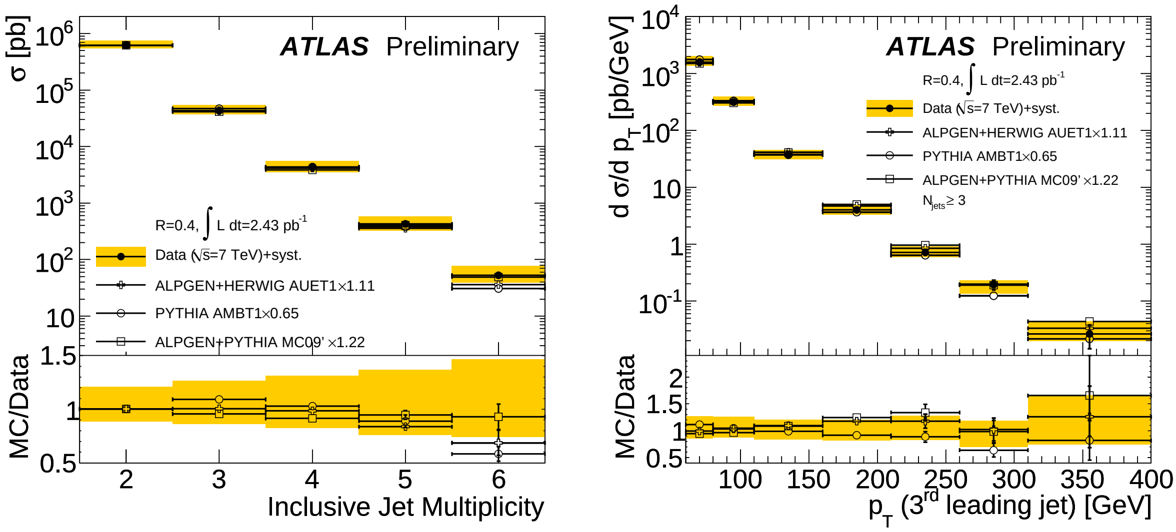
three jets have been performed and compared with Monte-Carlo generator predictions (Fig. 5.18). Results tend to show that the data are better reproduced with a matrix-element generator like ALPGEN than with a multi-purpose generator like PYTHIA .

5.2.5 Conclusions

Monte-Carlo generators are valuable tools to design physics analyses and study their performances. However, they fail to provide an accurate description of kinematic properties of the final state particles at high jet multiplicity for some of the most relevant processes in the context of top quark pair studies. These discrepancies originate mainly from the lack of QCD corrections beyond leading order. In such cases, Monte-Carlo generators should not be considered as reliable to estimate the number of selected background events and should be replaced by data-driven estimations whenever possible.

5.3 Monte-Carlo simulated event samples

In this section, samples of Monte-Carlo simulated events used throughout this thesis are reviewed. All these samples have been produced centrally by the CMS collabora-



(a) (b)

Figure 5.18: Upper plot : Measured total inclusive jet cross section as a function of the jet multiplicity (a) and differential cross section as a function of the third leading jet transverse momentum for events with at least three jets (b). Monte Carlo predictions are normalized to the measured two-jet inclusive jet multiplicity bin. The orange error bands correspond to the systematic uncertainties.
Lower plot : ratios of the different Monte Carlo simulations to the data [188].

tion during either the so-called Spring2010 or Summer2010 production round, using the CTEQ6L1 leading-order PDF's sets and the so-called D6T tune [189] for the modelling of the underlying event.

5.3.1 Default samples

The default set of samples used in this thesis consists of the following processes ;

- $t\bar{t} + jets$: top quark pair events, including the top quark decay into a b quark and a W boson, have been produced with up to three additional partons, using the MADGRAPH generator. The top quark mass was set to $m_{top} = 172.5$ GeV$/c^2$.
- $t + jets$: the three single top quark production channels have been generated separately with MADGRAPH . Samples for the s- and t-channels include the semi-leptonic decay of the W boson originating from the top quark decay, while top quarks are allowed to decay via all the possible channels in the sample for the tW production channel. The top quark mass is identical to the one used for $t\bar{t} + jets$ sample.
- $W + jets$: events containing a W boson produced in association with jets have been generated with up to four additional partons in the final state, using MADGRAPH . The event generation takes into account the W boson decay into a charged lepton and its corresponding neutrino.
- $Wc + jets$: events containing a W boson produced in association with a c quark and jets have been generated with up to three additional partons in the final state, using MADGRAPH . The event generation takes into account the W boson decay into a charged lepton and its corresponding neutrino.
- $Z + jets$: events containing a Z boson produced in association with jets have been generated with up to four additional partons in the final state, using MADGRAPH . During the generation, the Z boson was forced to decay into a pair of charged leptons with an invariant mass greater than 50 GeV$/c^2$.
- $V + Q\bar{Q}, V = W/Z$: these events, generated with MADGRAPH , contain a vector boson produced in association with a pair of b or c quarks. The final state contains up to two additional partons and include the leptonic decay of the vector boson. In case of a Z boson, the invariant mass of the lepton pair is greater than 50 GeV$/c^2$.
- $VV' + jets, V = W/Z$: events containing vector boson pairs, decaying leptonically, have been generated with up to one additonal parton in the final state, using MADGRAPH . In case of a Z boson, the invariant mass of the lepton pair is greater than 10 GeV$/c^2$.

- $pp \to \mu + X$: this multi-jet event sample has been produced by generating with PYTHIA $2 \to 2$ interactions between incoming partons. The transverse momentum of the outgoing partons in their centre-of-mass frame, $\hat{p}_T$, is greater than 20 GeV$/c$. In addition, events have been filtered at generator level to contain in the final state a muon with a transverse momentum greater than 15 GeV$/c$.

For the samples generated with MADGRAPH , the common renormalization and factorization scale was set to $\mu^2 = m_X^2 + \sum_i p_T^2$, where m_X is the top quark mass in top quark pair and single top quark events and m_X is the vector boson mass for events where a vector boson has been generated in association with jets. The sum runs over the transverse momentum of all the final state partons. The MADGRAPH generator has been interfaced with PYTHIA for the parton showering.

Table 5.4 shows an overview of the different processes considered in this thesis, together with the cross-sections used to normalize the numbers of selected events, the numbers of generated events and the corresponding integrated luminosity represented by each sample.

5.3.2 Heavy flavour mixing

Single vector boson samples, $W/Z+jets$ do not contain any heavy flavour quark in the final state of the hard interaction and need therefore to be completed with the dedicated sample $V + Q\bar{Q}$. However, it is impossible to simply add these two samples as heavy quarks might be produced during the parton showering of events contained in the $W/Z + jets$ samples, leading to a double counting of such events. To overcome such an issue, a procedure, called *Heavy flavor overlap removal* (HFOR), has been developed and validated within the CMS collaboration [191] ; the mixing procedure, identical to the MLM matching procedure, described in section 2.4.2, consists of taking events with heavy flavour quarks corresponding to well separated generator jets (cf. section 4.3.1) only from the $V + Q\bar{Q}$ sample for which explicit matrix-element calculations have been performed. Otherwise, heavy flavour quarks are considered as arising from parton showering and are taken from the $W/Z + jets$ samples. A similar procedure is applied to the $Wc + jets$ sample ; events with heavy flavour quarks which do not arise from matrix-element calculations are discarded. For each sample, a selection efficiency is calculated as the ratio of the number of simulated events to the number of simulated events kept after the HFOR. An effective

Process	Generator	Cross-section (pb)		Nb. of events	$\int \mathcal{L}$
$t\bar{t} + jets$:					
$(t \to Wb)$	MADGRAPH	157.5	(NLO)	1483404	9.4 fb^{-1}
$t + jets$, s-channel :					
$(t \to bl\nu)$	MADGRAPH	1.40	(NLO)	402055	287 fb^{-1}
$t + jets$, t-channel :					
$(t \to bl\nu)$	MADGRAPH	20.93	(NLO)	528593	25.3 fb^{-1}
$t + jets$, tW-channel :					
$(t \to bW)$	MADGRAPH	10.6	(NLO)	459589	43.4 fb^{-1}
$W + jets$:					
$(W \to l\nu)$	MADGRAPH	31314	(NNLO)	10068895	322 pb^{-1}
$Wc + jets$:					
$(W \to l\nu)$	MADGRAPH	606	(LO)	2838389	4.7 fb^{-1}
$Z + jets$:					
$(Z \to l^+l^-$, $m_{ll} > 50$ GeV$/c^2)$	MADGRAPH	3048	(NNLO)	1084921	356 pb^{-1}
$V + Q\bar{Q}$, $V = W/Z$, $Q = b/c$:					
$(W \to l\nu$, $Z \to l^+l^-)$	MADGRAPH	35.8	(LO)	936242	26.2 fb^{-1}
$VV' + jets$, $V = W/Z$:					
$(W \to l\nu$, $Z \to l^+l^-)$	MADGRAPH	7.5	(NLO)	102853	13.7 fb^{-1}
$pp \to \mu + X$ (multi-jets) :					
$(p_T^\mu > 15$ GeV$/c)$	PYTHIA	79688	(LO)	4357187	54.7 pb^{-1}

Table 5.4: Overview of the default set of simulated processes used in this thesis. Monte-Carlo generators, cross-sections, number of simulated events and corresponding amount of integrated luminosity are given. Next-to-next-to-leading order cross-sections for single vector boson production have been calculated with FEWZ [190]. Next-to-leading order cross-sections for the other processes have been calculated with MCFM , except for the $Wc + jets$ process whose cross-section has been calculated with the MADGRAPH generator.

cross-section for each sample is then calculated as the product of this selection efficiency and the cross-section calculated by the Monte-Carlo generator. The sum of these effective cross-sections is then rescaled to match the theoretical cross-section calculated at the next-to-next-to-leading order. This rescaling factor is equal to 1.36 for $W + jets$ events and to 1.31 for $Z + jets$.

5.3.3 Additional samples

Additional samples have been used in this thesis in order to evaluate the effects of different modellings of the signal and background processes. These samples, including top quark pair and single vector boson productions, have been also centrally produced by the CMS collaboration during the Spring2010 production round, using the `ALPGEN` generator. They have been generated with identical PDF's sets, underlying event tunes and factorization scales, compared to the default samples generated with `MADGRAPH` . Events with a single vector boson produced in association with jets from b or c quarks have been generated separately from events produced with jets originating from light flavour quarks and merged using the HFOR procedure.

5.4 Selection of single-muonic $t\bar{t}$+jets event candidates

As explained in section 5.1.2, the top quark pair decay channel of interest, called signal, in this thesis leads to a final state containing a muon, four jets and some missing transverse energy due to the neutrino. Unfortunately, as explained in section 5.2, other processes, called background, may lead to the same final-state. The first step of the signal selection strategy in this case consists in applying quality criteria on the reconstructed objects to ensure that they are not artefacts of the reconstruction algorithms and that their reconstructed kinematic properties are as accurate as possible. The second step consists in applying sequentially a set of topological and kinematic cuts to the final-state objects in order to increase the purity of the selected events.

5.4.1 Lepton and jet identification

The purpose of this section is to illustrate the quality criteria recommended by the CMS *Physics Object Group* concerning leptons and jets.

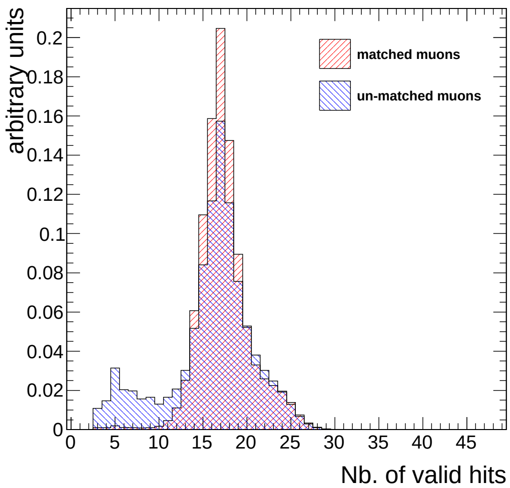
Lepton identification

With the CMS detector, muons are reconstructed with an efficiency of nearly $100\ \%$ due to the redundant informations provided both by the tracker and the muon system. This is of primary importance for the analysis presented in this thesis as the presence of an energetic muon in the final-state is a distinctive signature of the semi-muonic top quark pair decay channel. Requiring such a muon in the final-state allows to reject the overwhelming multi-jet background. It also allows to reduce the fully hadronic top quark pair decay channels. But what is also a matter of concern for this analysis is the quality and purity of the reconstructed muons. Indeed, objects wrongly reconstructed as muons or badly reconstructed muons, called fake muons, may arise from several sources ; in the case of muons from decays-in-flight of kaons or pions for instance, the association by the muon reconstruction algorithm of the tracker track from the meson itself and the muon track in the muon system would lead to a reconstructed muon with the wrong kinematic properties. In addition, fake muons may also originate from showers initiated in the hadronic calorimeter that are not totally absorbed by the calorimeter, called punch-through and therefore reach the muon system.

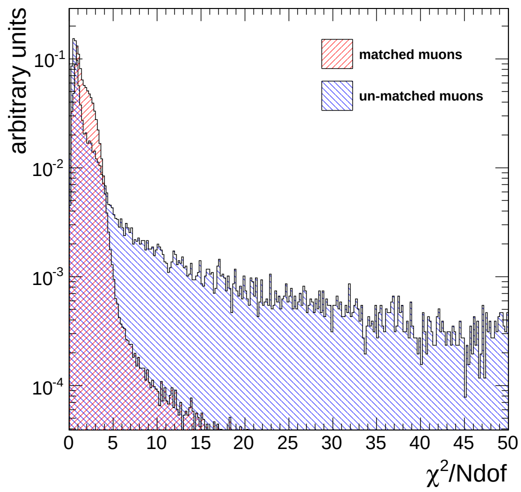
In order to distinguish between true and fake muons in Monte-Carlo simulated events, a matching in the $\eta - \phi$ space has been performed at generator level ; each reconstructed muon matching a generated muon with $\Delta R < 0.4$ is considered as a true muon (*matched muon*) or as a fake (*un-matched muon*) otherwise. Each time a generated muon is matched, it is removed from the list of the generated particles used as inputs to matching procedure in order to prevent multiple matching to the same generated muon.

Quality criteria for muons are based on the following quantities [192] :

- number of valid hits associated to the muon tracker track : tracks associated to real charged particles tend to have a higher number of valid hits than fake reconstructed tracks (Fig. 5.19a). The lower limit on the number of valid hits is set to 10.
- normalized χ^2 of the global fit performed simultaneously in the tracker and in the muon system ; for real muons, the normalized χ^2 value is closer to unity than for fake



(a)



(b)

Figure 5.19: Number of valid hits associated to reconstructed muon tracker tracks (a) and normalized χ^2 of the muon global fit (b) for matched and un-matched muons in Monte-Carlo simulated $t\bar{t} + jets$ events.

muons (Fig. 5.19b). The upper limit is set to 10. It is also required that the muon global track contains at least one valid hit in the muon system.

- muons have to be reconstructed both as global muons and tracker muons.

The performances of these variables have been tested with the first data collected by the CMS detector with proton-proton collisions at 7 TeV [193]. The overall agreement between the observed distributions for these variables and the one obtained with Monte-Carlo simulations is measured to be around $5 - 10\ \%$. However, this analysis concerns mainly muons with a low transverse momentum value. Nevertheless, this gives us confidence both in the validity of such identification strategy and in our knowledge about the muon reconstruction and identification with the CMS detector.

Jet reconstruction and identification

In this thesis, jets of particles are reconstructed using the anti-k_T clustering algorithm, with a R parameter value of 0.5 (cf. section 4.3.2). Detector noise in the calorimeter could produce fake signals reconstructed as if they were produced by particle energy deposits and therefore bias the calorimeter jet reconstruction and lead to un-physical jets. Based on studies performed with cosmic muon data (CRAFT09,[194]), a set of loose selection cuts have been defined in order to reject the fake jets :

- the fraction of the jet energy deposited in the electro-magnetic calorimeter has to be greater than 0.01, as no real jet can be detected in the hadron calorimeter only.
- the jet energy fraction contributed by the hybrid photo-diode (HPD) readout with the highest energy has to be lower than 0.98.
- the minimum number of calorimeter hits clustered into a jet which contribute to at least $90\ \%$ of the jet energy has to be greater than 1.

In this thesis, unless stated otherwise, these quality criteria are applied on any reconstructed calorimeter jets.

5.4.2 Event selection

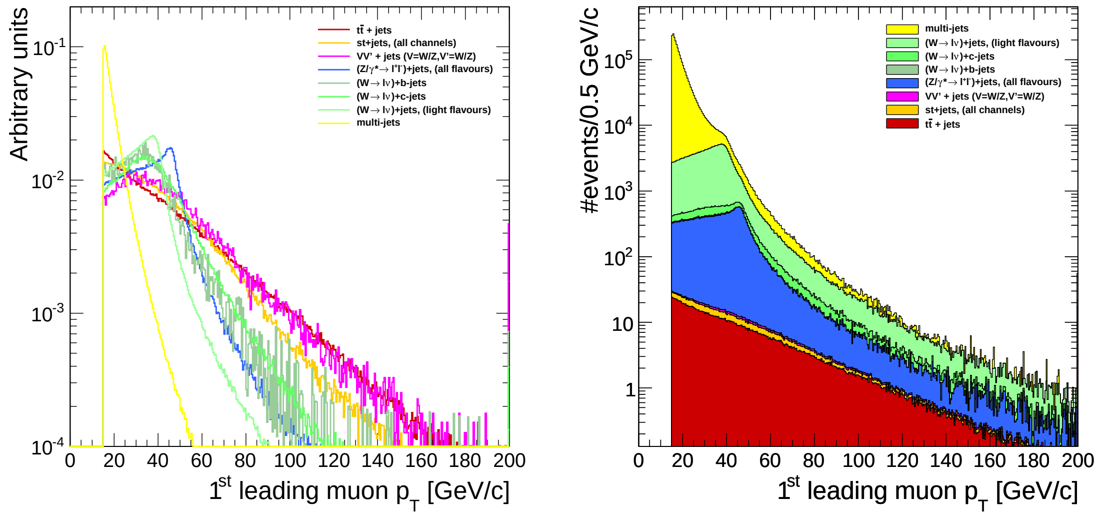
For the year 2010, the CMS detector has recorded around 40 pb^{-1} of integrated luminosity. Therefore, in order to allow an easy comparison with existing analyses as well as to demonstrate the performances of the analysis presented in this thesis with the already available data, event yields and results are rescaled to this amount of integrated luminosity whenever relevant.

Trigger : before any selection criteria are applied, events are required to be accepted by the single muon high-level trigger, *HLT_Mu9*. This trigger requires at least one level-3 muon with a transverse momentum higher than 9 GeV$/c$ and a minimal distance of 2 cm in the transverse plane with respect to the beam-spot. In addition, this level-3 muon has to be seeded by a level-2 muon with a transverse momentum higher than 7 GeV$/c$.

Primary vertex (P.V.) : the adaptive vertex fitting procedure (see section 4.1.2) is required to be successful and have more than four degrees of freedom as it is proportional to the sum of the track weights, ensuring that primary vertices are reconstructed on average with more than four highly compatible tracks. It is not possible to simply require a minimal number of tracks associated to the reconstructed vertex as the reconstruction procedure could lead to vertices reconstructed with a reasonable number of tracks although the tracks are incompatible with the vertex. Additional requirements on the primary vertex position are applied : $|z| < 15$ cm and $\rho < 2$ cm.

Muon selection : the transverse momentum distribution for muons originating from W boson[2] two-body decay exhibit a so-called Jacobian peak around a value which is half of the W boson mass (Fig. 5.20). This feature leads to a higher average muon transverse

[2] Similar behaviour is obviously observed for leptons originating from Z boson decay.



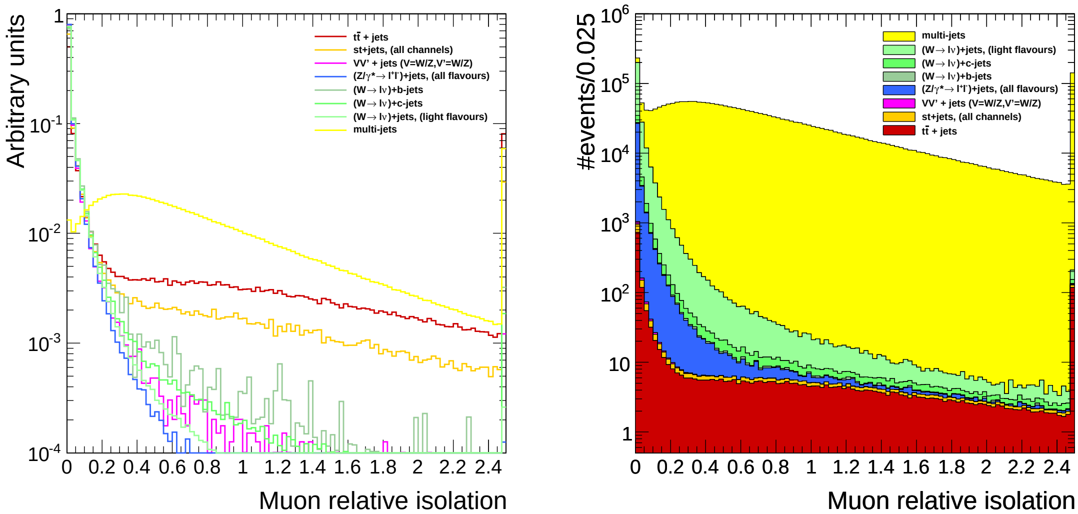
(a) **(b)**

Figure 5.20: Distribution of the leading muon transverse momentum value. Distributions are shown normalized to unity (a) and rescaled to 40 pb^{-1} of integrated luminosity (b).

momentum in this case than in the case of muons arising from multi-jet processes. However, this peak is smeared out, mainly because of the transverse momentum of the W boson. This smearing is therefore more pronounced for muons from semi-muonic top quark pair decay than for muons from single vector boson production processes. Nevertheless, final-state muons from multi-jet background processes still have on average a lower transverse momentum value than muons from $t\bar{t} + jets$ or $V + jets$ events. This allows to suppress a fraction of the multi-jet background events by imposing a lower cut of 20 GeV$/c$ on the transverse momentum value of the muon with the highest transverse momentum, called leading muon, as it is assumed that this muon in the case of semi-muonic $t\bar{t} + jets$ events is the one arising from the W boson decay.

Muon isolation : there is another consequence to the fact that the muon arises from the weak decay of a W boson for the signal and not as the decay product of hadrons contained in the collinear flow of particles from a quark or a gluon hadronization ; signal muons tend to be isolated from any hadronic activity. The degree of isolation is quantified by two variables :

- the calorimeter isolation (CaloIso) corresponds to the calorimeter activity around the muon. It is defined as the scalar sum of the transverse energy of the calorimeter towers contained in a cone of opening angle $\Delta R < 0.3$ around the muon track direction at the origin. Energy deposits due to the muon itself, contained in a cone of opening angle $\Delta R < 0.07$ and $\Delta R < 0.1$ for the electro-magnetic and the hadronic



(a) **(b)**

Figure 5.21: Distribution of the combined relative calorimeter and tracker isolation value, RelIso, for the leading muon. Distributions are shown normalized to unity (a) and rescaled to 40 pb^{-1} of integrated luminosity (b).

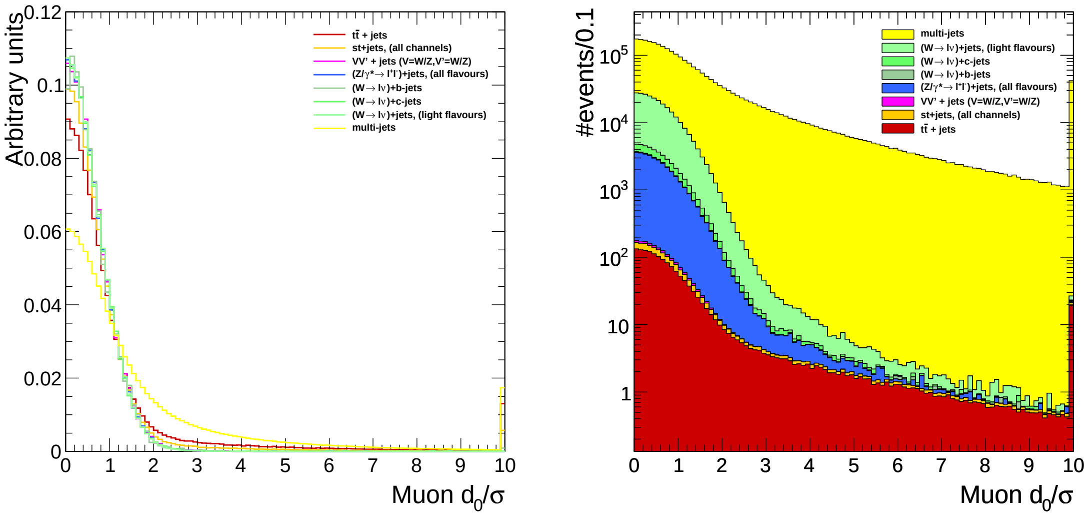
calorimeters respectively are removed from the sum. These opening angle values have been chosen in order to approximatively match a 3×3 matrix of crystals in the electro-magnetic calorimeter barrel and approximatively one tower in the hadronic calorimeter respectively.

- the tracker isolation (TrackIso) corresponds to the tracker activity around the muon. It is defined as the scalar sum of the track transverse momenta for all the tracks contained in a cone of opening angle $\Delta R < 0.3$ around the muon track direction at the origin. The muon track itself is excluded from the sum.

Studies performed in the Electro-weak Physics Analysis Group and in the Top Physics Analysis Group have shown that the rejection power against sources of non-isolated muons, like multi-jet background processes, increases when using relative isolation variables with respect to the muon transverse momentum. Additional rejection power may be obtained by combining these relative variables into a single variable, RelIso, defined as :

$$RelIso = \frac{CaloIso + TrackIso}{p_T^{\mu}} \tag{5.3}$$

In this analysis, muons are considered as isolated if their relative isolation value is lower than 0.1 (Fig. 5.21).



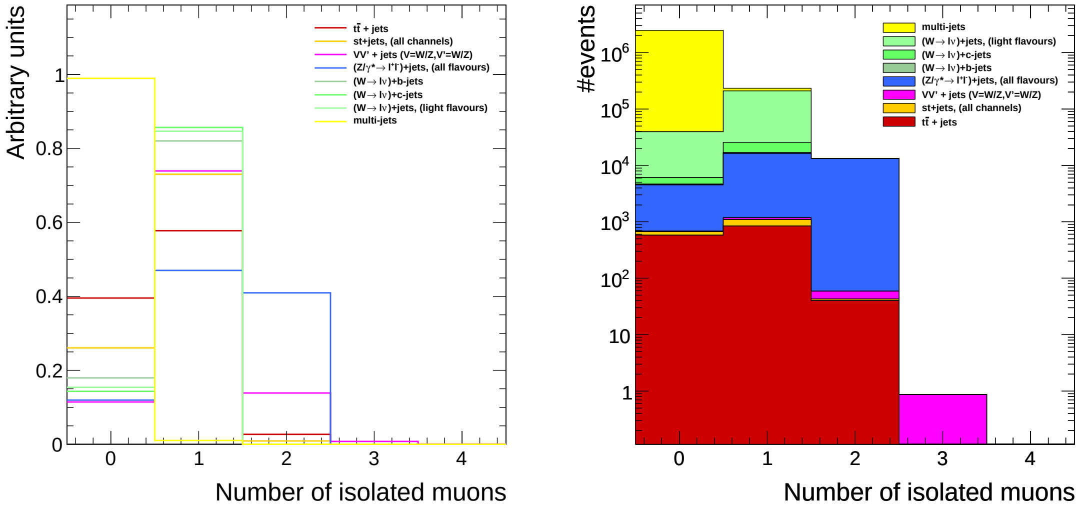
(a) **(b)**

Figure 5.22: Distribution of the leading muon transverse impact parameter significance. Distributions are shown normalized to unity (a) and rescaled to 40 pb^{-1} of integrated luminosity (b).

Finally, as explained in section 5.2.4, muons arising from long-lived particles have on average a larger impact parameter significance than muons which originate from the prompt decay of a W boson (Fig. 5.22). Therefore, this variable is helpful to reject multi-jet events. In this thesis, events are rejected if the transverse impact parameter significance of the leading muon is higher than 3.

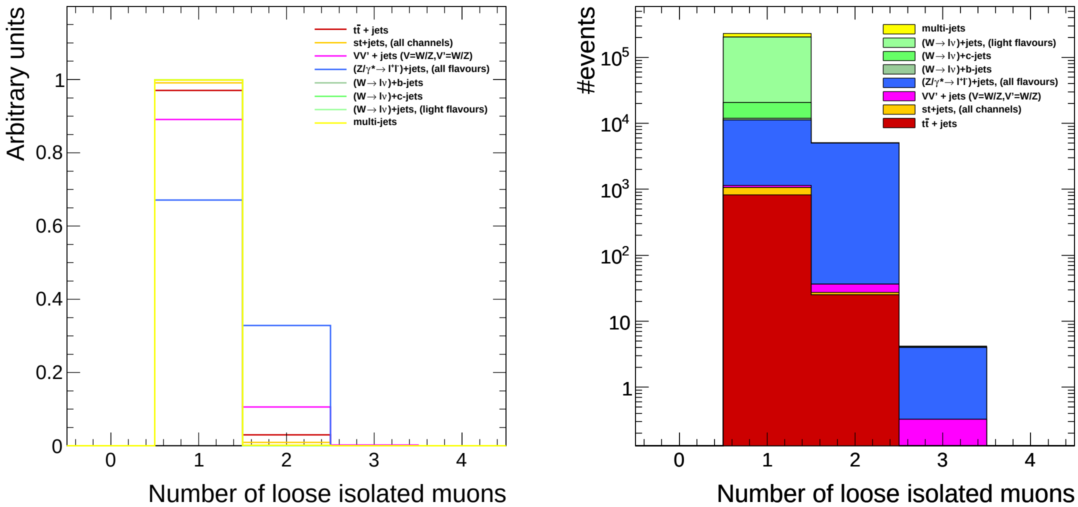
These requirements on the muon transverse momentum, relative isolation and transverse impact parameter significance allow to highly suppress multi-jet background events as well as events where the top quark pair decays via the fully hadronic channel. However, as explained in section 5.2, the production of vector boson associated with jets cannot be rejected in a similar manner as the leptonic decay of a vector boson leads to isolated leptons. Nevertheless, it is already possible to partially reject events where jets have been produced in association with a Z boson decaying into two muons by requiring exactly one such isolated muon (Fig. 5.23). This requirement rejects also events where leptonically decaying W or Z boson pairs have been produced.

Additional lepton veto : events where the top quark pair decays via the di-leptonic decay channel are also a source of background events as isolated leptons are produced in the final state. In order to reject these events, events with a second isolated muons (Veto 1, Fig. 5.24) or electrons (Veto 2, Fig. 5.25) are vetoed, with relaxed kinematic ($p_T > 10$ GeV$/c$ for muons and $E_T > 15$ GeV for electrons) and isolation criteria ($RelIso < 0.2$).



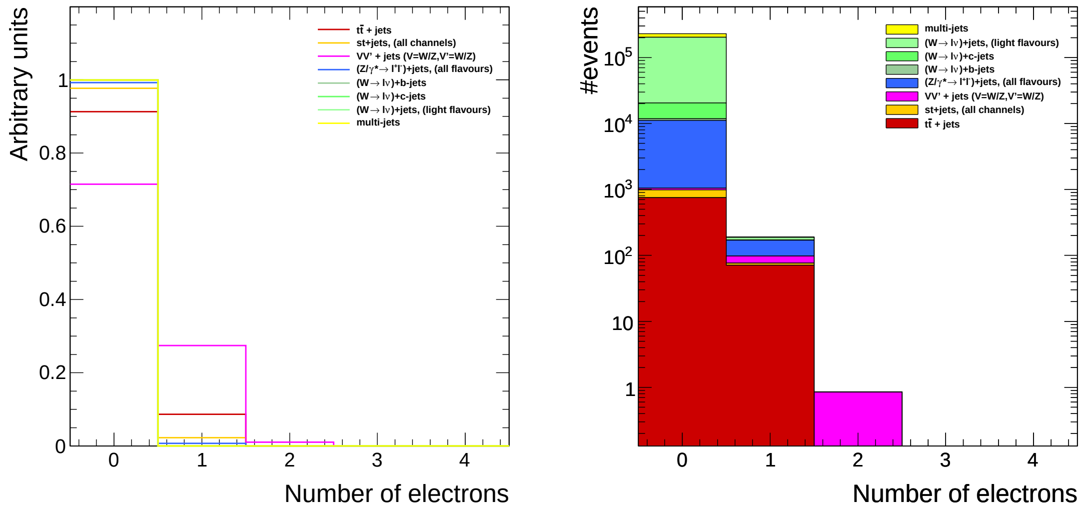
(a) (b)

Figure 5.23: Distribution of the number of isolated muon with a transverse momentum value higher than 20 GeV$/c$. Distributions are shown normalized to unity (a) and rescaled to 40 pb^{-1} of integrated luminosity (b).



(a) (b)

Figure 5.24: Distribution of the number of muons with a transverse momentum value higher than 10 GeV$/c$ and a relative isolation value lower than 0.2. Distributions are shown normalized to unity (a) and rescaled to 40 pb^{-1} of integrated luminosity (b).

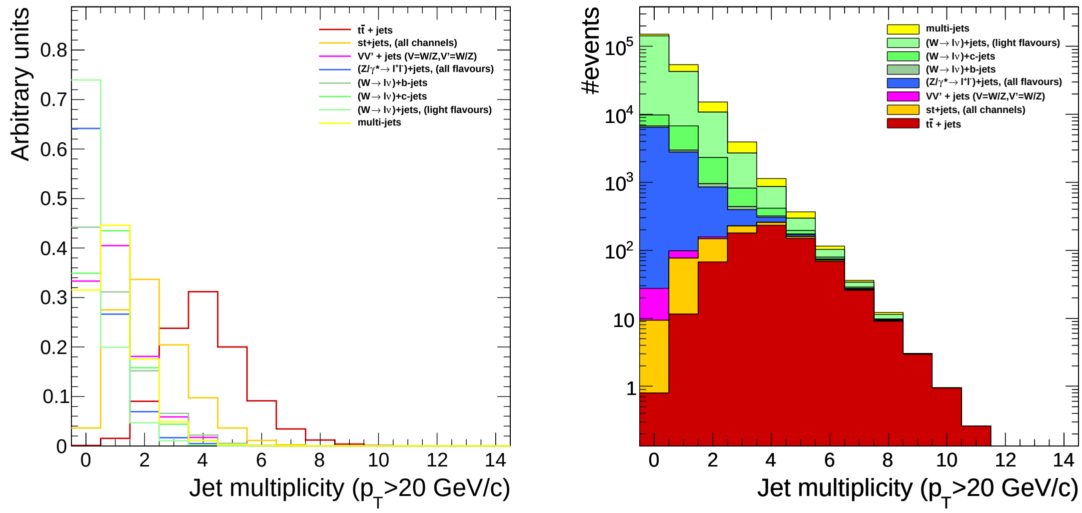


(a) **(b)**

Figure 5.25: Distribution of the number of electrons with a transverse energy value higher than 15 GeV$/c$ and a relative isolation value lower than 0.2. Distributions are shown normalized to unity (a) and rescaled to 40 pb^{-1} of integrated luminosity (b).

Jet selection : for events where the top quark pair decays via the semi-muonic decay channel, the four final-state quarks ideally lead to four reconstructed jets in addition to jets arising from initial and/or final-state radiations. In the case of vector boson production decaying leptonically, jets enter the final-state only via the emission of additional gluons and therefore are suppressed by a factor proportional to the strong coupling constant value. Therefore, assuming that the jet reconstruction efficiency does not strongly depend on the jet multiplicity of the event, the number of reconstructed jets tends to be higher for semi-muonic top quark events than for background events (Fig. 5.26).

Likewise the muons, final-state jets originating from the quarks produced by top quark pair decays have on average a higher transverse momentum value than jets originating from background processes (Fig. 5.27). Top quark events are thus preferentially selected by applying a threshold on the reconstructed jet transverse momentum. Within the CMS Top Physics Analysis Group, this threshold on the transverse momentum of the reconstructed jets is set to 30 GeV$/c$. However, as also shown in Figure 5.27d, applying such a threshold on the fourth leading jet transverse momentum lead to reject a significant fraction of top quark events. While increasing the purity of the selected event sample, this rejection also leads to an increase of the statistical uncertainty on quantities derived from the selected signal events. Nevertheless, in this thesis, a 30 GeV$/c$ threshold on jet transverse momenta has been chosen. Background estimation methods developed in this thesis could not



(a) (b)

Figure 5.26: Distribution of the number of reconstructed jet with a transverse momentum value higher than 20 GeV$/c$. Distributions are shown normalized to unity (a) and rescaled to 40 pb^{-1} of integrated luminosity (b).

only be used to measure the top quark pair production cross-section for instance but also to normalize distributions of variables such as the top quark mass or transverse momentum. In such cases, it is essential to ensure that the jets from the top quark decay are correctly reconstructed. But jet energy corrections are under-estimated for low transverse momentum jets from top quark pair decays (Fig. 5.28a). On average, the jet response, obtained by calculating, for each jet, the ratio of its transverse momentum value to the transverse momentum value of its closest parton in the $\eta - \phi$ space, deviates all the more from unity that the jet transverse momentum is low. Furthermore, the reconstructed jet transverse momentum resolution (Fig. 5.28b) and angular resolutions (Fig. 5.28c and Fig. 5.28d) are negatively correlated with the jet transverse momentum value for jets from top quark decays. The jet transverse momentum resolution is defined as the width of the jet response distribution. The width, obtained as the standard deviation of a Gaussian fit, is then divided by the mean of the jet response in order not to be biased by the jet energy mis-calibration. The angular resolutions are obtained in a similar way. From a general point of view, both the jet transverse momentum and angular resolutions improve as the jet transverse momentum value increases. The difference in resolution between the pseudo-rapidity and the azimuthal angle comes from the fact that, for the latter, particles produced during the hadronization are deflected from the original parton direction by the magnetic field. Similar behaviours have been observed whether the jets are reconstructed

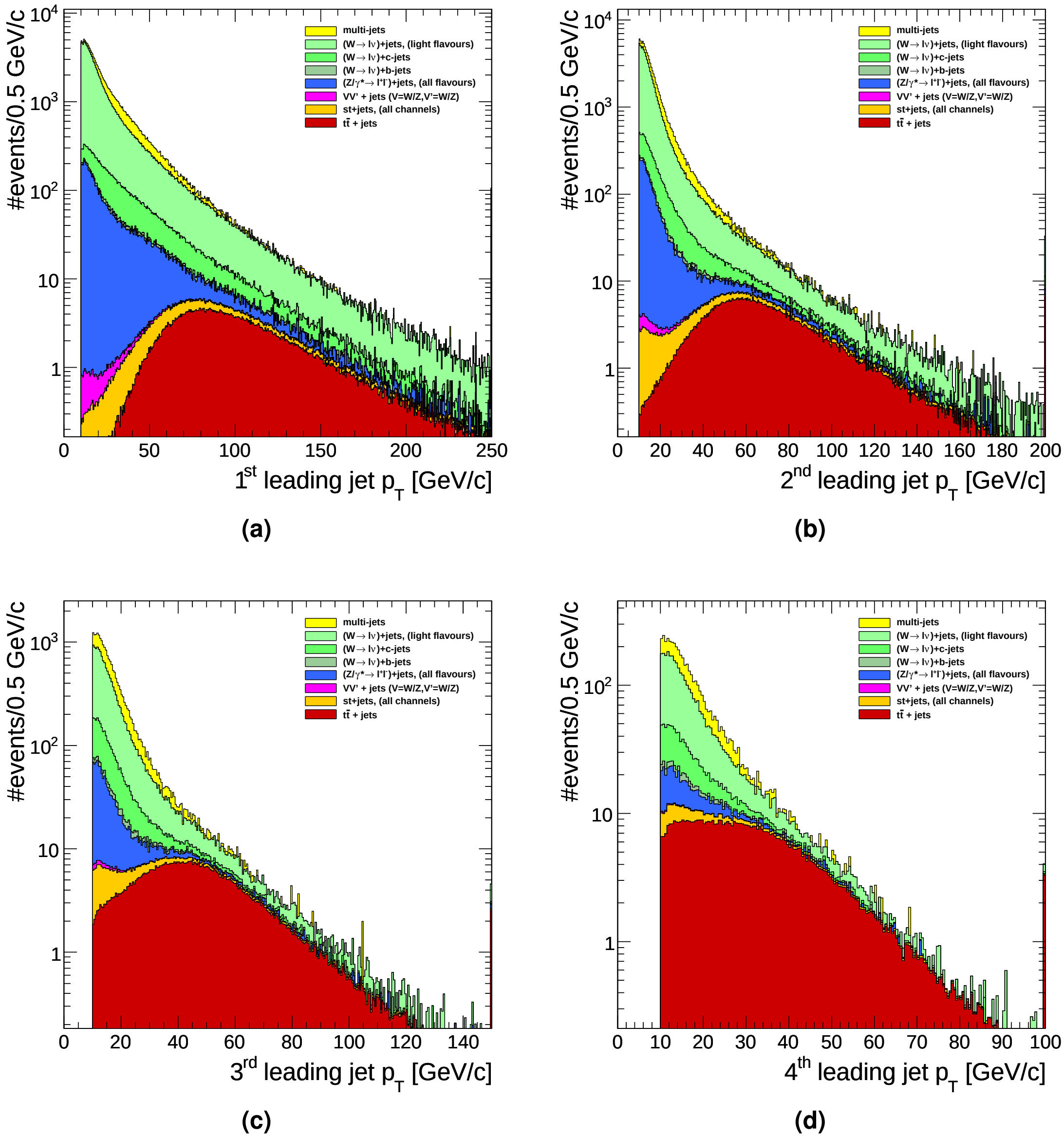


Figure 5.27: Distribution of the transverse momentum value of the first leading jet (a), second leading jet (b), third leading jet (c) and fourth leading jet (d) rescaled to 40 pb^{-1} of integrated luminosity (b). The N^{th} leading jet transverse momentum distribution takes into account a lower cut of 30 GeV$/c$ on the transverse momentum value of the $(M < N)^{th}$ leading reconstructed jets.

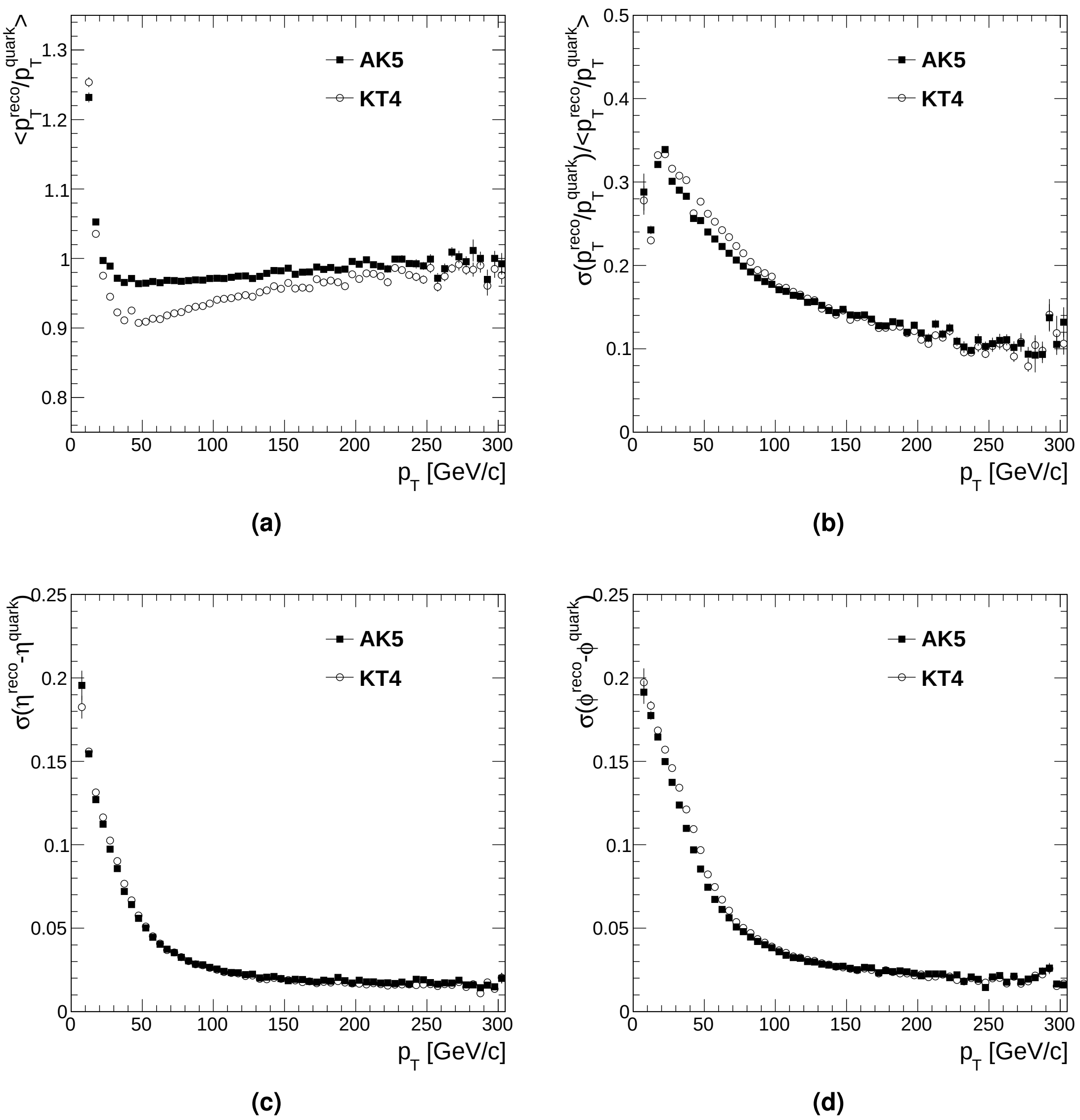


Figure 5.28: Mean jet energy response (a), jet transverse momentum resolution (b), jet pseudorapidity resolution (c) and jet azimuthal angle resolution (d) as a function of the generated jet transverse momentum for top quark pair events. Jets have been reconstructed either with the anti-k_T algorithm, using a R value of 0.5 (AK5) or with the k_T algorithm, using a R value of 0.4 (KT4).

with the anti-k_T algorithm using a value of 0.5 for the R parameter (AK5) or with the k_T algorithm, using a value of 0.4 for the R parameter (KT4).

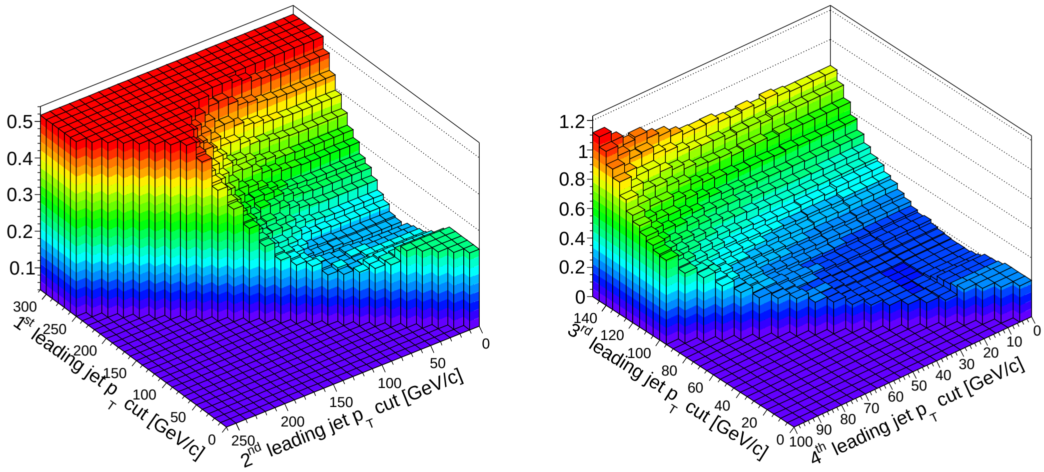
Asymmetrical cuts on jet transverse momentum

As future improvements in the understanding of the CMS detector performances will result in reduced uncertainties on the jet energy corrections factors and on the jet energy resolutions for jets with low transverse momentum and thus allow analyses to use reconstructed jets with lower transverse momentum thresholds than the current ones, the possibility to apply asymmetrical cuts on the jet transverse momentum has been investigated. As shown in Figure 5.27a, a higher cut than 30 GeV$/c$ on the first leading jet would allow to reject more background events while a lower cut on the third and fourth leading jet transverse momentum (Fig. 5.27c and Fig. 5.27d) would allow to reject less signal events. The best combination of jet transverse momentum cuts depends on the precision of the background estimations. As an example, the uncertainty on the estimated number of top quark pair events due to the correction for the remaining multi-jet events has been calculated as a function of the thresholds applied on the leading jet transverse momentum. The best combination is the combination minimizing this uncertainty. Assuming that the numbers of selected events follow a Poisson distribution, this uncertainty is expressed as following :

$$\frac{\Delta N_{t\bar{t}+jets}}{N_{t\bar{t}+jets}} = \frac{\sqrt{(\Delta N_{total})^2 + (\Delta N_{multi\text{-}jets})^2}}{N_{t\bar{t}+jets}} = \frac{\sqrt{N_{total} + ((1+(X/100)^2)N_{multi\text{-}jets})^2}}{N_{t\bar{t}+jets}} \tag{5.4}$$

where N_{total} is the total number of selected events and X is the percentage of systematic errors on the number of selected multi-jet events, $N_{multi\text{-}jets}$. Events have been selected with the default set of cuts, described in the previous except that the jet transverse momentum threshold has been lowered from 30 GeV$/c$ to 15 GeV$/c$. As an example, the uncertainty has been calculated for $X = 25$ %, varying the cuts on the jet transverse energy up to 400 GeV$/c$ for the two leading jets (Fig. 5.29a), by steps of 10 GeV$/c$ and up to 150 GeV$/c$ for the next leading jets, by steps of 5 GeV$/c$ (Fig 5.29b).

Table 5.5 shows the number of selected numbers of $t\bar{t} + jets$, $V + jets$ and multi-jet events for the best combination of cuts, as well as the uncertainty on the number of estimated $t\bar{t} + jets$ events after the full event selection. As expected, a lower uncertainty on the estimated number of $t\bar{t} + jets$ than the one obtained with a unique jet transverse momentum cut of 30 GeV$/c$ by applying asymmetrical cuts. While cuts higher than 30 GeV$/c$ are preferred for the two leading jets, cuts slightly lower than 30 GeV$/c$ are preferred for the next two leading jets. It has been also checked that modifications of the



(a) **(b)**

Figure 5.29: Uncertainty on the number of estimated $t\bar{t}+jets$ events, due to the multi-jet correction, assuming a systematic uncertainty of $25\ \%$ on that correction and corresponding to 40 pb^{-1} of integrated luminosity, as a function of the cut of the transverse momentum of the two leading jets (a) and of the next two leading jets (b). Cuts on the jet transverse momentum that are not used in the plots are set to their optimal values.

Jet multiplicity	Jet p_T cuts (GeV$/c$)	$\frac{\Delta N_{t\bar{t}+jets}}{N_{t\bar{t}+jets}}$	$t\bar{t}+jets$	$V+jets$	multi-jets
≥ 3	$30/30/30$	$56.9\ \%$	571.7	913.7	303.3
	$70/70/35$	$15.6\ \%$	173.5	160.4	18.3
≥ 4	$30/30/30/30$	$16.6\ \%$	319.4	179.5	44.6
	$70/50/25/25$	$12.2\ \%$	194.9	114.8	14.6

Table 5.5: Uncertainty on the number of estimated $t\bar{t}+jets$ events, due to the multi-jet correction, assuming a systematic uncertainty of $25\ \%$ on that correction for the optimal set of cuts on the jet transverse momentum for events with 3 or at least 4 reconstructed jets with a transverse momentum greater than 15 GeV$/c$. The number of $t\bar{t}+jets$, $V+jets$ and multi-jets, remaining after these cuts, are also shown for an integrated luminosity of 40 pb^{-1}.

level of systematic uncertainty, X, on the estimated number of multi-jet events, do not change the conclusions of this study ; asymmetrical cuts remain preferred.

Summary and event yields

Table 5.6 summarizes the expected numbers of selected events after the different selection criteria, using the default set of Monte-Carlo simulated samples, documented in section 5.3. Results have been rescaled to corresponds to an integrated luminosity of 40 pb^{-1}.

Before the muon selection, the sample of selected events based on the muon trigger decision is dominated by the multi-jet background. However, as expected, requiring an isolated muon with a transverse momentum value higher than 20 GeV$/c$ allows an efficient rejection of such background by a factor ~ 100. The veto on an additional isolated muon allows to discard about one third of the remaining events where a Z boson, mostly decaying into two muons, has been produced in association with jets while the veto on an additional isolated electrons do not affect the number of selected background events but reject part of the remaining events where top quark pairs decay via the di-lepton channel. Along these cuts, the ratio of single top events to top quark pair events remains approximatively the same (~ 0.25). At this stage of the event selection procedure, the sample of selected events is dominated by the production of vector boson in association with jets and, to a lesser extent, by the multi-jet background. However, it is possible to decrease the number of such background events to the order of the number of top quark pair events by requiring events to have at least three reconstructed jets with a transverse momentum higher than 30 GeV$/c$. It has also been shown that the number of multi-jet events can be decreases further by applying asymmetrical cuts on the jet transverse momentum. When requiring at least three jets with a transverse momentum higher than 30 GeV$/c$, the signal selection efficiency is then $\sim 9\%$, for a signal-over-background ratio of 0.45 and a signal purity of 31%. The signal selection efficiency decreases down to $\sim 5\%$ if at least four reconstructed jets are required, leading to a signal-over-background ratio of 1.34 and a signal purity of 57%.

	$t\bar{t}$+jets	st+jets	VV'+jets	Z+jets	W+b-jets	W+c-jets	W+light-jets	multi-jets
#events	$6300.0^{+0.0}_{-1.0}$	$1317.2^{+0.0}_{-1.0}$	$300.0^{+0.0}_{-1.0}$	$121924.0^{+0.0}_{-1.0}$	$4201.2^{+0.0}_{-1.0}$	$47303.6^{+0.0}_{-1.0}$	$1201056.0^{+0.0}_{-1.0}$	$3187520.0^{+0.0}_{-1.0}$
Trigger	$1884.7^{+36.5}_{-36.1}$	$404.5^{+16.9}_{-16.5}$	$125.5^{+8.6}_{-8.4}$	$35930.9^{+159.4}_{-159.0}$	$930.9^{+27.2}_{-26.6}$	$11337.9^{+93.1}_{-92.6}$	$244666.5^{+441.7}_{-441.1}$	$2540536.8^{+717.8}_{-718.4}$
P.V.	$1883.7^{+36.5}_{-36.1}$	$404.2^{+16.9}_{-16.5}$	$125.3^{+8.6}_{-8.4}$	$35859.6^{+159.3}_{-158.9}$	$929.8^{+27.2}_{-26.6}$	$11332.0^{+93.1}_{-92.6}$	$243858.9^{+441.1}_{-440.6}$	$2539735.0^{+718.1}_{-718.8}$
Muon sel.	$846.2^{+27.4}_{-26.7}$	$244.5^{+14.4}_{-13.8}$	$85.1^{+8.0}_{-7.6}$	$15071.0^{+115.3}_{-114.6}$	$660.6^{+23.9}_{-23.3}$	$8683.4^{+84.5}_{-83.9}$	$180710.7^{+392.2}_{-391.5}$	$25082.8^{+158.2}_{-157.3}$
Veto 1	$821.0^{+27.1}_{-26.4}$	$242.3^{+14.4}_{-13.7}$	$75.9^{+7.8}_{-7.3}$	$10112.7^{+96.7}_{-95.9}$	$659.7^{+23.9}_{-23.2}$	$8679.0^{+84.5}_{-83.9}$	$180639.8^{+392.1}_{-391.4}$	$25048.4^{+158.1}_{-157.2}$
Veto 2	$749.7^{+26.1}_{-25.3}$	$236.8^{+14.3}_{-13.6}$	$54.3^{+7.0}_{-6.3}$	$10040.0^{+96.4}_{-95.6}$	$659.7^{+23.9}_{-23.2}$	$8677.4^{+84.5}_{-83.9}$	$180624.2^{+392.1}_{-391.4}$	$25044.8^{+158.1}_{-157.1}$
$\geq$ 1 jets	$746.7^{+26.0}_{-25.3}$	$218.3^{+13.8}_{-13.2}$	$28.8^{+5.5}_{-4.7}$	$2253.3^{+47.5}_{-46.5}$	$247.4^{+15.7}_{-14.8}$	$3737.5^{+59.1}_{-58.2}$	$24993.9^{+156.9}_{-156.0}$	$10184.7^{+101.3}_{-100.3}$
$\geq$ 2 jets	$712.5^{+25.5}_{-24.8}$	$133.4^{+11.4}_{-10.6}$	$8.1^{+3.3}_{-2.4}$	$375.0^{+19.8}_{-18.8}$	$82.1^{+9.5}_{-8.5}$	$776.4^{+28.1}_{-27.2}$	$4059.5^{+64.1}_{-63.1}$	$1910.1^{+44.2}_{-43.2}$
$\geq$ 3 jets	$571.7^{+23.2}_{-22.4}$	$49.9^{+7.4}_{-6.5}$	$1.7^{+1.9}_{-0.9}$	$71.4^{+9.0}_{-8.0}$	$22.6^{+5.3}_{-4.3}$	$140.3^{+12.3}_{-11.3}$	$679.4^{+26.6}_{-25.6}$	$303.3^{+18.0}_{-17.0}$
$\geq$ 4 jets	$319.4^{+17.9}_{-17.0}$	$14.8^{+4.3}_{-3.4}$	$0.3^{+1.2}_{-0.2}$	$14.5^{+4.3}_{-3.3}$	$5.4^{+2.9}_{-1.9}$	$24.7^{+5.5}_{-4.5}$	$134.9^{+12.1}_{-11.1}$	$44.6^{+7.2}_{-6.2}$
$\geq$ 5 jets	$116.3^{+11.2}_{-10.2}$	$3.5^{+2.4}_{-1.4}$	$0.0^{+1.0}_{-0.0}$	$2.7^{+2.2}_{-1.2}$	$0.8^{+1.5}_{-0.5}$	$3.9^{+2.5}_{-1.5}$	$23.9^{+5.4}_{-4.4}$	$6.6^{+3.1}_{-2.1}$
$\geq$ 6 jets	$34.5^{+6.4}_{-5.4}$	$0.6^{+1.4}_{-0.4}$	$0.0^{+1.0}_{-0.0}$	$1.0^{+1.6}_{-0.6}$	$0.1^{+1.1}_{-0.1}$	$0.5^{+1.4}_{-0.4}$	$4.5^{+2.7}_{-1.7}$	$1.5^{+1.8}_{-0.8}$

Table 5.6: Expected numbers of selected events and associated statistical uncertainties, rescaled to an integrated luminosity of 40 pb^{-1}.

6

Multi-jet background estimation for semi-muonic $t\bar{t}$ studies

As in the past for preceding hadron collision experiments, multi-jet processes will be an important source of background for most channels that will be studied at the LHC, due to their huge cross-sections, typically $\mathcal{O}(10^9\ pb)$ for collisions at a centre-of-mass energy of 7 TeV, several orders of magnitude larger than the top quark pair production, for instance. The dominant multi-jet processes lead to jets with low transverse momentum in the final state and usual selection cuts suppress them strongly. However, due to the huge cross-sections, even tiny selection efficiencies still constitute a potential danger. Moreover, the remaining multi-jet background is most of the time difficult to estimate because of the poor theoretical knowledge of the multi-jet cross-section and the difficulty to simulate enough multi-jet events to estimate their selection probability with sufficient precision.

This chapter presents a possible method, the *ABCD* method, to estimate, from the data, the multi-jet background remaining after selection of semi-muonic top quark pair events. The principle of the method is explained in section 6.1 together with the choice of the input variables for the method as well as a procedure to check the statistical dependence between these variables. Then, the performances of the method are evaluated, using only a sample of multi-jet events simulated at a collision energy of 7 TeV in section 6.2, and using all the samples, including $t\bar{t}+jets$ and $V+jets$ simulated samples, in section 6.3. Finally, in section 6.4, the systematic uncertainties on the estimated number of multi-jet events are assessed.

6.1 The *ABCD* method

6.1.1 Principle of the method

This method aims to evaluate, from the data, the number of events from a particular background process in a given sample of selected events. It consists in identifying two statistically independent variables, X and Y, which discriminate between background events and other events, called here signal events. By applying a cut, X_0 and Y_0, on each of the variables, one defines four different regions (A,B, C and D) in the XY plane (Fig. 6.1). If the cuts are such that the signal is dominant in one of these regions, for example D, while the three others, called control regions, are background dominated, then the number of background events in the region D, N_D^{bckgd}, can be estimated by counting the number of events observed in the regions A,B and C where $N_\alpha \simeq N_\alpha^{bckgd}$ is assumed, with $\alpha =$A, B or C :

$$N_D^{bckgd} = \frac{N_B^{bckgd} \cdot N_C^{bckgd}}{N_A^{bckgd}} \simeq \frac{N_B \cdot N_C}{N_A} \tag{6.1}$$

The statistical uncertainty on the estimated number of events has been calculated assuming that the joint distribution of N_A, N_B, N_C and N_D is a multinomial distribution for which the probability for an event to belong to the region α, p_α, has been estimated using the ratio N_α/N_{total}, with $N_{total} = N_A + N_B + N_C + N_D$. The correlations between the numbers of events in two different regions have also been taken into account.

6.1.2 Choice of the variables

As the estimation made by the ABCD method relies on the possibility to define simultaneously a signal dominated region and three background dominated ones, the chosen variables must discriminate as much as possible against this particular kind of background events. Attempts to find satisfactory variables have been made, using variables related to

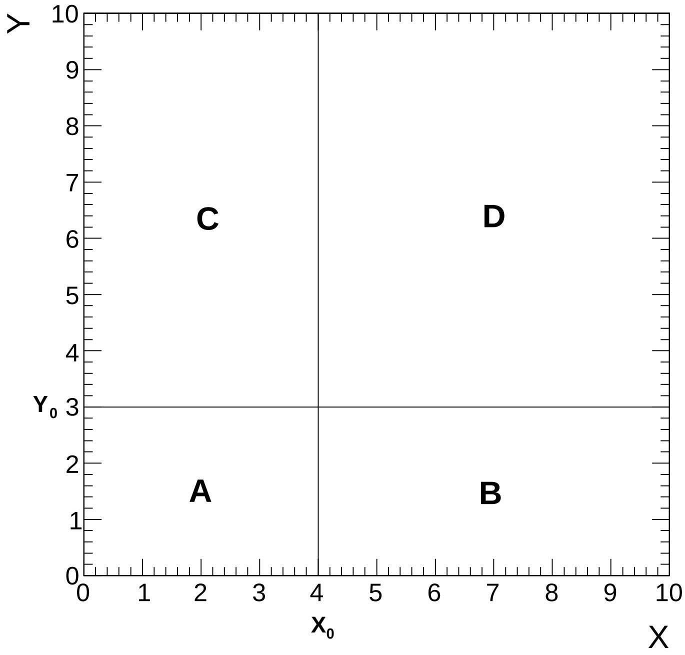


Figure 6.1: Definition of the regions in the ABCD method in the XY plane.

either the selected muon or the selected jets. Among all the processes considered in this thesis (cf. section 5.2), the multi-jet process is the only one that does not lead in its final state to a muon from the decay of a vector boson. As discussed in section 5.4.2, selected muons in multi-jet events tend thus to be less energetic, less isolated and have a larger transverse impact parameter significance than the selected muons from other processes. Similarly, jets from multi-jet events also tend to have a lower transverse momentum. The muon transverse momentum, p_T^{μ}, the muon relative isolation, RelIso, and the muon transverse impact parameter significance, d_0/σ, are therefore candidates for the ABCD method variables, together with the transverse momentum of the third or fourth jet, $p_T^{3rd/4th\ jet}$, and the scalar sum of the transverse momentum of the selected jets, H_T.

All the possible combinations of two variables have been tested with the ABCD method. However, as $p_T^{3rd/4th\ jet}$ and H_T are obviously correlated, this pair of variable has been discarded :

- p_T^{μ} and d_0/σ
- p_T^{μ} and RelIso
- p_T^{μ} and $p_T^{3rd/4th\ jet}$
- p_T^{μ} and H_T
- d_0/σ and RelIso
- d_0/σ and $p_T^{3rd/4th\ jet}$
- d_0/σ and H_T
- RelIso and $p_T^{3rd/4th\ jet}$
- RelIso and H_T

All the selection cuts defined in the previous chapter have been applied, except for those applied on the variables used in the ABCD method. Depending on the fact that the method aims to estimate the number of multi-jet events with at least three or four jets, either $p_T^{3rd\ jet}$ or $p_T^{4th\ jet}$ is used. The transverse momentum threshold for this particular jet is then lowered from 30 GeV/c to 20 GeV/c.

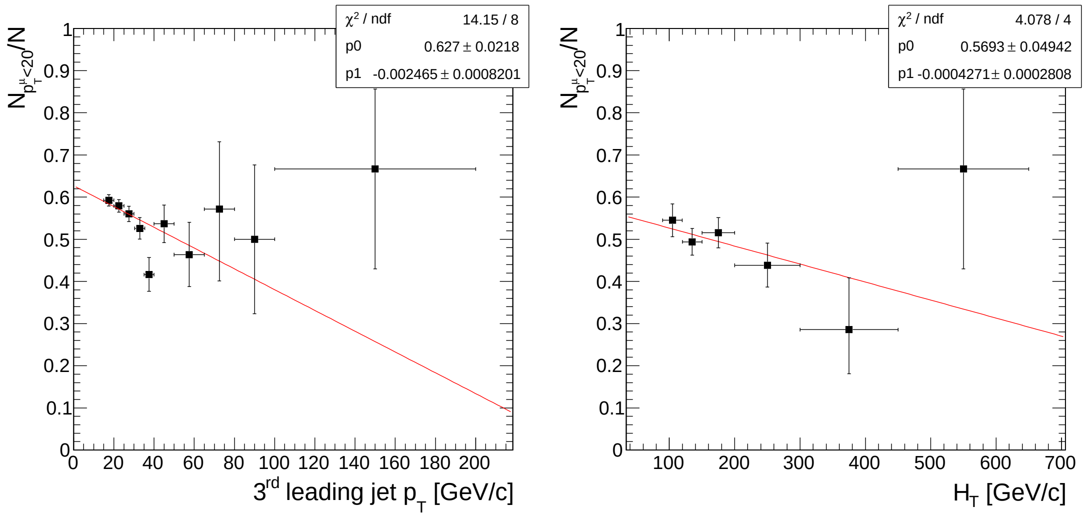
6.1.3 Control of the statistical dependence between variables

As explained in the previous section, the estimation made by the ABCD method is valid only if the two variables X and Y are independent. A simple test of this assumption consists in dividing the range of each variable, for example X, into several intervals and for each interval, calculate the fraction of events satisfying the cut in the other variable, $Y > Y_0$. If the two variables are independent, this fraction must be constant over the X range and this, whatever the values chosen for Y_0.

When using data, it is possible to test, at least partially, the assumption of independence between the two variables by calculating the aforementioned fractions for intervals contained in X (or Y) range expected to be dominated by the multi-jet background. In addition, it is possible to compare the fractions obtained with the data to the expected one from Monte-Carlo simulations. Although a constant fraction as well as an agreement between the obtained and the expected fraction do not guarantee that this fraction would remain constant in the signal region, it would nevertheless give us confidence in the possibility to rely on this method to estimate the multi-jet background.

This test has been performed on each pair of variables listed in the previous section, using simulated multi-jet events.

For events with at least three reconstructed jets within the detector acceptance with a transverse momentum greater than 30 GeV/c, the fraction of multi-jet events with $p_T^{\mu} <$ 20 GeV/c has been calculated as a function of $p_T^{3rd\ jet}$ (Fig. 6.2a) and H_T (Fig. 6.2b). In



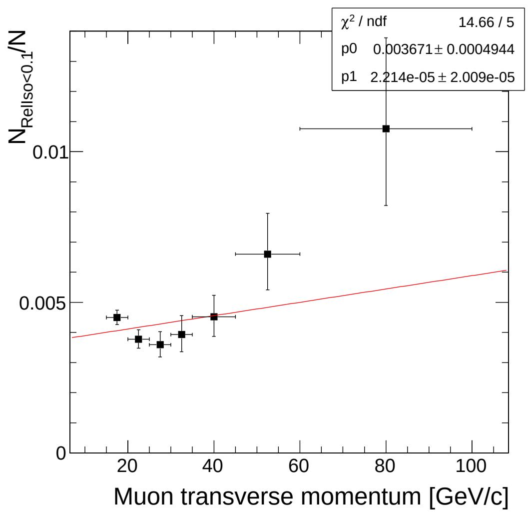
(a) **(b)**

Figure 6.2: Fraction of multi-jet events with a muon transverse momentum lower than 20 GeV$/c$ as a function of the third leading jet transverse momentum (a) and as a function of the H_T variable (b). Multi-jet events have been selected if they contain at least two (a) or three (b) reconstructed jets with a transverse momentum greater than 30 GeV$/c$ in the detector acceptance. Uncertainties have been calculated for an integrated luminosity of 40 pb^{-1}.

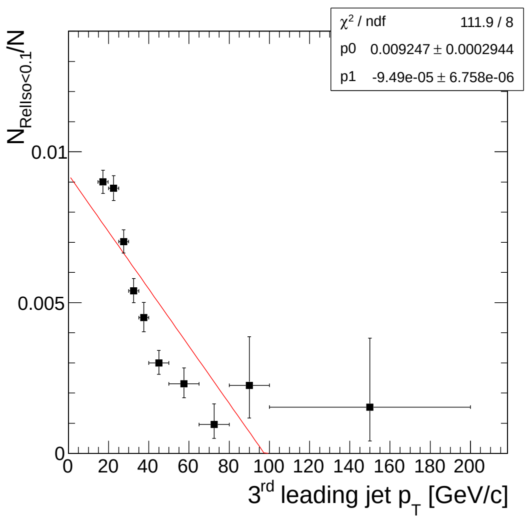
both cases, it has been observed that this fraction decreases when the binned variable increases.

The fraction of multi-jet events with an isolated muon, $RelIso < 0.1$, has also been calculated as a function of p_T^{μ} (Fig. 6.3a), $p_T^{3rd\ jet}$ (Fig. 6.3b) and H_T (Fig. 6.3c). In the first case, as the relative isolation is inversely proportional to the muon transverse momentum, the fraction of isolated muons increases when p_T^{μ} increase. In the latter cases, as $p_T^{3rd\ jet}$ or H_T increases, the activity in the calorimeter is enhanced, leading to less isolated muons and therefore to smaller fractions of events with an isolated muons. None of these pairs of variables has thus been considered as a candidate for the ABCD method.

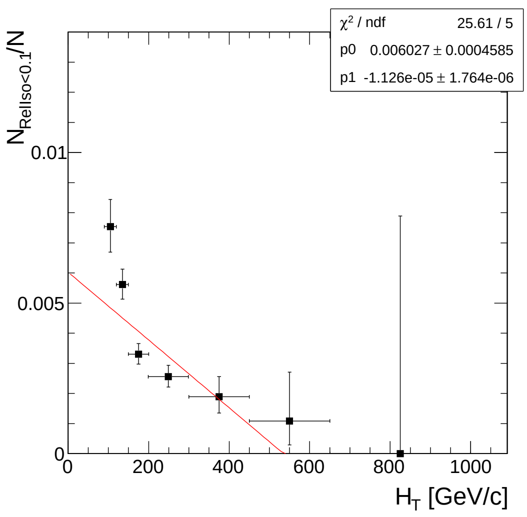
Pursuing the study of all the possible pair combinations of variables, the fraction of multi-jet events with the muon emitted at the primary vertex, $d_0/\sigma < 3$, as been calculated as a function of $p_T^{3rd\ jet}$ as well as the fraction of multi-jet events with $p_T^{3rd\ jet} < 30$ GeV$/c$ as a function of d_0/σ (Fig. 6.4). Both fractions of multi-jet events seem to be rather constant within their statistical uncertainties over the considered range. However, it seems also that the fraction of multi-jet events with $p_T^{3rd\ jet} < 30$ GeV$/c$ is systematically lower for $d_0/\sigma > 3$ than for $d_0/\sigma < 3$. This effect is more visible when replacing $p_T^{3rd\ jet}$ by H_T (Fig. 6.5) , both



(a)

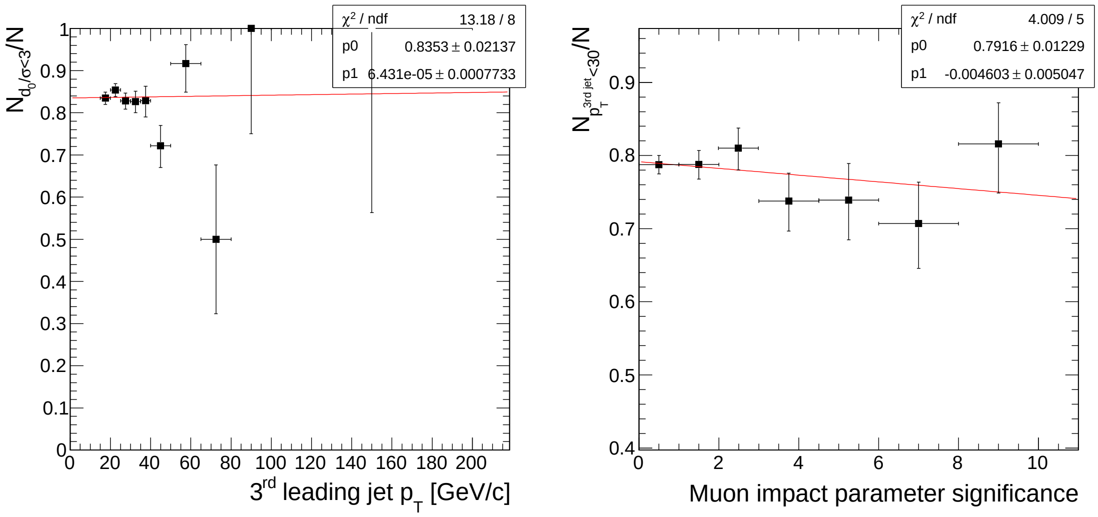


(b)



(c)

Figure 6.3: Fraction of multi-jet events with an isolated muon, $RelIso < 0.1$ as a function of the leading muon transverse momentum (a), as a function of the third leading jet transverse momentum (b) and as a function of the H_T variable (c). Multi-jet events have been selected if they contain at least two (b) or three (a,c) reconstructed jets with a transverse momentum greater than 30 GeV$/c$ in the detector acceptance. Uncertainties have been calculated for an integrated luminosity of 40 pb^{-1}.

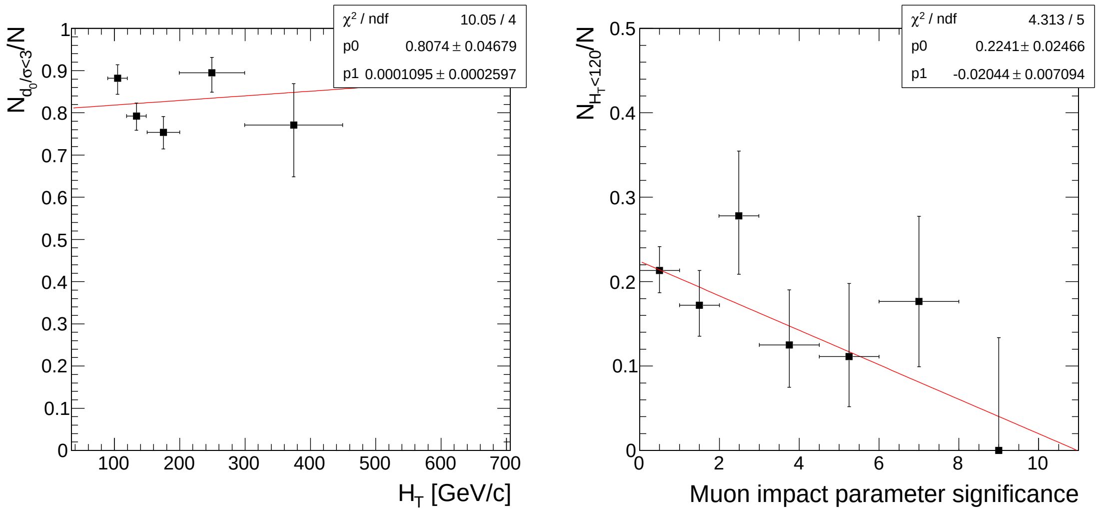


(a) **(b)**

Figure 6.4: Fraction of multi-jet events with a muon impact parameter significance greater than 3 as a function of the third leading jet transverse momentum (a) and fraction of multi-jet events with a third leading jet transverse momentum lower than 30 GeV$/c$ as a function of the muon impact parameter significance (b). Multi-jet events have been selected if they contain at least two reconstructed jets with a transverse momentum greater than 30 GeV$/c$ in the detector acceptance. Uncertainties have been calculated for an integrated luminosity of 40 pb^{-1}.

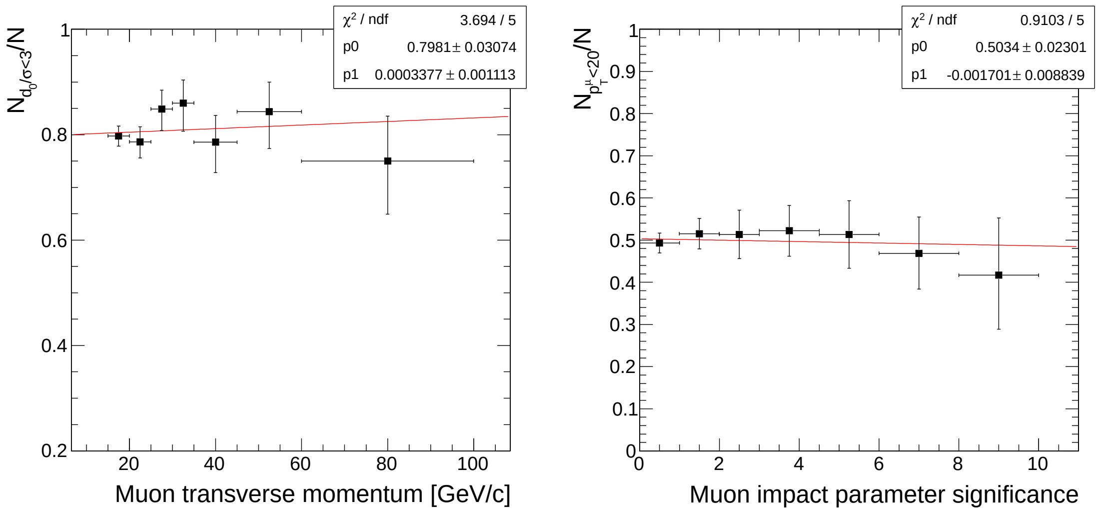
variables being highly correlated. Therefore, none of these pairs of variables has been considered as a candidate for being used with the ABCD method.

As shown in Figure. 6.6, the fraction of multi-jet events with $p_T^\mu < 20$ GeV$/c$ has been calculated as a function of d_0/σ as well as the fraction of events with $d_0/\sigma < 3$ as a function of p_T^μ. Both distributions are compatible with a horizontal line (see fit parameter p_1, giving the slope of the line), indicating that this pair of variables may be a good candidate for the ABCD method. However, as the cut on the muon relative isolation has to be applied on the events before the ABCD method is performed with these two variables, only a small fraction of multi-jet events are left. It is thus impossible to define regions where the multi-jet background is dominant with respect to the other processes. Furthermore, the only sample of Monte-Carlo simulated multi-jet events available in this thesis does not contain a leading muon with a transverse momentum lower than 15 GeV$/c$ and therefore does not allow us to study the correlation between p_T^μ and d_0/σ in the region $p_T^\mu < 15$ GeV$/c$. Although considered as promising, this pair of variables has not been considered any further in this study.



(a) **(b)**

Figure 6.5: Fraction of multi-jet events with a muon impact parameter significance greater than 3 as a function of H_T (a) and fraction of multi-jet events with $H_T < 120$ GeV$/c$ as a function of the muon impact parameter significance (b). Multi-jet events have been selected if they contain at least three reconstructed jets with a transverse momentum greater than 30 GeV$/c$ in the detector acceptance. Uncertainties have been calculated for an integrated luminosity of 40 pb^{-1}.



(a) **(b)**

Figure 6.6: Fraction of multi-jet events with a muon impact parameter significance greater than 3 as a function of the muon transverse momentum (a) and fraction of multi-jet events with a muon transverse momentum greater than 20 GeV$/c$ as a function of the muon impact parameter significance (b). Multi-jet events have been selected if they contain at least three reconstructed jets with a transverse momentum greater than 30 GeV$/c$ in the detector acceptance. Uncertainties have been calculated for an integrated luminosity of 40 pb^{-1}.

Finally, the fraction of multi-jet events with $d_0/\sigma < X_0$ has been calculated as a function of $RelIso$ for three different cuts, $X_0 = 2$, 3 and 4 as well as the fraction of multi-jet events with $RelIso < Y_0$ as a function of d_0/σ for three different cuts, $Y_0 = 0.05$, 0.1 and 0.15 (Fig. 6.7). Within the ranges studied and for the different cut values considered, the fractions of multi-jet events have been found to be constant within their statistical uncertainties, indicating that the $RelIso$ and d_0/σ variables are to a large extent statistically independent and are therefore suitable candidate variables for the ABCD method.

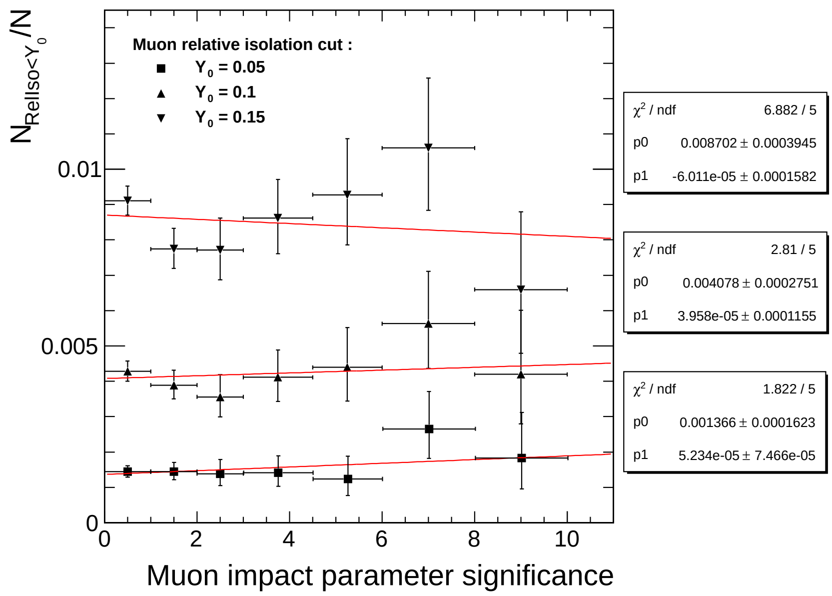
6.2 Performances with multi-jet background only

In this section, the performances of the ABCD method have been studied, using the muon relative isolation and impact parameter significance variables, as these variables appear to be independent.

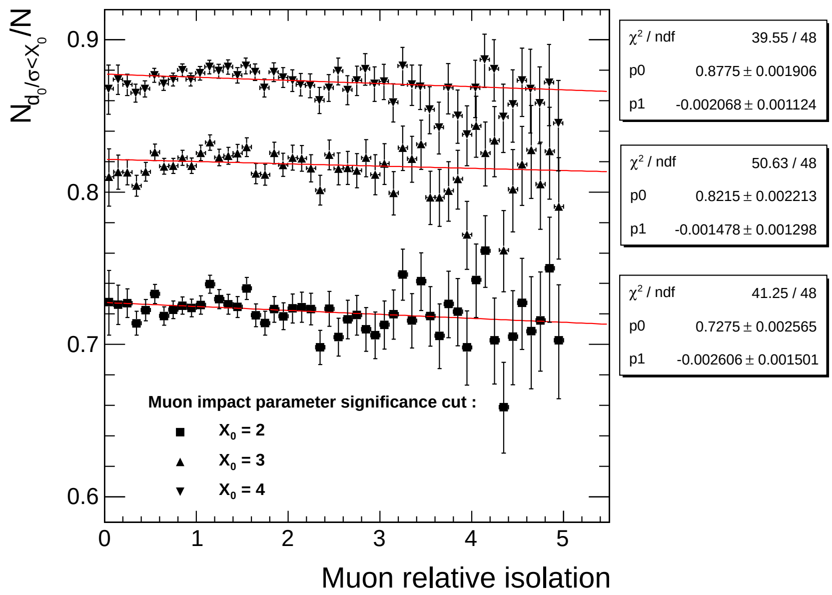
6.2.1 Results

Multi-jet background events, simulated at a centre-of-mass energy of 7 TeV, have been selected with at least three reconstructed jets within the detector acceptance with a transverse momentum greater than 30 GeV/c. Table 6.1 shows the number of selected multi-jet events in the control and signal regions, defined by applying a cut at 0.1 on $RelIso$ and a cut at 3 on d_0/σ, as well as the estimation in the signal region ; the ABCD method estimates (303 ± 37) multi-jet background events in the signal region for a dataset representing 40 pb^{-1} of integrated luminosity. This result is in agreement with the expected number from Monte-Carlo simulations : (303 ± 17) events. Both the expectation and its error have been rescaled to 40 pb^{-1}.

The reconstructed jet multiplicity depends on the jet algorithm as well as on the parameters used to reconstruct the jets in each event. Therefore, in order to check the robustness of the ABCD method, the number of multi-jet events in the signal region has been estimated as a function of the jet multiplicity and compared to the generated number of multi-jet events. However, due to the limited size of the current sample of simulated multi-jet events used in this thesis, it has not been possible to extent this study to a jet multiplicity higher than 4 jets. Table 6.2 shows the number of selected multi-jet events in the control and signal regions, as well as the estimation, for events with exactly three reconstructed jets



(a)



(b)

Figure 6.7: Fraction of multi-jet events with a muon relative isolation lower than 0.05, 0.1 and 0.15 as a function of the impact parameter significance (a) and fraction of multi-jet events with an impact parameter significance greater than 2, 3 and 4 as a function of the relative isolation (b). Multi-jet events have been selected if they contain at least three reconstructed jets with a transverse momentum greater than 30 GeV$/c$ in the detector acceptance. Uncertainties have been calculated for an integrated luminosity of 40 pb^{-1}.

$N_{jets} \geq 3$		$0 < RelIso < 0.1$	$0.1 < RelIso$
	$0 < d_0/\sigma < 3$	303 ± 17 (303 ± 37)	79230 ± 282
	$3 < d_0/\sigma$	68 ± 8	17805 ± 133

Table 6.1: Numbers of multi-jet events simulated in the A, B, C and D regions, with at least three jets reconstructed within the detector acceptance with a transverse momentum greater than 30 GeV$/c$. The regions are defined by cuts at 0.1 on the muon relative isolation ($RelIso$) and at 3 on the impact parameter significance (d_0/σ). The estimation made by the ABCD method in the signal region is shown in parentheses. The numbers of events and their statistical errors have been calculated for an integrated luminosity of 40 pb^{-1}.

$N_{jets} = 3$		$0 < RelIso < 0.1$	$0.1 < RelIso$
	$0 < d_0/\sigma < 3$	259 ± 16 (271 ± 35)	63102 ± 251
	$3 < d_0/\sigma$	60 ± 8	13977 ± 118

Table 6.2: Numbers of multi-jet events simulated in the A, B, C and D regions, with **exactly** three jets reconstructed within the detector acceptance with a transverse momentum greater than 30 GeV$/c$. The regions are defined by cuts at 0.1 on the muon relative isolation ($RelIso$) and at 3 on the impact parameter significance (d_0/σ). The estimation made by the ABCD method in the signal region is shown in parentheses. The numbers of events and their statistical errors have been calculated for an integrated luminosity of 40 pb^{-1}.

within the detector acceptance with a transverse momentum greater than 30 GeV$/c$; the method estimates (271 ± 35) multi-jet background events in the signal region for a dataset representing 40 pb^{-1} of integrated luminosity. This result is in agreement with the expected number from Monte-Carlo simulations : (259 ± 16) events. Table 6.3 shows the same numbers but for events with at least four jets. In this case too, the estimated number of multi-jet events, (34 ± 12), is in agreement with the expected number, (45 ± 7) events, within the statistical error of the estimation. Nevertheless, after having required at least four reconstructed jets, only few events are left in the control region with $RelIso < 0.1$ and $d_0/\sigma > 3$, leading to a relative error on the estimation of $\sim 35\,\%$.

In order to check the stability of these results with respect to the cut on $RelIso$ and d_0/σ, these cuts have been varied independently and for each pair of cuts, the ABCD

$N_{jets} \geq 4$		$0 < RelIso < 0.1$	$0.1 < RelIso$
	$0 < d_0/\sigma < 3$	45 ± 7 (34 ± 12)	16128 ± 127
	$3 < d_0/\sigma$	8 ± 3	3827 ± 62

Table 6.3: Numbers of multi-jet events simulated in the A, B, C and D regions, with at least four jets reconstructed within the detector acceptance with a transverse momentum greater than 30 GeV$/c$. The regions are defined by cuts at 0.1 on the muon relative isolation ($RelIso$) and at 3 on the impact parameter significance (d_0/σ). The estimation made by the ABCD method in the signal region is shown in parentheses. The numbers of events and their statistical errors have been calculated for an integrated luminosity of 40 pb^{-1}.

$N_{jets} \geq 3$	Cut values		Number of multi-jet background events	
	$RelIso$	d_0/σ	expected	estimated
	0.05	2	91 ± 10	90 ± 15
	0.05	3	101 ± 10	107 ± 22
	0.05	4	107 ± 10	129 ± 30
	0.10	2	274 ± 17	253 ± 26
	0.10	3	303 ± 17	303 ± 37
	0.10	4	323 ± 18	340 ± 48
	0.15	2	578 ± 24	540 ± 38
	0.15	3	644 ± 25	636 ± 53
	0.15	4	688 ± 26	680 ± 68

Table 6.4: Numbers of selected multi-jet events as expected by Monte-Carlo simulations and estimated by the ABCD method with the muon relative isolation ($RelIso$) and impact parameter significance (d_0/σ) variables for several cut values. Selected events have at least three jets reconstructed within the detector acceptance with a transverse momentum greater than 30 GeV$/c$. The numbers and their statistical errors have been calculated for an integrated luminosity of 40 pb^{-1}.

method has been applied to estimate the number of multi-jet background events with at least three jets in the signal region. Results are summarized in Table 6.4. In each case, the estimated number of multi-jet events matches the expected number within its statistical uncertainty.

6.2.2 Veto on a second isolated lepton

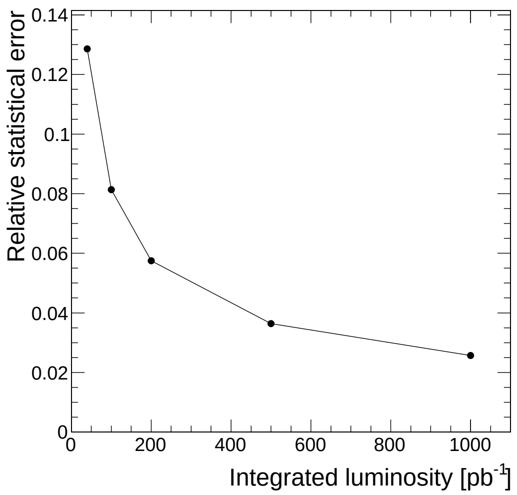
In order not to bias the estimation in the signal region, all the cuts applied on events contained in the signal region must be applied on events contained in the control regions, except for the cuts on the variables used to define these regions. In the case of the veto on a second isolated lepton in the final state (see section 5.4.2), it is necessary to apply this veto also on events for which the muon relative isolation is greater than 0.1 (control regions). But as this cut defines non-isolated muons, it is impossible. Nevertheless, it has been decided to estimate the number of multi-jet events without this veto and apply it on the selected signal events after the ABCD method is performed. It has been checked that the number of multi-jet events with at least three reconstructed jets within the detector acceptance with a transverse momentum greater than 30 GeV$/c$, passing all the selection cuts, and without any veto on a second isolated lepton, is identical to the number of multi-jet events obtained when applying this veto.

6.2.3 Statistical properties of the estimator

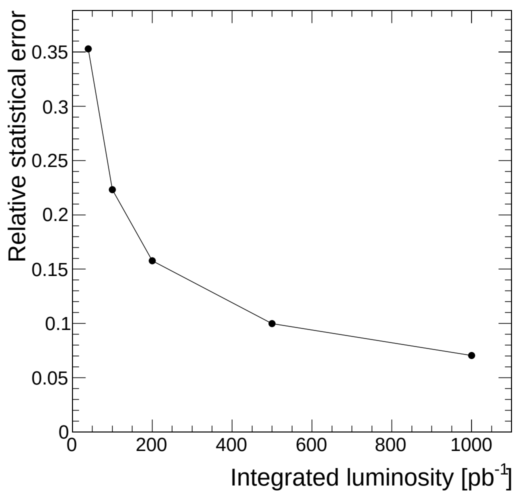
The relative statistical error on the estimation of the number of multi-jet background events decreases as more and more data are collected ; as shown in Figure 6.8, it ranges from $\sim 13\%$ to $\sim 2\%$ when the integrated luminosity increases from 40 pb^{-1} to 1000 pb^{-1} for selected events with exactly three jets reconstructed within the detector acceptance with a transverse momentum greater than 30 GeV$/c$. For selected events with at least four jets, the error increases to 35% and 7% over the same integrated luminosity range.

In order to further study the statistical properties of this estimator, pseudo-experiments have been carried out. Each pseudo-experiment consists in randomizing the numbers of multi-jet events in each of the signal and control regions simultaneously according to Poisson distributions with means equal to the number of events in each region, as expected from Monte-Carlo simulations. For each pseudo-experiment i, the ABCD method has been applied, leading to a set of different estimations of the number of multi-jet background events in the signal region, $N^{Est,i}_{\text{multi-jet}}$, and the pull has been calculated as following :

$$pull = \frac{N^{Est,i}_{\text{multi-jet}} - \bar{N}^{Est,i}_{\text{multi-jet}}}{\Delta N^{Est,i}_{\text{multi-jet}}} \tag{6.2}$$



(a)



(b)

Figure 6.8: Relative statistical error on the estimated number of multi-jet background events as a function of the integrated luminosity. Selected events have exactly three (a) or at least four (b) reconstructed jets within the detector acceptance with a transverse momentum greater than 30 GeV/c.

where $\Delta N^{Est,i}_{\text{multi-jet}}$ is the associated error on $N^{Est,i}_{\text{multi-jet}}$ and $\bar{N}^{Est,i}_{\text{multi-jet}}$ is the arithmetic mean of all the estimations. In the case of an unbiased estimator with properly calculated errors, the pull of the estimation is distributed according to a Gaussian function with a mean equal to zero and a standard deviation equal to one respectively. The pull distribution of the multi-jet background estimation has been calculated using multi-jet events selected with exactly three reconstructed jets within the detector acceptance with a transverse momentum greater than 30 GeV/c. A total of 50000 pseudo-experiments have been performed in each case and for each pseudo-experiment, the numbers of selected multi-jet events have been calculated using Monte-Carlo simulated data representing two different integrated luminosities, 40 pb^{-1} and 1000 pb^{-1}.

For an integrated luminosity of 40 pb^{-1}, the pull distribution (Fig. 6.9a) exhibits an asymmetry, resulting in a mean of (-0.062 ± 0.005) for the Gaussian fit and a poor χ^2/Ndf value. This asymmetry decreases as the integrated luminosity increases to 1000 pb^{-1} (Fig. 6.9b), leading to a mean of (-0.011 ± 0.005) for the Gaussian fit and a χ^2/Ndf value which indicates the pull distribution is compatible with a Gaussian distribution. In this case, the standard deviation of the Gaussian fit is compatible with one : (1.002 ± 0.003). This shows that the error on the estimation of the number of multi-jet background events with at least three jets has been calculated correctly. Attempts have been made to calculate the pull distribution of the estimated numbers of multi-jet events with at least four jets.

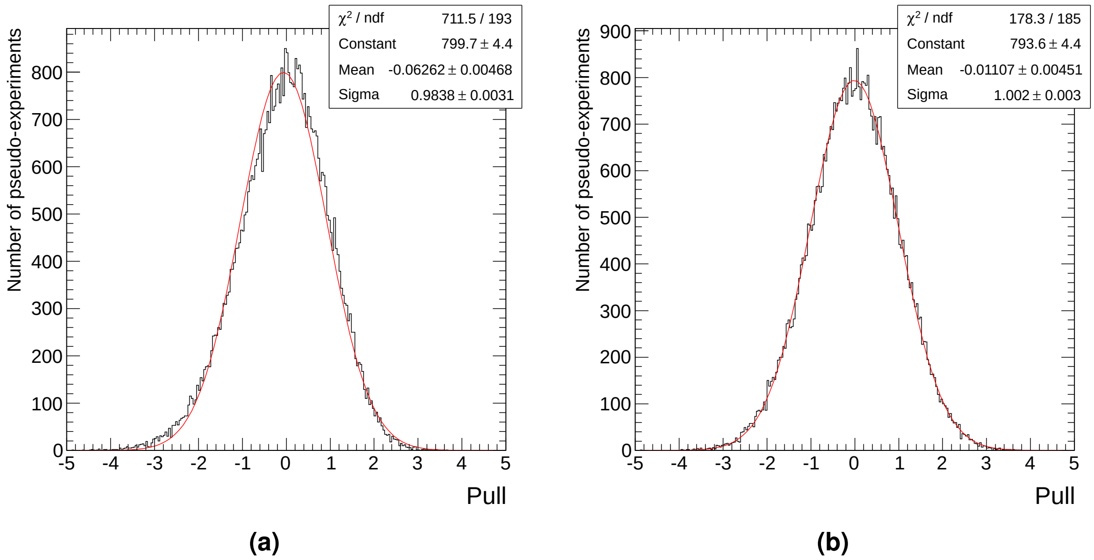


Figure 6.9: Pull distribution of the estimation of the number of multi-jet background events in the signal region, using 50000 pseudo-experiments, for a dataset representing 40 pb^{-1} (a) and 1000 pb^{-1} (b) of integrated luminosity. Selected events have exactly three jets reconstructed within the detector acceptance with a transverse momentum greater than 30 GeV$/c$.

But, due to the small numbers of events selected with 40 pb^{-1} of integrated luminosity in the control region, $RelIso < 0.1$ and $d_0/\sigma > 3$, the dispersion of the random variable representing the number of selected events in this region, which is an integer, is such that it leads to gaps between the estimations produced in different pseudo-experiments. The obtained pull distribution exhibits a spiked structure which does not allow any comparison with a Gaussian distribution. Nevertheless, it has been checked that, at higher integrated luminosity, the pull distribution recovers its Gaussian shape.

6.3 Performances with signal and other background processes

With real data, the regions A, B, C and D will not only contain multi-jet background events but also events from other possible processes not yet rejected or estimated, mainly $V + jets$ and of course the $t\bar{t} + jets$ signal. Their presence might spoil the performances of the ABCD method as it requires to define three regions dominated by the background to estimate. Therefore, to allow a correct estimation of the number of multi-jet background events in the signal region, it is needed to reject regions in which both multi-jet events and

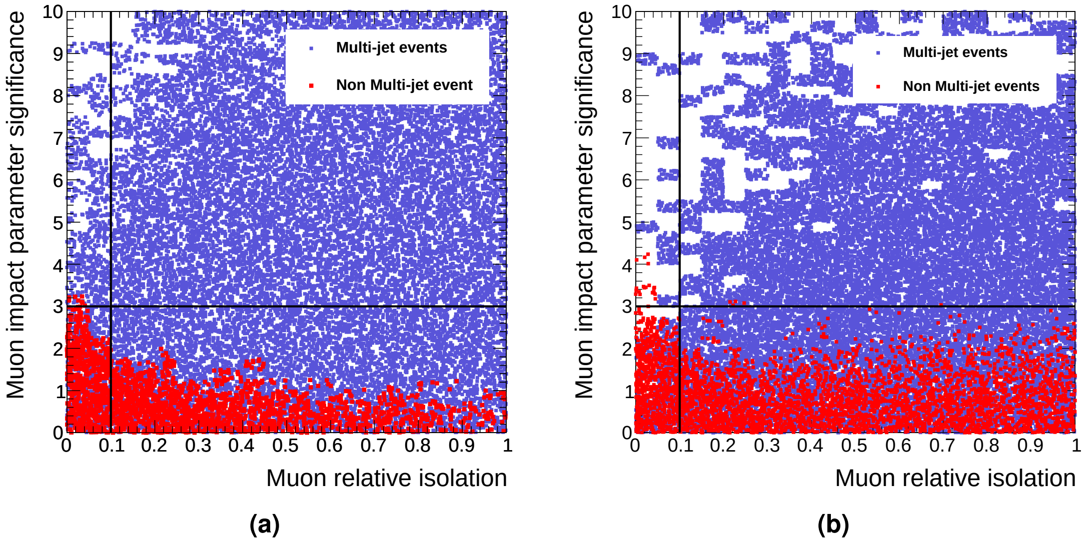


Figure 6.10: Muon impact parameter significance as a function of the muon relative isolation for selected events with three jets (a) and at least four jets (b) reconstructed within the detector acceptance with a transverse momentum greater than 30 GeV$/c$.

events from other processes are present, by defining non contiguous A, B, C and D regions where needed.

In this section, all the processes defined in the previous chapter, including $t\bar{t} + jets$, single top quark events, $V + jets$ and di-boson events have been simulated, in addition to the multi-jet events.

6.3.1 Reducing the signal contamination

Table 6.5 shows the numbers of selected events with at least three jets, from all processes, as well as the numbers of multi-jet events, in the control and signal regions, defined by applying a cut at 0.1 on $RelIso$ and a cut at 3 on d_0/σ ; the fractions of non-multi-jet events in the control regions defined by $RelIso > 0.1$ are both below the percent level and therefore would not affect the estimation of the number of multi-jet events in the signal region. However, this fraction is ~ 6 % for the control region defined by $RelIso < 0.1$ and $d_0/\sigma > 3$. It is therefore necessary to modify the cut on d_0/σ defining this control region control.

In order to find the appropriate cut on d_0/σ, the ratio of the number of selected events from all processes with $RelIso < 0.1$, (S+B), to the number of selected multi-jet events

$N_{jets} \geq 3$		$0 < RelIso < 0.1$	$0.1 < RelIso$
	$0 < d_0/\sigma < 3$	1999 ± 45 Multi-jet : 303 ± 17	79678 ± 282 Multi-jet : 79230 ± 282
	$3 < d_0/\sigma$	72 ± 8 Multi-jet : 68 ± 8	17875 ± 133 Multi-jet : 17805 ± 133

Table 6.5: Numbers of simulated events from all processes, including $t\bar{t} + jets$ and $V + jets$ events, in the A, B, C and D regions, with at least three jets reconstructed within the detector acceptance with a transverse momentum greater than 30 GeV$/c$. The regions are defined by cuts at 0.1 on the muon relative isolation ($RelIso$) and at 3 on the impact parameter significance (d_0/σ). The number of multi-jet events present in these regions are also shown. The numbers of events and their statistical errors have been calculated for an integrated luminosity of 40 pb^{-1}.

with $RelIso < 0.1$, (B) has been calculated as a function of the cut on d_0/σ (Fig. 6.11). As already mentioned, due to the limited size of the sample of simulated multi-jet events, it has not been possible to calculate this ratio for events with more than four jets. However, it is expected, and already noticeable by comparing Figure 6.11a and Figure 6.11b, that the fraction of non-multi-jet events in the control region increases with the jet multiplicity, as non-multi-jet events have on average a higher jet multiplicity than the multi-jet events (cf. section 5.4.2). As CMS has recorded more than 1 fb^{-1} of integrated luminosity for the year 2011, it should be envisaged to use the ABCD method to estimate the number of multi-jet events with more than four jets and thus, defining non contiguous regions becomes important at high jet multiplicity where Monte-Carlo generators, known not to provide an accurate description of the multi-jet events, need to be superseded by data-driven estimations.

As observed in Figure 6.11, a cut at 4 on d_0/σ would already bring the ratio (S+B)/B closer to one for events with three jets. For events with at least four jets, a higher cut should be envisaged. However, the number of selected multi-jet events left in the control region has also to be taken into account in order not to increase too much the statistical uncertainty on the estimation. Therefore, in both cases, cuts higher than 4 have not been considered.

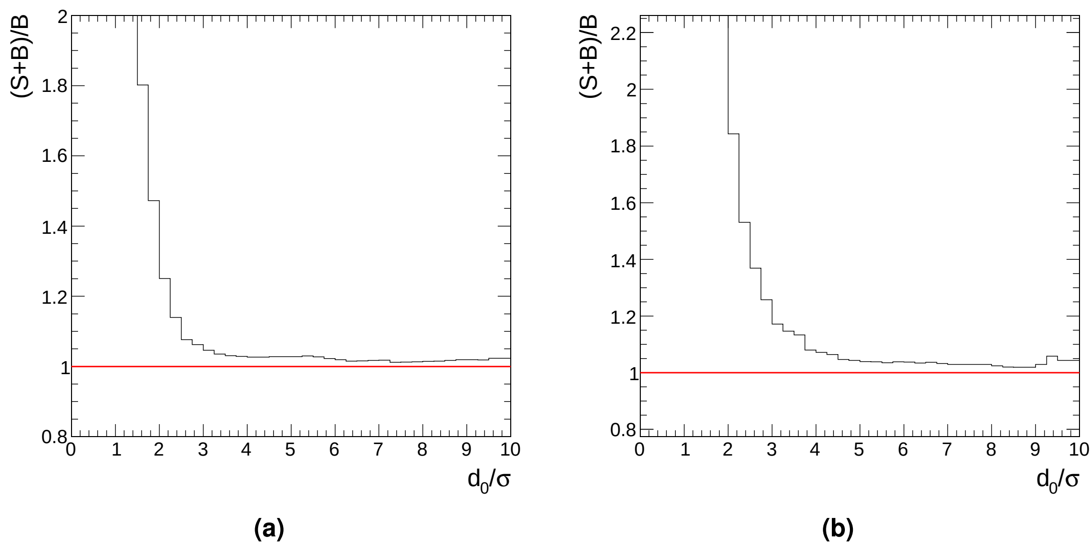


Figure 6.11: Ratio of the number of selected events from all processes with $RelIso < 0.1$, (S+B), to the number of selected multi-jet events with $RelIso < 0.1$, (B), as a function of the cut on d_0/σ. Selected events have exactly three (a) or at least four (b) jets reconstructed within the detector acceptance with a transverse momentum greater than 30 GeV$/c$.

6.3.2 Results

In this section, the performances of the ABCD method using non contiguous regions have been studied. Table 6.6 shows the number of selected multi-jet events with three jets in the control region defined by applying a cut at 0.1 on $RelIso$ and at 4 on d_0/σ. It also shows the number of selected events and the estimation of the number of multi-jet events in the signal region defined by cuts at 0.1 on $RelIso$ and at 3 on d_0/σ. Events contained in the region defined by $3 \leq d_0/\sigma \leq 4$ have not been used in the estimation in order to prevent the contamination of non-multi-jet events in the control regions. The ABCD method estimates 283 ± 43 multi-jet events in the signal region. This results is in agreement within its statistical uncertainty with the expected number of multi-jet events, 259 ± 16.

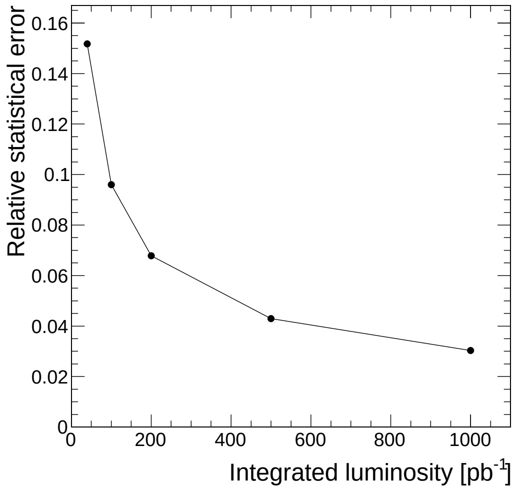
The same quantities as in Table 6.6 are reported in Table 6.7 for events with at least four jets. In this case, the method estimates (48 ± 17) multi-jet background events, in agreement with the expected number, (45 ± 7) events, within the statistical error of the estimation. It appears as possible to apply the ABCD method with non contiguous regions. However, by discarding a fraction of the events contained in the control regions, the statistical uncertainty on the estimation increases.

$N_{jets} = 3$		$0 < RelIso < 0.1$	$0.1 < RelIso$
	$0 < d_0/\sigma < 3$	1398 ± 37 (283 ± 43)	63296 ± 252
	$4 < d_0/\sigma$	44 ± 7	9768 ± 99

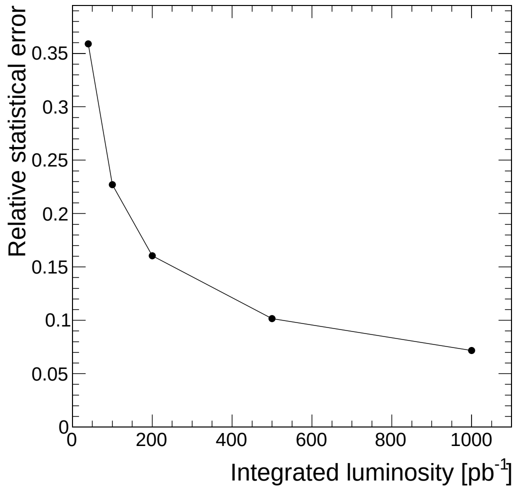
Table 6.6: Numbers of simulated events from all processes, including $t\bar{t} + jets$ and $V + jets$ events, in the A, B, C and D regions, with exactly three jets reconstructed within the detector acceptance with a transverse momentum greater than 30 GeV$/c$. The regions are defined by cuts at 0.1 on the muon relative isolation ($RelIso$) and at 3, for the signal region, and at 4, for the control regions, on the impact parameter significance (d_0/σ). The estimation made by the ABCD method in the signal region is shown in parentheses. The numbers of events and their statistical errors have been calculated for an integrated luminosity of 40 pb^{-1}

$N_{jets} \geq 4$		$0 < RelIso < 0.1$	$0.1 < RelIso$
	$0 < d_0/\sigma < 3$	602 ± 25 (47 ± 17)	16382 ± 128
	$4 < d_0/\sigma$	8 ± 3	2716 ± 52

Table 6.7: Numbers of simulated events from all processes, including $t\bar{t} + jets$ and $V + jets$ events, in the A, B, C and D regions, with at least four jets reconstructed within the detector acceptance with a transverse momentum greater than 30 GeV$/c$. The regions are defined by cuts at 0.1 on the muon relative isolation ($RelIso$) and at 3, for the signal region, and at 4, for the control regions, on the impact parameter significance (d_0/σ). The estimation made by the ABCD method in the signal region is shown in parentheses. The numbers of events and their statistical errors have been calculated for an integrated luminosity of 40 pb^{-1}



(a)

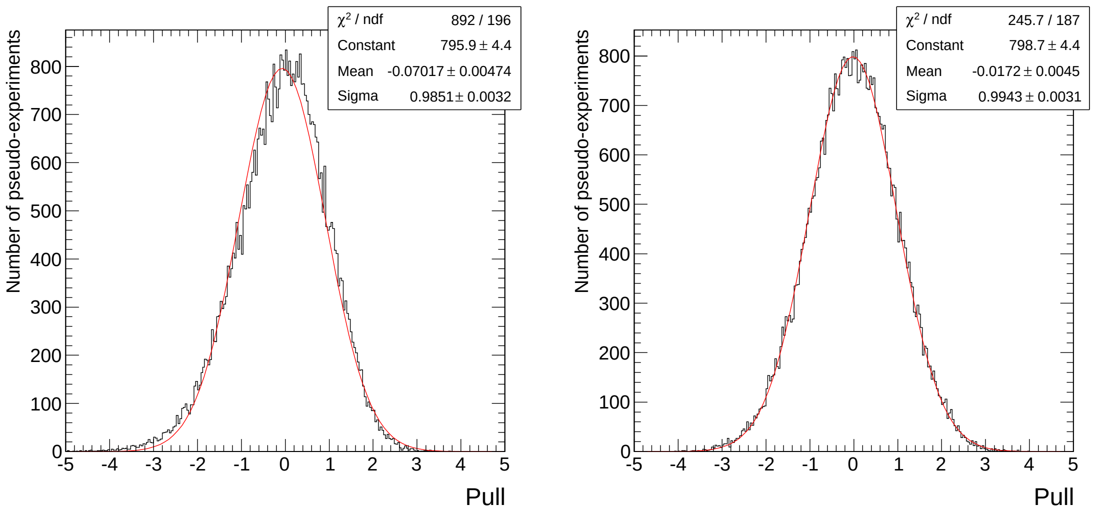


(b)

Figure 6.12: Relative statistical error on the estimated number of multi-jet background events as a function of the integrated luminosity. Selected events have exactly three (a) or at least four (b) jets reconstructed within the detector acceptance with a transverse momentum greater than 30 GeV$/c$.

6.3.3 Statistical properties of the estimator

The relative statistical error on the estimation of the number of multi-jet background events has been calculated as a function of the integrated luminosity. As shown in Figure 6.12, it ranges from 15% to 3% when the integrated luminosity increases from 40 pb^{-1} to 1000 pb^{-1} for selected events with three jets reconstructed within the detector acceptance with a transverse momentum greater than 30 GeV$/c$. For selected events with at least four jets, the error increases to 36% and 7% over the same integrated luminosity range. Pull distributions have been calculated using the same procedure as in section 6.2.3 and show similar behaviours (Fig. 6.12). The distribution obtained for an integrated luminosity of 40 pb^{-1} is asymmetric which leads to a mean of -0.070 ± 0.005 and a poor χ^2/Ndf value. This asymmetry decreases as the integrated luminosity increases ; with 1000 pb^{-1} of integrated luminosity, the pull distribution is compatible with a Gaussian function whose mean is equal to (-0.017 ± 0.005).



(a) (b)

Figure 6.13: Pull distribution of the estimated numbers of multi-jet background events in the signal region, using 50000 pseudo-experiments, for a dataset including events from $t\bar{t}+jets$ and $V+jets$ processes, representing 40 pb^{-1} (a) and 1000 pb^{-1} (b) of integrated luminosity. Selected events have exactly three jets reconstructed within the detector acceptance with a transverse momentum greater than 30 GeV$/c$.

6.4 Systematic uncertainties

To study the systematic uncertainties on the estimated number of selected multi-jet events, the relevant parameters have been varied within conservative limits. Three different sources of systematic uncertainties have been identified ; the first one is related to the uncertainty on the jet energy calibration. In order to assess this uncertainty, the four-momenta of jets have been rescaled up and down with a constant factor $1+\alpha$, prior to the application of the event selection criteria used in this thesis. The results of the ABCD method with the nominal jet energy calibration have been compared with the results obtained with $\alpha = \pm 10$ %. Table 6.8 shows the expected numbers of selected multi-jet events from simulations, $N_{multi\text{-}jets}$, as well as the estimation of the ABCD method performed in presence of the $t\bar{t}+jets$ and all the other processes, $\hat{N}_{multi\text{-}jets}$. As expected, upward and downward variations of the jet energy scale, $\alpha = \pm 10$ %, lead also to a modified number of selected multi-jet events, modifying the relative bias between the expected and the estimated number of multi-jet events by $\pm 5-6$ % for events with three jets and by $0-4$ % for events with at least four jets. However, it is worth noticing that the statistical uncertainties on the estimated numbers of multi-jet events being significantly larger than the observed biases with a modified energy scale, it is impossible to draw any conclusion. Instead, these results are presented both to

Jet multiplicity		$N_{multi\text{-}jets}$	$\hat{N}_{multi\text{-}jets}$	Rel. bias $(\%)$
	$\alpha = -10\ \%$	166	190 ± 30	$+14\ \%$
$N_{jets} = 3$	Nominal	259	283 ± 37	$+9\ \%$
	$\alpha = +10\ \%$	376	388 ± 44	$+3\ \%$
	$\alpha = -10\ \%$	26	28 ± 11	$+8\ \%$
$N_{jets} \geq 4$	Nominal	45	47 ± 15	$+4\ \%$
	$\alpha = +10\ \%$	74	77 ± 19	$+4\ \%$

Table 6.8: Numbers of selected multi-jet events as expected by Monte-Carlo simulations and estimated by the ABCD method with the muon relative isolation ($RelIso$) and impact parameter significance (d_0/σ) variables as a function of the number of reconstructed jets within the detector acceptance with a transverse momentum greater than 30 GeV$/c$ and as a function of the jet energy rescaling factor, α. The numbers have been calculated for an integrated luminosity of 40 pb^{-1} but the statistical uncertainties on the estimation originate from the limited amount of simulated multi-jet events.

give an idea of the magnitude of the systematic uncertainty associated to the jet energy scale and to explicit the methodology.

The second source of systematic uncertainties arise from the uncertainties on the variables used in the ABCD method. In the case of perfectly independent variables, the estimation of the number of selected multi-jet events is invariant with respect to the cuts applied to define the control regions. However, this is no longer the case if the variables are slightly correlated. Furthermore, the relative isolation and the impact parameter significance variables have been found to be independent to a large extent for the multi-jet but not for the other processes. Variations of the cuts applied to define the control regions would then result in variations of the contamination of the control regions by the non-multi-jet processes and thus bias the estimation.

In order to define the range of the cut variations, it is necessary to assess the magnitude of the possible uncertainties on RelIso and d_0/σ. The uncertainty on the muon transverse momentum, p_T^μ is negligible compared to those associated to the calorimeter isolation, CaloIso, and to the tracker isolation, TrackIso. The uncertainty on the relative isolation, defined as $(CaloIso + TrackIso)/p_T^\mu$, is thus proportional to the uncertainties on CaloIso and TrackIso. For a muon transverse momentum ranging from 20 GeV$/c$ to 100 GeV$/c$, a cut at 0.1 on RelIso yields an upper limit on CaloIso+TrackIso of $2 - 10$ GeV. For this energy range, the energy resolution of the calorimeter as well as the track momentum resolution

Jet multiplicity		$N_{multi\text{-}jets}$	$\hat{N}_{multi\text{-}jets}$	$\Delta N^{Syst.}_{multi\text{-}jets}$
	$d_0/\sigma = 3.5$	213	238 ± 31	$+12$ %
$N_{jets} = 3$	Nominal	259	283 ± 37	$+9$ %
	$d_0/\sigma = 4.5$	213	258 ± 38	$+21$ %
	$d_0/\sigma = 3.5$	37	33 ± 11	-11 %
$N_{jets} \geq 4$	Nominal	45	47 ± 15	$+4$ %
	$d_0/\sigma = 4.5$	37	38 ± 14	$+3$ %

Table 6.9: Numbers of selected multi-jet events as expected by Monte-Carlo simulations and estimated by the ABCD method with the muon relative isolation ($RelIso$) and impact parameter significance (d_0/σ) variables as a function of the number of reconstructed jets within the detector acceptance with a transverse momentum greater than 30 GeV$/c$ and as a function of the cut applied on d_0/σ. The numbers have been calculated for an integrated luminosity of 40 pb^{-1} but the statistical uncertainties on the estimation originate from the limited amount of simulated multi-jet events.

of the tracker are of the order of the percent and therefore negligible. But, as discussed in section 4.3.4, pile-up effects are expected to yield an energy offset of ~ 0.2 GeV to the jet transverse energy reconstructed in a cone of opening angle of $\Delta R = 0.5$. Extrapolating these results to CaloIso and TrackIso, defined as the transverse energy measured in a cone of 0.3 in the calorimeter and in the tracker, an offset of ~ 0.3 GeV for CaloIso+TrackIso has been considered. Concerning the transverse impact parameter, as shown in [195] and in more recent CMS studies, the resolution on this parameter is below 20 µm and 8 µm for muons with a transverse momentum higher than 20 GeV$/c$ and 100 GeV$/c$ respectively. However, it has been impossible to estimate the uncertainty of the transverse impact parameter significance, d_0/σ, as the uncertainty of the impact parameter uncertainty, σ, is unknown. Therefore, a simple variation of ± 0.5 on the d_0/σ cut used to define the control regions has been envisaged to assess the related systematic uncertainty.

As for the assessment of the systematic uncertainties related to the jet energy scale, the results of the ABCD method performed with the additional energy offset for $RelIso$ and the modified cuts on d_0/σ have been compared with the nominal results. Table 6.9 shows the expected numbers of selected multi-jets from simulations as well as the estimation of the ABCD method performed in presence of the $t\bar{t} + jets$ and all the other processes, with modified cuts on $RelIso$ and d_0/σ. As expected, the number of multi-jet events in the signal region, defined by $RelIso < 0.1$, is lower than the nominal number due to the offset on $RelIso$ and upward and downward variations on the d_0/σ have yielded additional biases on

Jet multiplicity		$N_{multi\text{-}jets}$	$\hat{N}_{multi\text{-}jets}$	$\Delta N^{Syst.}_{multi\text{-}jets}$
	$-10\ \%$	259	274 ± 37	$+6\ \%$
$N_{jets} = 3$	Nominal	259	283 ± 37	$+9\ \%$
	$+10\ \%$	259	283 ± 36	$+9\ \%$
	$-10\ \%$	45	44 ± 14	$-2\ \%$
$N_{jets} \geq 4$	Nominal	45	47 ± 15	$+4\ \%$
	$+10\ \%$	45	48 ± 15	$+6\ \%$

Table 6.10: Numbers of selected multi-jet events as expected by Monte-Carlo simulations and estimated by the ABCD method with the muon relative isolation ($RelIso$) and impact parameter significance (d_0/σ) variables as a function of the number of reconstructed jets within the detector acceptance with a transverse momentum greater than 30 GeV$/c$ and as a function of the rescaling factor applied on the cross-sections of all the non-multi-jet processes. The numbers have been calculated for an integrated luminosity of 40 pb^{-1} but the statistical uncertainties on the estimation originate from the limited amount of simulated multi-jet events.

the estimations. However, as for the jet energy scale, the large statistical uncertainties on the estimated numbers of multi-jet events due to the limited amount of simulated multi-jet events, prevent to conclude on the size of these systematic uncertainties.

The third source systematic uncertainties is related to the uncertainties on the production cross-sections of the various processes considered in this thesis. These uncertainties affect the expected numbers of non-multi-jet events in the control regions and therefore, lead to additional uncertainties on the estimation of the number of multi-jet events based on the observed numbers of events in these control regions. In order to evaluate these uncertainties, the cross-sections of all the non-multi-jet processes have been varied and the results of the ABCD method compared with the nominal results. Some of the cross-sections used in this thesis have calculated at next-to-next-to leading order in QCD corrections and have therefore small uncertainties, but these uncertainties concern the inclusive processes and are likely to underestimate the cross-section uncertainties when at least three jets are required in the final state. It has thus been decided to vary these cross-section coherently by an arbitrary factor of $\pm 10\ \%$. Table 6.10 shows the expected numbers of selected multi-jets from simulations as well as the estimation of the ABCD method performed in presence of the $t\bar{t} + jets$ and all the other processes, with rescaled cross-sections. As previously, small additional biases on the estimated number of multi-jet events have been observed but remain smaller than the observed statistical uncertainties.

6.5 Conclusions

In this chapter, a method, called the ABCD method, to estimate the number of multi-jet background events passing the top quark pair event selection criteria, as defined in section 5.4.2, has been presented for proton collisions at a centre-of-mass energy of 7 TeV at LHC using the CMS detector. The ABCD method consists in identifying two independent variables which discriminate between the multi-jet events and events from other processes. Thus, pairs of potential variables have been studied and the methodology to control their statistical dependence, possibly with data, has been introduced. A pair of suitable variables has been identified, namely the muon relative isolation and the muon transverse impact parameter significance, and the performances of the method have been evaluated, first with simulated multi-jet events only and then with events from all processes, including $t\bar{t} + jets$ and $V + jets$. It has also been shown that the presence of non-multi-jet events in the control regions, which can potentially bias the estimation, can be reduced by defining non-contiguous signal and control regions. Finally, the statistical uncertainty on the estimated number of multi-jet events has been calculated as a function of the amount of integrated luminosity, showing that, for 1000 pb^{-1}, a relative statistical uncertainty of 3 % is obtained for events with three jets reconstructed within the detector acceptance with a transverse momentum higher than 30 GeV$/c$. For events with at least four jets, the uncertainty increases to 7 %. Despite the limited amount of simulated multi-jet events, an attempt has been made to calculate the systematic uncertainties on the estimation, related to various sources. Emphasize has been put on the methodology rather than on the results as the calculated systematic uncertainties are smaller than the statistical uncertainties on the estimation due to the limited amount of simulated multi-jet events.

7

Estimation of the vector boson background for semi-muonic $t\bar{t}$ studies

As described in chapter 5, the main source of remaining background events after the top quark pair event candidate selection is the production of a weak vector boson associated with jets. It is a source of irreducible background for the top quark pair semi-leptonic decay channel when the vector boson decays leptonically and, therefore, methods developed to estimate instrumental backgrounds like the one presented in the previous chapter cannot be used. Finally, as already emphasized in chapter 5, it is difficult to simulate reliably this process with a high jet multiplicity in the final state. Consequently, an accurate measurement of the number of top quark pair events produced relies strongly on an accurate estimation, from data, of the number of background events due to the production of a weak vector boson associated with jets.

This chapter presents an original method allowing to estimate such a background from data, using the number of b-tagged jets. The method is described in section 7.1. In section 7.2, both the stability and the statistical precision on the estimated parameters are assessed for different amounts of integrated luminosity using Monte-Carlo simulations. A possible extension of the method to improve the accuracy of the estimation is presented in section 7.3. Finally, section 7.4 describes the treatment of the systematic uncertainties associated to the method.

7.1 The estimation method

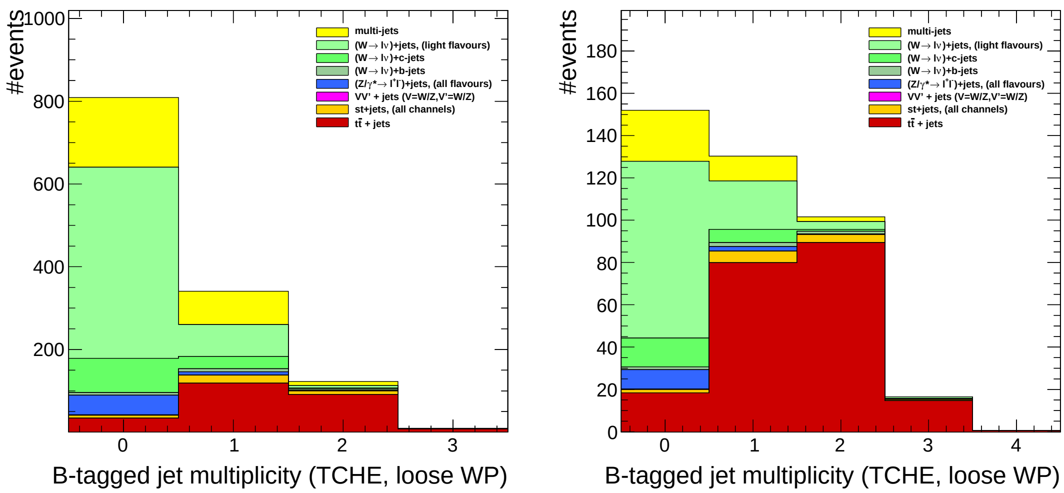
7.1.1 Principle of the method

This method aims to estimate, from data, the number of remaining events with a vector boson associated with jets, called $V+jets$ events, after the criteria used to select top quark pair events have been applied. It exploits the presence of two b-quarks from the top quark pair decays, leading to two b-jets in the final state. In the case of jets associated with the production of a vector boson, they mainly arise from the fragmentation and hadronization of gluons. Nevertheless, a small fraction of the gluons might split into a pair of b/anti-b quarks. As a consequence, within this method, events are rather classified as :

- $t\bar{t}$-$like$ events for events with two b-quarks in the final state,
- V-$like$ events for events without any b-quark in the final state.

As it is possible, to a certain extent, to identify jets from b-quark hadronization (section 4.3.3), the number of b-tagged jets per event has thus been used as the key variable to estimate the number of $V+jets$ events in a given data set (Fig. 7.1. The number of $t\bar{t}$-$like$ and V-$like$ events in that set are estimated from the number of events observed with 0, 1 and 2 b-tagged jets, using an extended maximum likelihood technique [196, 197, 198]. To do so, it is necessary to calculate the probability density function of the random variable whose outcome represents the number of b-tagged jets per event as a function of the number of b-quarks present in the final state. As this density probability function is also a function of the b-tagging and mis-tagging efficiencies of the algorithm used to identify the b-jets, it is possible to either use the efficiency values derived from Monte-Carlo simulations or to estimate them as well. Both cases have been studied in this thesis.

Events which cannot be classified either as V-$like$ or $t\bar{t}$-$like$ have to be subtracted before the method is performed. The number of multi-jet events to be subtracted is evaluated using the ABCD method, as described in Chapter 6. The contribution from single top quark events is contained in both $t\bar{t}$-$like$ and V-$like$ categories. Indeed, single top quark production processes (section 5.2.1) leads to one or two b-quarks in the final state, depending on the production channel and on the flavour of the interacting quark in



(a) **(b)**

Figure 7.1: Numbers of simulated events as a function of the b-tagged jet multiplicity, after event selection with three (a) or four (b) jets reconstructed within the detector acceptance with a transverse momentum greater than 30 GeV$/c$. Jets from quarks are identified using the *TrackCountingHighEfficiency* b-tagging algorithm with a b-discriminant (b-disc.) threshold of 2.03.

the initial state. The possibility to modify the present method in order to include events with only one b-quark as a third category has been investigated but including this third category introduces too many free parameters to lead to a stable and precise estimation. Instead, the single top quark contribution is subtracted using Monte-Carlo simulations, before the method is performed. This decision is also motivated by the fact that the single top quark cross-sections for the different production channels are known at least at the next-to-leading order in QCD corrections. In addition, the number of remaining events from this process after event selection represents less than 5% and 7% of the number of background events with at least three and four reconstructed jets within the detector acceptance with a transverse momentum greater than 30 GeV$/c$ respectively.

The vector boson production associated with a single b-quark in the final-state, $W + b$ and $Z + b$, is heavily suppressed because of smallness of the V_{ub} and V_{cb} CKM matrix elements[1] and by the smallness of the proton b-quark distribution function respectively. Their contribution has thus been neglected.

[1] $|V_{ub}| = (3.89 \pm 0.44) \times 10^{-3}$ and $|V_{cb}| = (40.6 \pm 1.3) \times 10^{-3}$ [108]

7.1.2 Probability density function of the b-tagged jet multiplicity

Assuming than each event belongs to one of the aforementioned categories ($t\bar{t}$-$like$ or V-$like$) which are mutually exclusive, it is possible, thanks to the total probability theorem, to write the probability for an event to have x jets tagged as b-jets, $p(x\ b\text{-}tagged\ jets)$[2], as a sum of two conditional probabilities, multiplied respectively by their prior probability :

$$\begin{aligned} p(x\ b\text{-}tagged\ jets) = {} & p(x\ b\text{-}tagged\ jets|t\bar{t}\text{-}like) \times p_{t\bar{t}\text{-}like} \\ & + p(x\ b\text{-}tagged\ jets|V\text{-}like) \times p_{V\text{-}like} \end{aligned} \tag{7.1}$$

with :

- $p(x\ b\text{-}tagged\ jets|t\bar{t}\text{-}like)$, probability for an event to have x b-tagged jets, given that this event belongs to the $t\bar{t}$-$like$ category,
- $p_{t\bar{t}\text{-}like}$, prior probability for an event to belong to the $t\bar{t}$-$like$ category, i.e. to have two b-quarks in the final state,

and

- $p(x\ b\text{-}tagged\ jets|V\text{-}like)$, probability for an event to have x b-tagged jets, given that this event belongs to the V-$like$ category,
- $p_{V\text{-}like}$, prior probability for an event to belong to the V-$like$ category, i.e. to have zero b-quark in the final state.

Simple estimators of these prior probabilities are given by the fraction of $t\bar{t}$-$like$ and V-$like$ events in the set of selected events, N_{total} :

$$\hat{p}_{t\bar{t}\text{-}like} = \frac{N_{t\bar{t}\text{-}like}}{N_{total}} \tag{7.2}$$

$$\hat{p}_{V\text{-}like} = \frac{N_{V\text{-}like}}{N_{total}} \tag{7.3}$$

The probability for an event to have x b-tagged jets is now parametrized as a function of the number of $t\bar{t}$-$like$ and V-$like$ events. Moreover, each conditional probability is also expressed as a function of the following parameters :

- ϵ_{btag}, the probability for a jet to be tagged as a b-jet, given that it is a jet coming from the hadronization of a b-quark (*b-tagging efficiency*).

[2] In this method, $p(3\ b\text{-}tagged\ jets)$ represents the probability for an event to have 3 or more b-jets in the final state.

- ϵ_{mistag}, the probability for a jet to be tagged as a b-jet, given that it is a jet coming from the hadronization of a light quark or a gluon (*mis-tagging efficiency*),

Finally, for a given event, the probability to mis-tag non b-jets is evaluated for each jet. The probability for an event to have x b-tagged jets due to mis-tagging is then a function of the jet multiplicity, n.

It is now possible to associate a probability density function, *f*, to the discrete random variable X, which represents the number of b-tagged jets per event, for events with n reconstructed jets in the final state :

$$f(x;\theta,n) = \frac{1}{C}\sum_{i=0}^{3} p(x\ b\text{-}tagged\ jets)\ \delta(x - x_i) \tag{7.4}$$

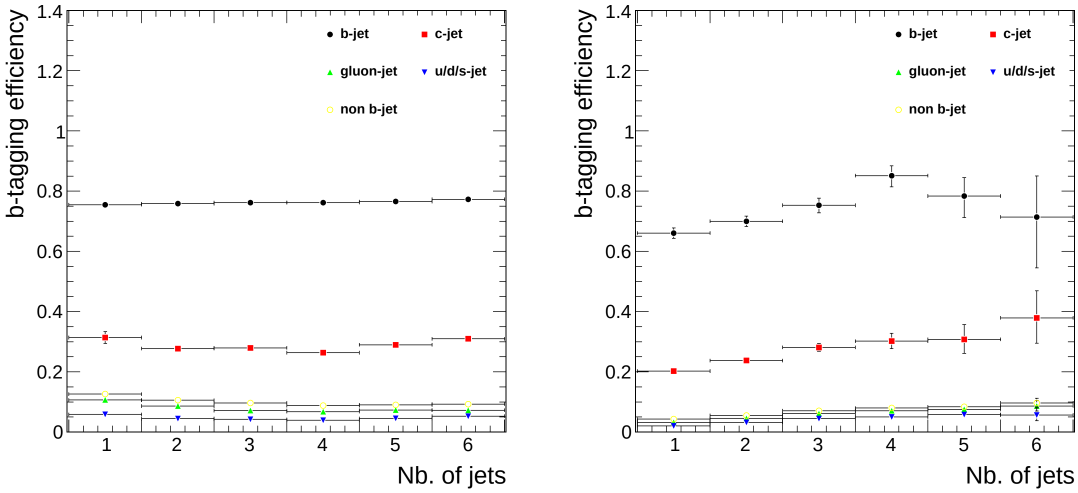
where $\theta = \{\epsilon_{btag}, \epsilon_{mistag}, p_{V\text{-}like}, p_{t\bar{t}\text{-}like}\}$, C is a constant which takes care that the integral of this function is normalized to unity, δ, the Dirac delta function[3] and $x_0, \ldots x_3$, the discrete values accessible to the random variable X.

Using the total number of selected events, N_{total}, it is now possible to estimate, from the data, the value of the parameters ϵ_{btag}, ϵ_{mistag}, $N_{t\bar{t}\text{-}like}$ and $N_{V\text{-}like}$ for different jet multiplicities.

The performances of the b-tagging algorithms, measured by their b-tagging and mis-tagging efficiencies for a given b-discriminant threshold, are dependent on the reconstructed jet transverse momentum and pseudo-rapidity. But, in this method, the number of expected b-tagged jets is inferred from the number of reconstructed jets within the detector acceptance with a transverse momentum greater than 30 GeV$/c$, without regard to their individual transverse momentum or pseudo-rapidity value. A parametrization of the b-tagging and mis-tagging efficiencies as a function of the jet transverse momentum and pseudo-rapidity would lead to the introduction of new degrees of freedom and therefore would degrade the accuracy of the estimations. This implies that ϵ_{btag} and ϵ_{mistag} are estimators of the inclusive b-tagging and mis-tagging efficiency respectively.

The mis-tagging efficiency differs whether the jet comes from the hadronization of a gluon or a quark. It also depends on the quarks flavour (Fig. 7.2), referred to as jet parton flavour hereafter. Gluon jets and u/d/s-jets have a low mis-tagging efficiency ; it ranges

[3]The Dirac delta function δ is defined by : $\begin{array}{ll} \delta(x) = 0 & \text{if } x \neq 0 \\ \int_{-\epsilon}^{+\epsilon} \delta(x) = 1 & \forall \epsilon > 0 \end{array}$



(a) **(b)**

Figure 7.2: B-tagging or mis-tagging efficiency for different jet parton flavours, using the *TrackCountingHighEff* b-tagging algorithm and a b-discriminant threshold of 2.03 (loose working point). The efficiencies are shown as a function of the number of reconstructed jets within the detector acceptance with transverse momentum greater than 30 GeV$/c$ in events where a top quark pair (a) or a vector boson associated with jets (b) has been produced. Errors are due to the limited amount of simulated events.

from 4% to 10% and decreases with the jet multiplicity for $t\bar{t} + jets$ events. It ranges from 2% to 8% for $V + jets$ events and increases with the jet multiplicity. Jets from c quark have a higher mis-tagging efficiency ; it is comprised between 28% and 32% for $t\bar{t} + jets$ events. For $V + jets$ events, it ranges from 20% to 28% and increases with the jet multiplicity. Therefore, the averaged mis-tagging efficiency for a given event depends on the jet parton flavour composition of its final state. In the case of $t\bar{t} + jets$ events, the hadronically decaying W boson leads to c-quarks in the final state while in the case of $V + jets$ events, the jets associated to the vector boson production are mainly gluon jets. In order to account for this difference, it is necessary to introduce two different estimators for the mis-tagging efficiency (ϵ_{mistag} for $t\bar{t} + jets$ events and ϵ'_{mistag} for $V + jets$).

7.1.3 B-tagged jet multiplicity probability

For a selected V-$like$ event, having x b-tagged jets among the n reconstructed jets in the final state means that x jets out of n have been mis-tagged while the remaining $n - x$ jets have not. Therefore, the expression of the probability for a selected V-$like$ event to have x

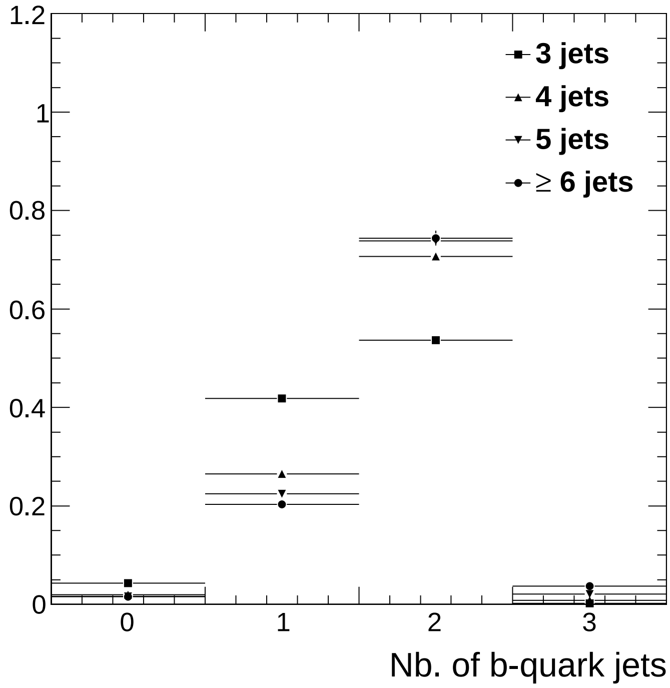


Figure 7.3: Fraction of $t\bar{t} + jets$ events with 0, 1, 2 and 3 b-quark jets, reconstructed within the detector acceptance with a transverse momentum greater than 30 GeV$/c$ for different jet multiplicities. Errors are due to the limited amount of simulated events.

b-tagged jets in the final state is simply :

$$p(x\ b\text{-}tagged\ jets|V\text{-}like) = \binom{x}{n} (\epsilon'_{mistag})^x (1 - \epsilon'_{mistag})^{n-x} \tag{7.5}$$

where the binomial coefficient $\binom{x}{n}$ represents the number of possible subsets of x elements out of a set of n distinct elements. For a selected $t\bar{t}\text{-}like$ event, this expression is more complicated as there are two different potential sources of b-tagged jets : b-jets which are tagged as such and non b-jets which are mis-tagged. Furthermore, after the jet reconstruction, b-jets may not be selected because of the kinematic cuts on the jet transverse momentum and/or on the detector acceptance. The fraction of $t\bar{t} + jets$ events with a given number of selected b-jets in the final state (Fig. 7.3) increases up to 2 b-jets. The fraction of events with more than 2 b-jets, due to gluon splitting of initial or final state radiations, is negligible. Therefore, it is necessary to define three mutually exclusive categories for $t\bar{t}\text{-}like$ events, according to the number of b-jets present in the final state $(0, 1$ and $2)$. For each category with i selected b-jets, the probability for an event to have x b-tagged jets in the final state, $p(x\ b\text{-}tagged\ jets|i\ b\text{-}jets)$, is expressed in terms of b-tagging and mis-tagging efficiencies and the total probability to have x b-tagged jets is then calculated using the total probability theorem :

$$p(x\ b\text{-}tagged\ jets|t\bar{t}\text{-}like) = \sum_{i=0}^{2} p(x\ b\text{-}tagged\ jets|i\ b\text{-}jets) \times p(i\ b\text{-}jets|t\bar{t}\text{-}like) \tag{7.6}$$

As an example, here is given the expression of the different elements needed to calculate $p(1\ b\text{-}tagged\ jet|t\bar{t}\text{-}like)$, for n jets in the final state :

- $p(1\ b\text{-}tagged\ jets|2\ b\text{-}jets)$ is the probability for an event to have 1 b-tagged jet, given that there are 2 selected b-jets in the final state ; either one of the two b-jets is tagged as such, the other one not and there is no mis-tagged light jet or none of the two b-jets is tagged as such and only one light jet is mis-tagged.

$$\begin{aligned} p(1\ b\text{-}tagged\ jets|2\ b\text{-}jets) = {} & 2\ \epsilon_{btag}(1-\epsilon_{btag}) \times (1-\epsilon_{mistag})^{n-2} \\ & + (1-\epsilon_{btag})^2 \times (n-2)\ \epsilon_{mistag}\ (1-\epsilon_{mistag})^{n-3} \end{aligned} \tag{7.7}$$

- $p(1\ b\text{-}tagged\ jets|1\ b\text{-}jet)$ is the probability for a $t\bar{t}\text{-}like$ event to have 1 b-tagged jet, given that there is 1 selected b-jet in the final state : either the b-jet is tagged as such and there is no mis-tagged light jet or the b-jet is not tagged as such and only one light jet is mis-tagged.

$$\begin{aligned} p(1\ b\text{-}tagged\ jets|1\ b\text{-}jet) = {} & \epsilon_{btag} \times (1-\epsilon_{mistag})^{n-1} \\ & + (1-\epsilon_{btag}) \times (n-1)\epsilon_{mistag}(1-\epsilon_{mistag})^{n-2} \end{aligned} \tag{7.8}$$

- $p(1\ b\text{-}tagged\ jets|0\ b\text{-}jet)$ is the probability for a $t\bar{t}\text{-}like$ event to have 1 b-tagged jet, given that there is no selected b-jet in the final state ; only one light jet is mis-tagged.

$$p(1\ b\text{-}tagged\ jets|0\ b\text{-}jet) = n\ \epsilon_{mistag}(1-\epsilon_{mistag})^{n-1} \tag{7.9}$$

The generic expressions of the conditional probabilities $p(X\ b\text{-}tagged\ jets|t\bar{t}\text{-}like)$ and $p(X\ b\text{-}tagged\ jets|V\text{-}like)$ can be found in Appendix A.1.1.

For a given jet multiplicity, the probability for a $t\bar{t}\text{-}like$ event to have x selected b-jets in the final state is estimated from Monte-Carlo simulations, using simple estimators :

$$\hat{p}(x\ b\text{-}jets|t\bar{t}\text{-}like) = \frac{N^{x\ b\text{-}jets}_{t\bar{t}\text{-}like}}{N_{t\bar{t}\text{-}like}} \tag{7.10}$$

where $N^{x\ b\text{-}jets}_{t\bar{t}\text{-}like}$ is the number of $t\bar{t}\text{-}like$ events with x b-jets in the final state and $N_{t\bar{t}\text{-}like}$ is the total number of selected $t\bar{t}\text{-}like$ events.

These parameters are fixed, using values calculated with Monte Carlo simulations. The sensitivity of the method to these values has been studied and the results included in the calculation of the systematic uncertainties.

7.2 Performances

With a too small amount of integrated luminosity, it might not be possible to estimate the numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events, together with the b-tagging and mis-tagging efficiencies as the numbers of events at high b-tagged jet multiplicities might be too low. Therefore, different scenarios have been considered, depending on the number of free parameters in addition to the numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events :

Scenario 0 : the b-tagging and mis-tagging efficiencies have been fixed to their nominal values derived from Monte-Carlo simulations. The remaining free parameters are then the numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events, $N_{V\text{-}like}$ and $N_{t\bar{t}\text{-}like}$.

Scenario I-ϵ_{btag} : the mis-tagging efficiency for $V\text{-}like$ and $t\bar{t}\text{-}like$ events, ϵ'_{mistag} and ϵ_{mistag}, have been fixed to their nominal values derived from Monte-Carlo simulations. The remaining free parameters are then the numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events, $N_{V\text{-}like}$ and $N_{t\bar{t}\text{-}like}$, and the b-tagging efficiency, ϵ_{btag}.

Scenario I-ϵ_{mistag} : the b-tagging efficiency, ϵ_{btag}, and the mis-tagging efficiency for $V\text{-}like$ events, ϵ'_{mistag}, have been fixed to their nominal values derived from Monte-Carlo simulations. The remaining free parameters are then the numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events, $N_{V\text{-}like}$ and $N_{t\bar{t}\text{-}like}$, and the mis-tagging efficiency for $t\bar{t}\text{-}like$ events, ϵ_{mistag}.

Scenario I-ϵ'_{mistag} : the b-tagging efficiency, ϵ_{btag}, and the mis-tagging efficiency for $t\bar{t}\text{-}like$ events, ϵ_{mistag}, have been fixed to their nominal values derived from Monte-Carlo simulations. The remaining free parameters are then the numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events, $N_{V\text{-}like}$ and $N_{t\bar{t}\text{-}like}$, and the mis-tagging efficiency for $V\text{-}like$ events, ϵ'_{mistag}.

Scenario II-ϵ_{btag}-ϵ_{mistag} : the mis-tagging efficiency for $V\text{-}like$ events, ϵ'_{mistag}, has been fixed to its nominal value derived from Monte-Carlo simulations. The remaining free parameters are then the numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events, $N_{V\text{-}like}$ and $N_{t\bar{t}\text{-}like}$, the b-tagging efficiency , ϵ_{btag}, and the mis-tagging efficiency for $t\bar{t}\text{-}like$ events, ϵ_{mistag}.

The performances of the method have been evaluated with Monte-Carlo simulated pseudo-data representing 40, 200 and 1000 pb^{-1} of integrated luminosity respectively. The nominal values for the different parameters, $N_{V\text{-}like}$, $N_{t\bar{t}\text{-}like}$, ϵ_{btag}, ϵ'_{mistag} and ϵ_{mistag} (Tab. 7.1) have been calculated with Monte-Carlo simulated events passing the standard selection criteria defined in section 5.4.2. The processes taken into account for each event category are the following :

$\boldsymbol{N_{t\bar{t}\text{-}like}}$ **events** (2 b-quarks) : $t\bar{t} + jets$ events and $Vb\bar{b} + jets$ events

$\boldsymbol{N_{V\text{-}like}}$ **events** (0 b-quark) : $V +$ light $jets$, $Wc + jets$, $Vc\bar{c} + jets$ and $VV' + jets$

The removal of the multi-jet and single top background events, before the estimation, has been accounted for in the calculation of the systematic uncertainties associated to the estimation.

7.2.1 Statistical properties of the estimators

The stability of the method for various scenarios has been investigated using the distributions of the pulls. The statistical uncertainties have been evaluated from the distributions of the estimated parameters. These distributions have been obtained using pseudo-experiments. Each pseudo-experiment, for a given integrated luminosity, consists in sampling a data set from the probability density function defined in Eq. (7.4) whose parameters have been set to their nominal values, calculated with Monte-Carlo simulations. The value of each free parameter is then estimated using an extended likelihood technique. For a given integrated luminosity, 10000 pseudo-experiments have been performed and for each pseudo-experiment, the number of events in the data set is randomized according to a Poisson distribution whose mean is equal to the expected total number of events.

Study of the pull distribution

The pull of a given parameter α has been calculated for a pseudo-experiment i as following :

$$pull = \frac{\alpha^{Est,i} - \alpha^{nom.}}{\Delta\alpha^{Est,i}} \tag{7.11}$$

N_{jets}	$N_{t\bar{t}\text{-}like}$	$N_{V\text{-}like}$	B-disc. threshold	ϵ_{btag}	ϵ_{mistag}	ϵ'_{mistag}
3	256.7 ± 1.4	706.4 ± 9.0	2.03	0.76 ± 0.02	0.101 ± 0.007	0.062 ± 0.005
			3.20	0.66 ± 0.02	0.053 ± 0.007	0.025 ± 0.003
			4.38	0.57 ± 0.02	0.035 ± 0.006	0.015 ± 0.003
4	196.3 ± 1.1	141.3 ± 4.1	2.03	0.76 ± 0.02	0.095 ± 0.007	0.079 ± 0.012
			3.20	0.66 ± 0.02	0.053 ± 0.007	0.025 ± 0.003
			4.38	0.57 ± 0.02	0.035 ± 0.006	0.015 ± 0.003
5	77.5 ± 0.6	24.1 ± 1.7	2.03	0.77 ± 0.02	0.100 ± 0.008	0.086 ± 0.039
			3.20	0.66 ± 0.02	0.053 ± 0.007	0.025 ± 0.003
			4.38	0.57 ± 0.02	0.035 ± 0.006	0.015 ± 0.003
≥ 6	32.2 ± 0.4	5.9 ± 0.9	2.03	0.77 ± 0.02	0.108 ± 0.004	0.074 ± 0.056
			3.20	0.66 ± 0.02	0.053 ± 0.007	0.025 ± 0.003
			4.38	0.57 ± 0.02	0.035 ± 0.006	0.015 ± 0.003

Table 7.1: Numbers of selected $t\bar{t}\text{-}like$ and $V\text{-}like$ events, b-tagging and mis-tagging efficiencies and their associated uncertainties derived from Monte-Carlo simulations, as a function of the number of reconstructed jets within the detector acceptance with a transverse momentum greater than 30 GeV$/c$. The numbers of selected events have been calculated for an integrated luminosity of 40 pb^{-1}. The b-tagging and mis-tagging efficiencies have been calculated using the *TrackCountingHighEfficiency* b-tagging algorithm with a b-discriminant (b-disc.) threshold of 2.03, 3.2 and 4.38. The uncertainties stem from the limited amount of simulated events available.

where $\alpha^{Est,i}$ is the estimated value of the parameter α for the i^{th} pseudo-experiment, $\Delta\alpha^{Est,i}$ its associated error and $\alpha^{nom.}$ the nominal value of the parameter. In the case of an unbiased estimator with properly calculated errors, the pull of the estimation is distributed according to a Gaussian function with a mean equal to zero and a standard deviation equal to one. In the following, pull distributions are fitted with a Gaussian function and the compatibility with a Gaussian distribution evaluated through the value of the χ^2 per degree of freedom returned by the fit.

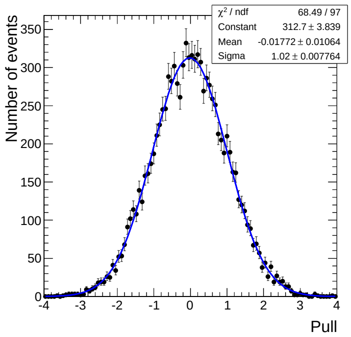
All the pull distributions have been calculated using the *TrackCountingHighEfficiency* b-tagging algorithm with a b-discriminant threshold of 2.03 for different jet multiplicities, from 3 jets to ≥ 6 jets, with pseudo-data corresponding to different amounts of integrated luminosity.

Scenario 0 :

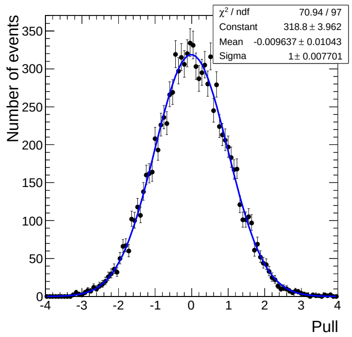
the Gaussian fits to the pull distributions of the two free parameters, $N_{V\text{-}like}$ and $N_{t\bar{t}\text{-}like}$ (Fig 7.4 and Fig. 7.5 respectively), have a χ^2 per degree of freedom smaller than 3 for all the different jet multiplicities and integrated luminosities considered. However, the χ^2 value tends to increase when the jet multiplicity increases, mainly for the distributions using events with at least 6 jets and luminosities of 40 pb^{-1} and 200 pb^{-1}. An asymmetry is also visible in the pull distributions for 5 and ≥ 6 jets at a luminosity of 40 pb^{-1} and 200 pb^{-1}, leading to a mean of the Gaussian fit slightly non compatible with zero, indicating a bias towards negative values of ~ 4 % to ~ 6 %. No significant bias is observed at lower jet multiplicities or at higher integrated luminosities. As the amount of integrated luminosity collected by CMS in 2011 already exceeds 1000 pb^{-1}, this scenario is worth to be tried. The uncertainties on the estimated parameters seem to be calculated correctly as the standard deviations of all the Gaussian fits are compatible with unity.

Scenario I-$\boldsymbol{\epsilon_{btag}}$:

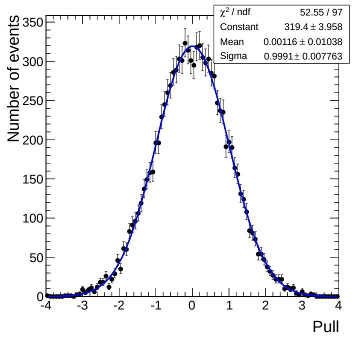
as there are more free parameters in this scenario than in the previous one, the method becomes unstable at high jet multiplicity for an integrated luminosity of 40 pb^{-1}. Therefore, this luminosity will not be considered further in this study. For higher integrated luminosities, 200 pb^{-1} and 1000 pb^{-1}, all the pull distributions of $N_{V\text{-}like}$, $N_{t\bar{t}\text{-}like}$ and ϵ_{btag} (Fig. 7.6) are compatible with Gaussian distributions of standard deviations equal to unity, the χ^2 per degree of freedom returned by the fit being ~ 1. Furthermore, the pull distributions for $N_{t\bar{t}\text{-}like}$ and ϵ_{btag} do not show any bias. However, a $4-7$ % bias towards negative values is observed for the $N_{V\text{-}like}$ pull distributions with more than 3 jets at 200 pb^{-1} of integrated luminosity. This bias becomes compatible



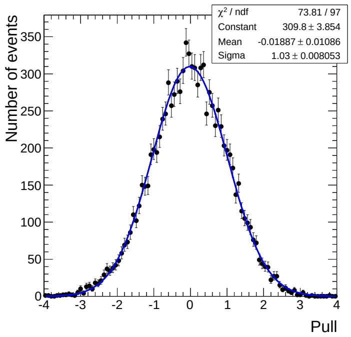
(a) 3 jets, 40 pb^{-1}



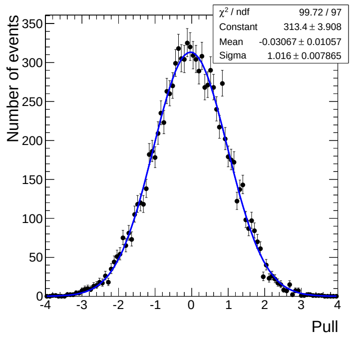
(b) 3 jets, 200 pb^{-1}



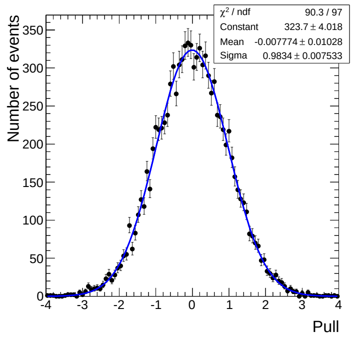
(c) 3 jets, 1000 pb^{-1}



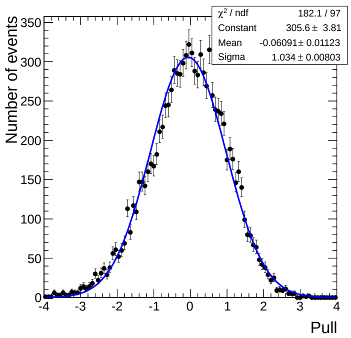
(d) 4 jets, 40 pb^{-1}



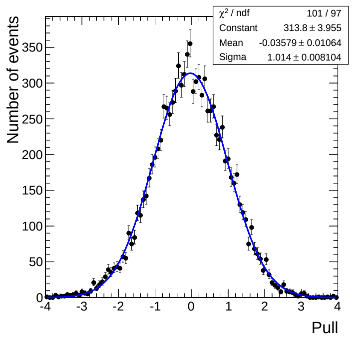
(e) 4 jets, 200 pb^{-1}



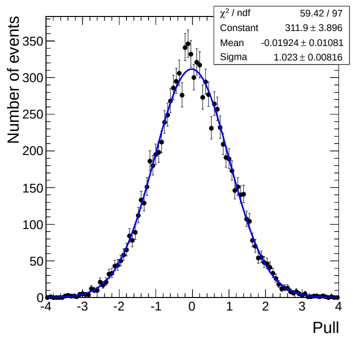
(f) 4 jets, 1000 pb^{-1}



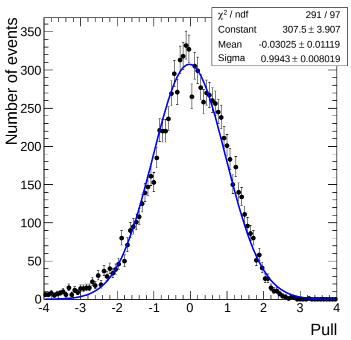
(g) 5 jets, 40 pb^{-1}



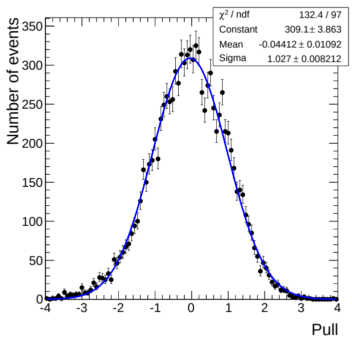
(h) 5 jets, 200 pb^{-1}



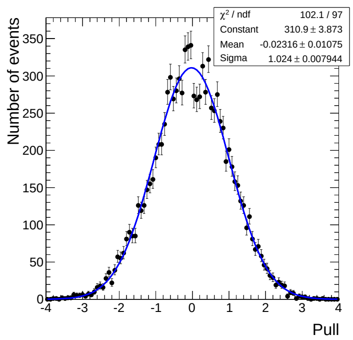
(i) 5 jets, 1000 pb^{-1}



(j) ≥ 6 jets, 40 pb^{-1}

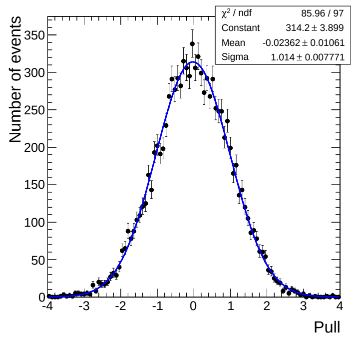


(k) ≥ 6 jets, 200 pb^{-1}

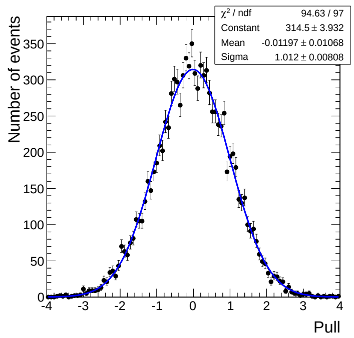


(l) ≥ 6 jets, 1000 pb^{-1}

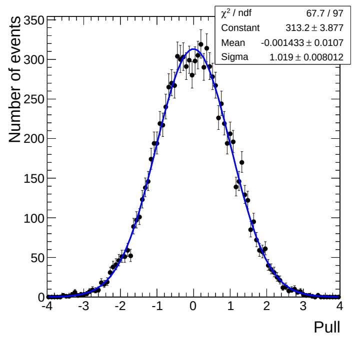
Figure 7.4: Scenario 0 : Pull distributions of the estimated number of $V\text{-}like$ events for different jet multiplicities and integrated luminosities, using 10000 pseudo-experiments and the *TrackCountingHighEfficiency* b-tagging algorithm with a discriminator threshold of 2.03.



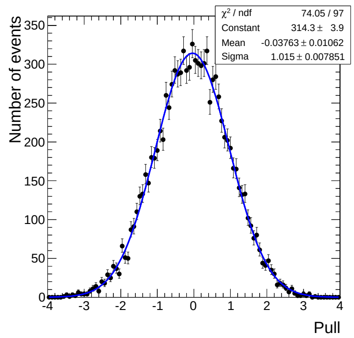
(a) 3 jets, 40 pb^{-1}



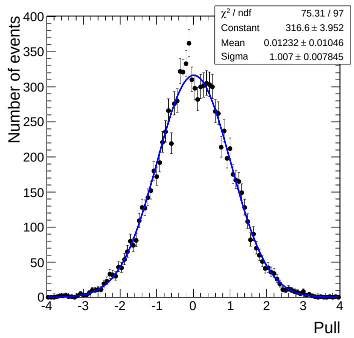
(b) 3 jets, 200 pb^{-1}



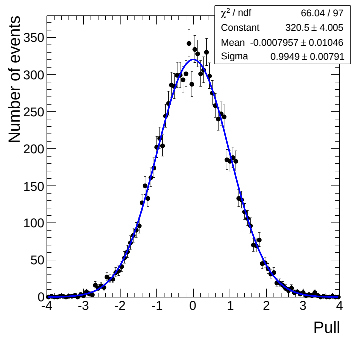
(c) 3 jets, 1000 pb^{-1}



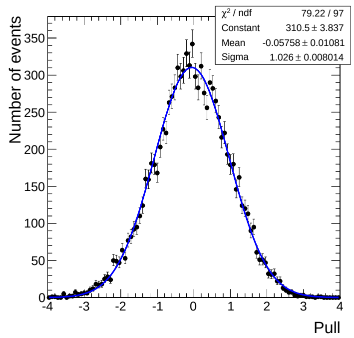
(d) 4 jets, 40 pb^{-1}



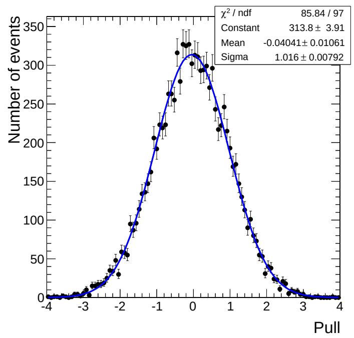
(e) 4 jets, 200 pb^{-1}



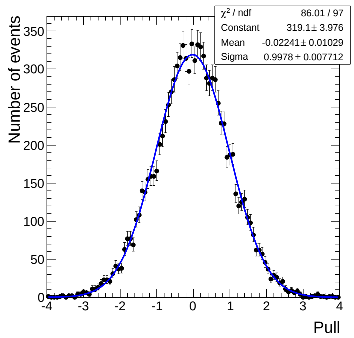
(f) 4 jets, 1000 pb^{-1}



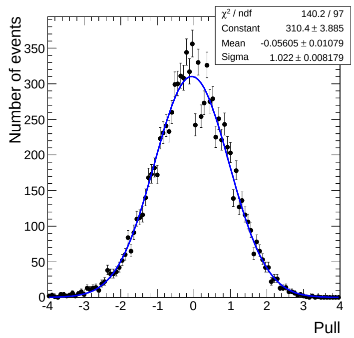
(g) 5 jets, 40 pb^{-1}



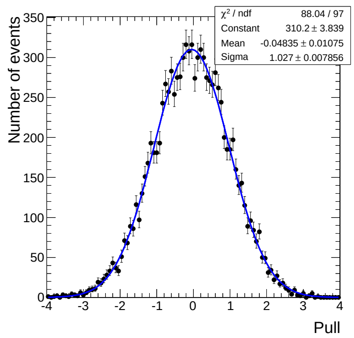
(h) 5 jets, 200 pb^{-1}



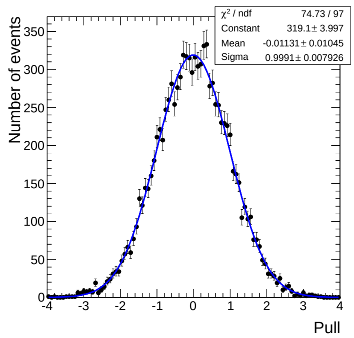
(i) 5 jets, 1000 pb^{-1}



(j) ≥ 6 jets, 40 pb^{-1}



(k) ≥ 6 jets, 200 pb^{-1}



(l) ≥ 6 jets, 1000 pb^{-1}

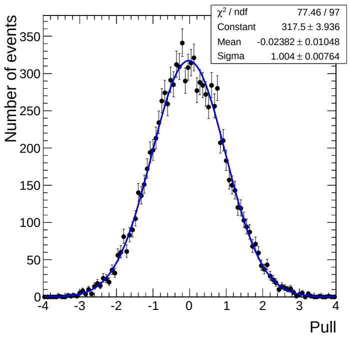
Figure 7.5: Scenario 0 : Pull distributions of the estimated number of $t\bar{t}$-$like$ events for different jet multiplicities and integrated luminosities, using 10000 pseudo-experiments and the *TrackCountingHighEfficiency* b-tagging algorithm with a discriminator threshold of 2.03.

with zero at 1000 pb^{-1}, except for at least 6 jets for which it remains of the order of 3 %. Nevertheless, it should not prevent this scenario to be tried already with the integrated luminosity recorded by CMS in 2011.

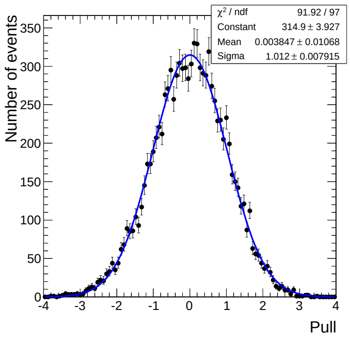
Scenario I-ϵ_{mistag} :
the pull distributions for $N_{V\text{-}like}$, $N_{t\bar{t}\text{-}like}$ and ϵ_{mistag} (Fig. 7.7) are already compatible with a Gaussian distribution at 200 pb^{-1} of integrated luminosity for all the different jet multiplicities, except for the ϵ_{mistag} pull distributions with at least 5 jets due to a small asymmetry towards the negative values. At higher integrated luminosities, the compatibility of these distributions with a Gaussian distribution is restored. The statistical uncertainty on these estimations seems to calculated correctly as the standard deviation of all the pull distributions are compatible with unity. Finally, while no bias is observed in the ϵ_{mistag} pull distributions, the $N_{V\text{-}like}$ and $N_{t\bar{t}\text{-}like}$ pull distributions show a $4-5$ % bias towards negative values at 200 pb^{-1} of integrated luminosity for events with at least 5 jets. This bias vanishes at higher integrated luminosities.

Scenario I-ϵ'_{mistag} :
at 200 pb^{-1} of integrated luminosity, the χ^2 per degree of freedom of the Gaussian fit is ~ 1 for the $N_{V\text{-}like}$, $N_{t\bar{t}\text{-}like}$ and ϵ'_{mistag} pull distributions with less than 5 jets (Fig. 7.8). At higher jet multiplicity, the method fails to provide stable estimations for the free parameters. At 1000 pb^{-1} of integrated luminosity, only the pull distributions with 5 jets are compatible with Gaussian distributions and a small bias of ~ 4 % towards negative values is observed for $N_{t\bar{t}\text{-}like}$ with $5 jets$. However, it has been checked that all the pull distributions with at least 6 jets become also compatible with Gaussian distributions when the integrated luminosity increases to 5000 pb^{-1}. Nevertheless, it is still worth to apply this scenario on the data collected by CMS in 2011 for jet multiplicities up to 5. For events with at least 6 jets, it is possible to set the mis-tagging efficiency either to the value obtained with 5 jets as its dependency on the jet multiplicity is expected to be small (see Fig. 7.2b) or to the value calculated from Monte-Carlo simulations. In both cases, with only two free parameters remaining, it has already been shown that stable estimations of $N_{V\text{-}like}$ and $N_{t\bar{t}\text{-}like}$ are possible at 1000 pb^{-1} of integrated luminosity (cf. scenario 0).

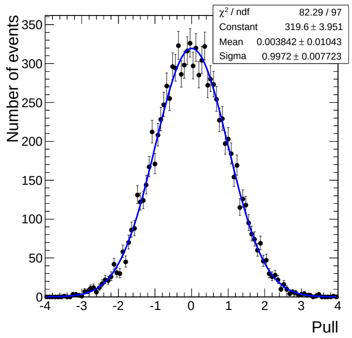
Scenario II-ϵ_{btag}-ϵ_{mistag} :
in this scenario, as the mis-tagging efficiency for $V\text{-}like$ events remains a constant, the method is able to provide stable estimations for $N_{V\text{-}like}$ at any jet multiplicity with 1000 pb^{-1} of integrated luminosity. However, for ϵ_{btag} and ϵ_{mistag}, the pull distributions become incompatible with Gaussian distributions for events with at least 6 jets



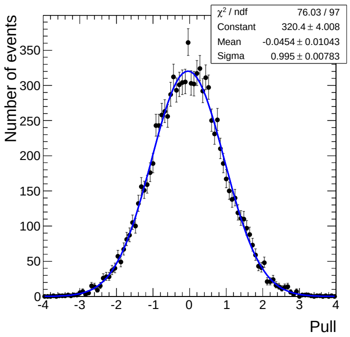
(a) $N_{V\text{-}like}$, 3 jets



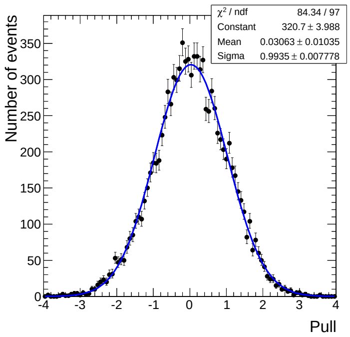
(b) $N_{t\bar{t}\text{-}like}$, 3 jets



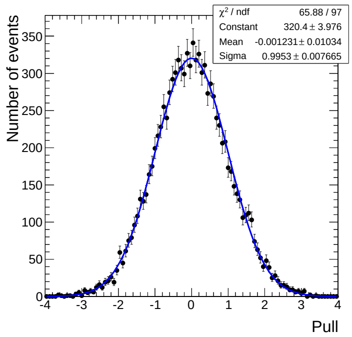
(c) ϵ_{btag}, 3 jets



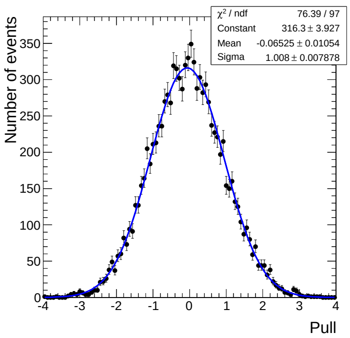
(d) $N_{V\text{-}like}$, 4 jets



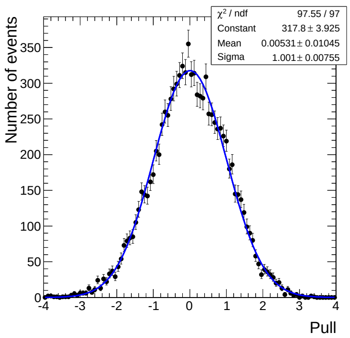
(e) $N_{t\bar{t}\text{-}like}$, 4 jets



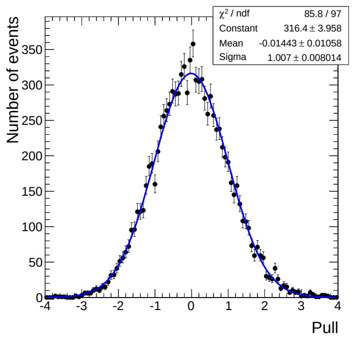
(f) ϵ_{btag}, 4 jets



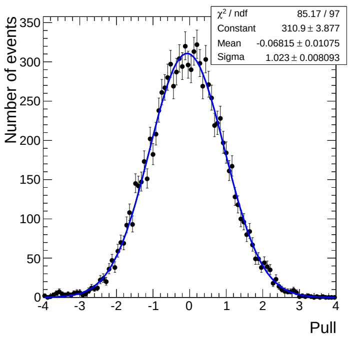
(g) $N_{V\text{-}like}$, 5 jets



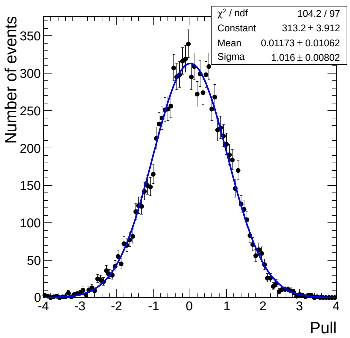
(h) $N_{t\bar{t}\text{-}like}$, 5 jets



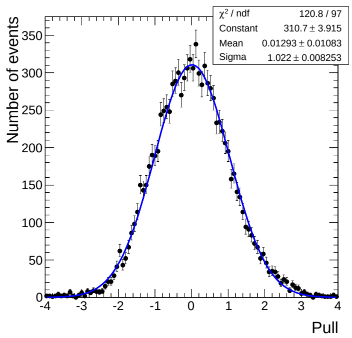
(i) ϵ_{btag}, 5 jets



(j) $N_{V\text{-}like}$, ≥ 6 jets

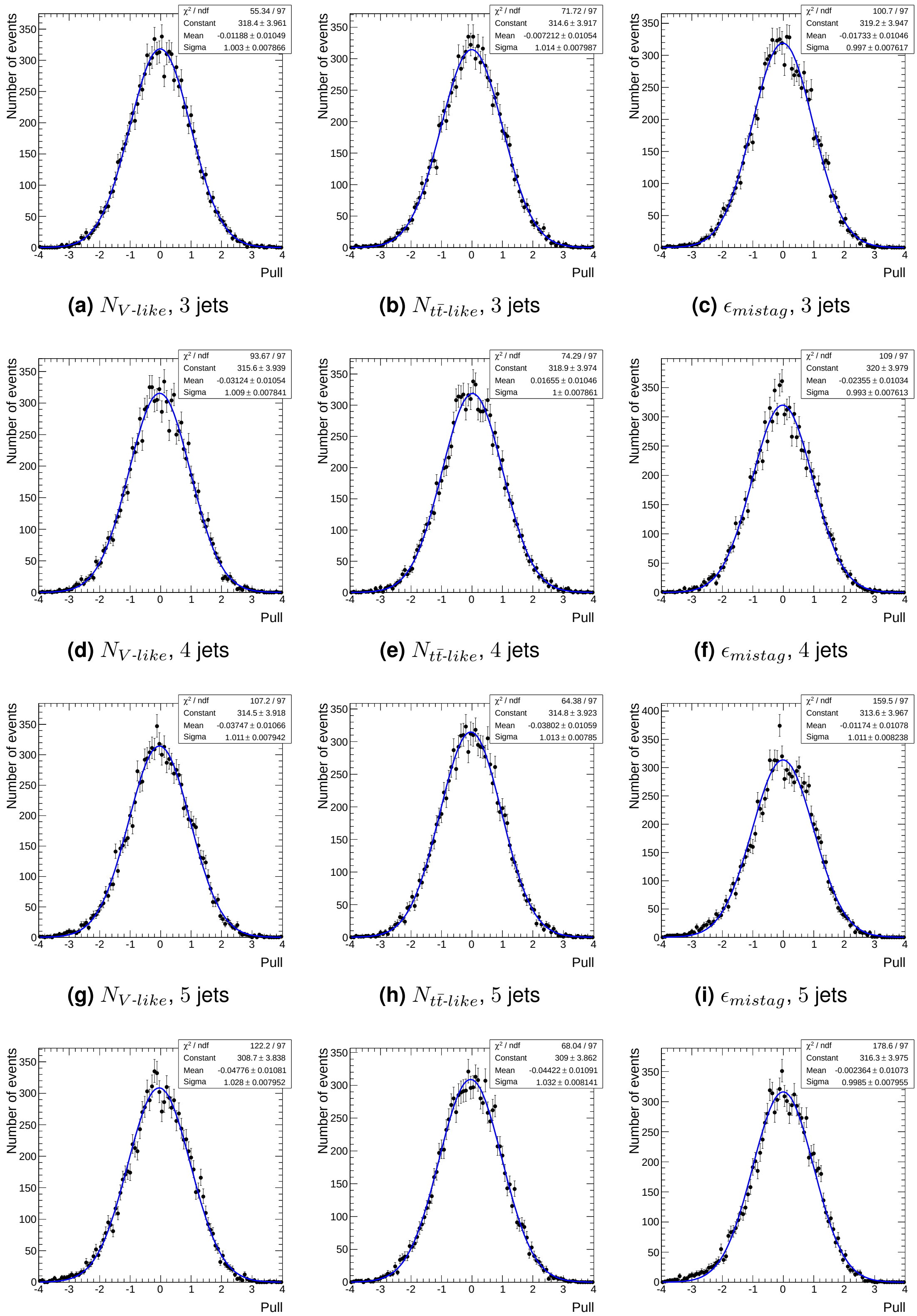


(k) $N_{t\bar{t}\text{-}like}$, ≥ 6 jets



(l) ϵ_{btag}, ≥ 6 jets

Figure 7.6: Scenario I-ϵ_{btag} : Pull distributions of the estimated numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events and of the estimated b-tagging efficiency, at 200 pb^{-1} of integrated luminosity for different jet multiplicities, using 10000 pseudo-experiments and the *TrackCountingHigh-Efficiency* b-tagging algorithm with a discriminator threshold of 2.03.



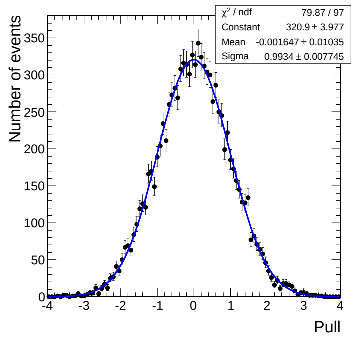
(a) $N_{V\text{-}like}$, 3 jets **(b)** $N_{t\bar{t}\text{-}like}$, 3 jets **(c)** ϵ_{mistag}, 3 jets

(d) $N_{V\text{-}like}$, 4 jets **(e)** $N_{t\bar{t}\text{-}like}$, 4 jets **(f)** ϵ_{mistag}, 4 jets

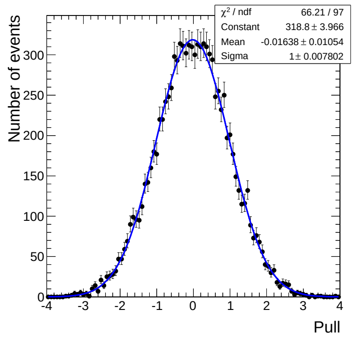
(g) $N_{V\text{-}like}$, 5 jets **(h)** $N_{t\bar{t}\text{-}like}$, 5 jets **(i)** ϵ_{mistag}, 5 jets

(j) $N_{V\text{-}like}$, ≥ 6 jets **(k)** $N_{t\bar{t}\text{-}like}$, ≥ 6 jets **(l)** ϵ_{mistag}, ≥ 6 jets

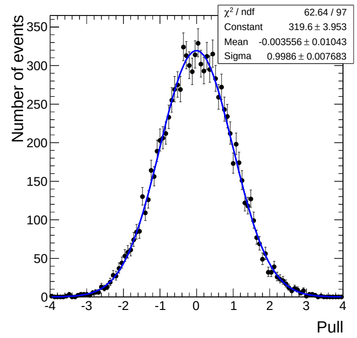
Figure 7.7: Scenario I-ϵ_{mistag} : Pull distributions of the estimated numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events and of the estimated mis-tagging efficiency for $t\bar{t}\text{-}like$ events, at 200 pb^{-1} of integrated luminosity, for different jet multiplicities using 10000 pseudo-experiments and the *TrackCountingHighEfficiency* b-tagging algorithm with a discriminator threshold of 2.03.



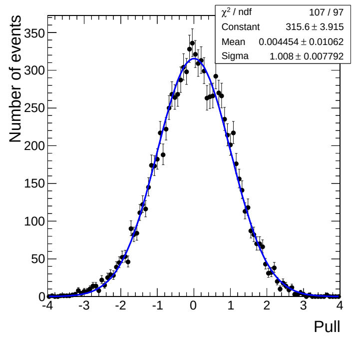
(a) $N_{V\text{-}like}$, 3 jets



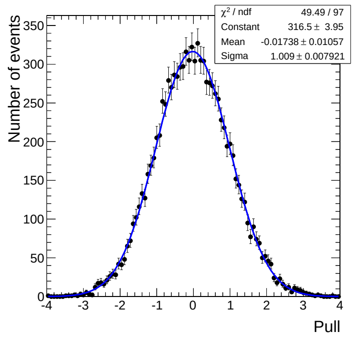
(b) $N_{t\bar{t}\text{-}like}$, 3 jets



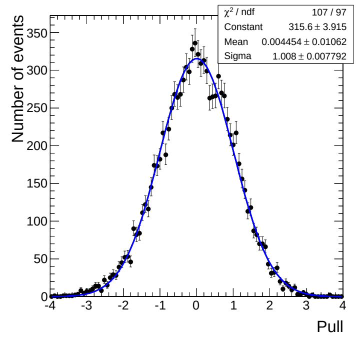
(c) ϵ'_{mistag}, 3 jets



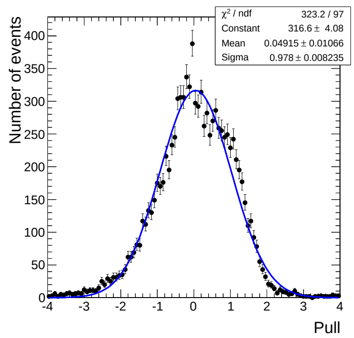
(d) $N_{V\text{-}like}$, 4 jets



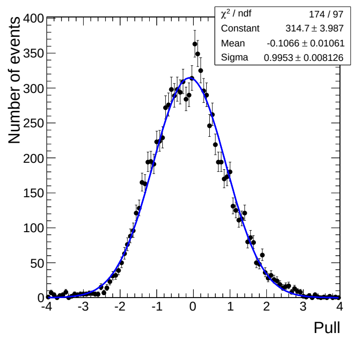
(e) $N_{t\bar{t}\text{-}like}$, 4 jets



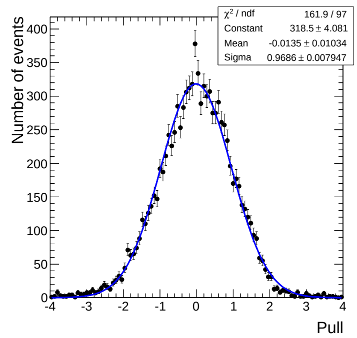
(f) ϵ'_{mistag}, 4 jets



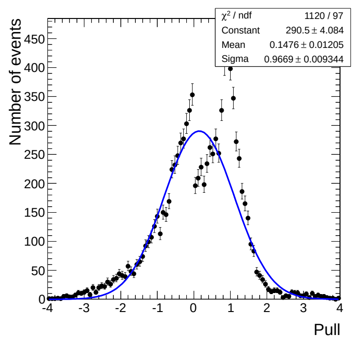
(g) $N_{V\text{-}like}$, 5 jets



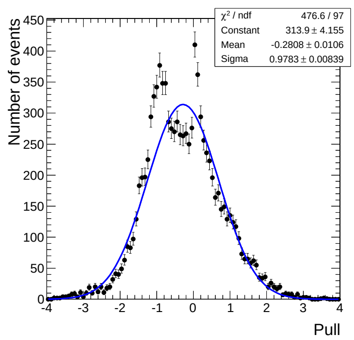
(h) $N_{t\bar{t}\text{-}like}$, 5 jets



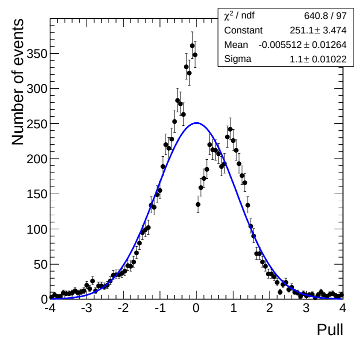
(i) ϵ'_{mistag}, 5 jets



(j) $N_{V\text{-}like}$, ≥ 6 jets



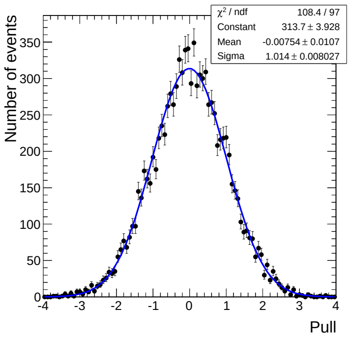
(k) $N_{t\bar{t}\text{-}like}$, ≥ 6 jets



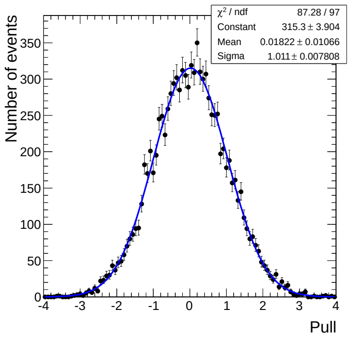
(l) ϵ'_{mistag}, ≥ 6 jets

Figure 7.8: Scenario I-ϵ'_{mistag} : Pull distributions of the estimated numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events and of the estimated mis-tagging efficiency for $V\text{-}like$ events at 200 pb^{-1} of integrated luminosity, for different jet multiplicities using 10000 pseudo-experiments and the *TrackCountingHighEfficiency* b-tagging algorithm with a discriminator threshold of 2.03.

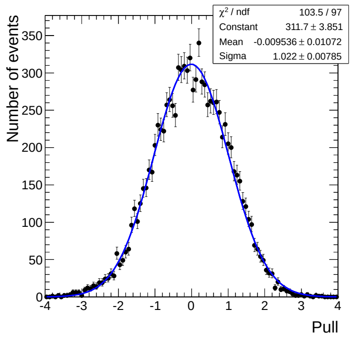
(Fig. 7.9). Nevertheless, the method remains able to provide stable estimations for $N_{t\bar{t}\text{-}like}$, even for events with at least 6 jets for which an under-estimated b-tagging efficiency is compensated by an over-estimated mis-tagging efficiency, as shown in the pull distributions.



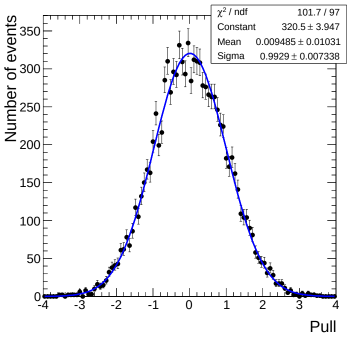
(a) $N_{t\bar{t}\text{-}like}$, 3 jets



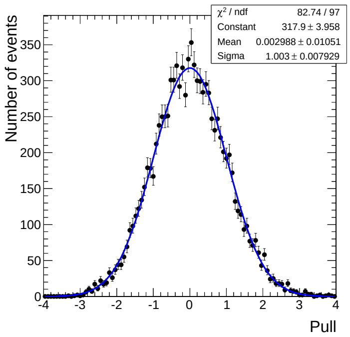
(b) ϵ_{btag}, 3 jets



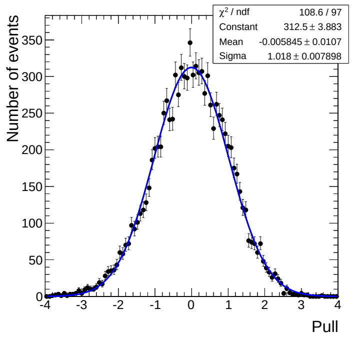
(c) ϵ_{mistag}, 3 jets



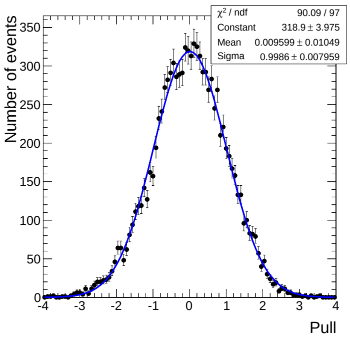
(d) $N_{t\bar{t}\text{-}like}$, 4 jets



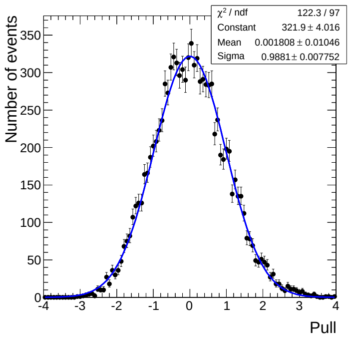
(e) ϵ_{btag}, 4 jets



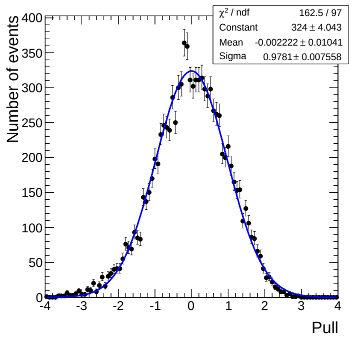
(f) ϵ_{mistag}, 4 jets



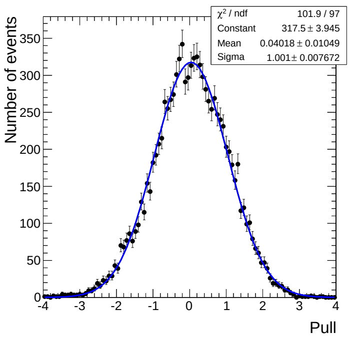
(g) $N_{t\bar{t}\text{-}like}$, 5 jets



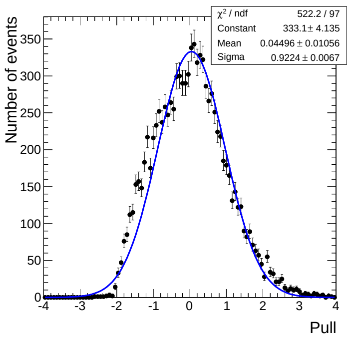
(h) ϵ_{btag}, 5 jets



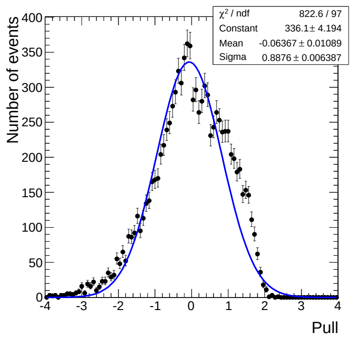
(i) ϵ_{mistag}, 5 jets



(j) $N_{t\bar{t}\text{-}like}$, ≥ 6 jets



(k) ϵ_{btag}, ≥ 6 jets



(l) ϵ_{mistag}, ≥ 6 jets

Figure 7.9: Scenario II-ϵ_{btag}-ϵ_{mistag} : Pull distributions of the estimated numbers of $t\bar{t}$-$like$ events and of the estimated b-tagging and mis-tagging efficiencies for $t\bar{t}$-$like$ events at 1000 pb^{-1} of integrated luminosity, for different jet multiplicities using 10000 pseudo-experiments and the *TrackCountingHighEfficiency* b-tagging algorithm with a discriminator threshold of 2.03.

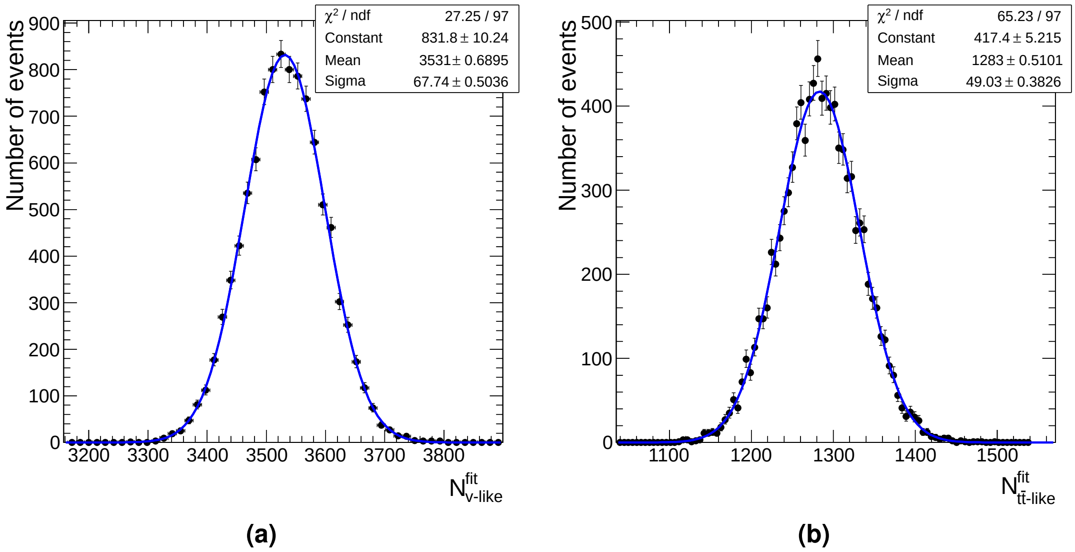


Figure 7.10: Scenario 0 : Distributions of the estimated numbers of $V\text{-}like$ (a) and $t\bar{t}\text{-}like$ (b) events with three reconstructed jets within the detector acceptance with a transverse momentum greater than 30 GeV$/c$, using 10000 pseudo-experiments and the *TrackCountingHighEfficiency* b-tagging algorithm with a discriminator threshold of 2.03. Each pseudo-experiment uses pseudo-data representing 200 pb^{-1} of integrated luminosity.

Estimation of the statistical uncertainty

For each pseudo-experiment, the value of each free parameter is estimated. The statistical uncertainties on these estimations are finally considered as equal to the standard deviation of the distributions of the estimated values, obtained by fitting these distributions with a Gaussian function (Fig 7.10). These statistical uncertainties have been calculated at an integrated luminosity of 200 pb^{-1} and 1000 pb^{-1} for different jet multiplicities, using the *TrackCountingHighEfficiency* b-tagging algorithm with a discriminator threshold of 2.03.

Scenario 0 :

Table 7.2 shows the estimated numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events and their associated statistical uncertainties as a function of the jet multiplicity and of the integrated luminosity. The last column also shows the overall relative statistical uncertainty at a given luminosity. As expected, the relative statistical uncertainty increases with the jet multiplicity but as the number of events decreases quite quickly with the jet multiplicity, the overall relative uncertainty remains low ; for $N_{V\text{-}like}$, it amounts to 1.8 % with an integrated luminosity of 200 pb^{-1} and to 0.8 % with 1000 pb^{-1}. These numbers become 2.4 and 1.1 for $N_{t\bar{t}\text{-}like}$.

	Int. lum. (pb^{-1})	Jet multiplicity 3	4	5	6	Rel. Uncert.
$N_{V\text{-}like}$	200	3531 ± 68	706 ± 35	120 ± 17	29 ± 8	1.8 %
	1000	17659 ± 153	3532 ± 76	601 ± 37	148 ± 19	0.8 %
$N_{t\bar{t}\text{-}like}$	200	1283 ± 49	982 ± 38	387 ± 23	160 ± 15	2.4 %
	1000	6419 ± 110	4908 ± 85	1937 ± 52	805 ± 32	1.1 %

Table 7.2: Scenario 0 : Estimated numbers of V-$like$ and $t\bar{t}$-$like$ events and their associated statistical uncertainties for different jet multiplicities and integrated luminosities, using the *TrackCountingHighEfficiency* b-tagging algorithm with a discriminator threshold of 2.03.

Scenario I-ϵ_{btag} :

Table 7.3 shows the estimated numbers of V-$like$ and $t\bar{t}$-$like$ events, as well as the estimated b-tagging efficiency and their associated statistical uncertainties as a function of the jet multiplicity and of the integrated luminosity. The last column also shows the overall relative statistical uncertainty associated to $N_{V\text{-}like}$ and $N_{t\bar{t}\text{-}like}$ at a given luminosity. As expected, the overall uncertainty on the estimated number of V-$like$ events is only marginally affected by the fact that the b-tagging efficiency is now treated as a free parameter with respect to the scenario 0 : 2.3 % and 1.0 % for an integrated luminosity of 200 pb^{-1} and 1000 pb^{-1} respectively. To a lesser extent, the overall uncertainty on the estimated number of $t\bar{t}$-$like$ events remains also low : 3.3 % and 1.5 % for 200 pb^{-1} and 1000 pb^{-1} respectively. A statistical uncertainty on the estimated b-tagging efficiency below 5 % and 2 % is achievable for 200 pb^{-1} and 1000 pb^{-1} respectively.

Scenario I-ϵ_{mistag} :

Table 7.4 shows the estimated numbers of V-$like$ and $t\bar{t}$-$like$ events, as well as the estimated mis-tagging efficiency for $t\bar{t}$-$like$ events and their associated statistical uncertainties as a function of the jet multiplicity and of the integrated luminosity. The last column also shows the overall relative statistical uncertainty associated to $N_{V\text{-}like}$ and $N_{t\bar{t}\text{-}like}$ at a given luminosity. As in the previous scenario, the uncertainty on the number of V-$like$ events is expected not to be affected by the fact that the mis-tagging efficiency for $t\bar{t}$-$like$ events is now treated as a free parameter with respect to the scenario 0. Indeed, the estimated values are almost identical to those of scenario 0 : 1.9 % and 0.8 % for 200 pb^{-1} and 1000 pb^{-1} of integrated luminosity respectively.

	Int. lum.	Jet multiplicity				
	(pb^{-1})	3	4	5	≥ 6	Rel. uncert.
$N_{V\text{-}like}$	200	3529 ± 86	704 ± 46	120 ± 22	31 ± 14	$2.3\ \%$
	1000	17658 ± 192	3531 ± 101	600 ± 53	148 ± 27	$1.0\ \%$
$N_{t\bar{t}\text{-}like}$	200	1285 ± 72	984 ± 49	388 ± 29	161 ± 17	$3.3\ \%$
	1000	6418 ± 162	4909 ± 109	1939 ± 64	806 ± 38	$1.5\ \%$
ϵ_{btag}	200	0.76 ± 0.03	0.76 ± 0.03	0.77 ± 0.04	0.77 ± 0.07	
	1000	0.76 ± 0.01	0.76 ± 0.01	0.77 ± 0.02	0.77 ± 0.03	

Table 7.3: Scenario I-ϵ_{btag} : Estimated numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events and b-tagging efficiencies with their associated statistical uncertainties for different jet multiplicities and integrated luminosities, using the *TrackCountingHighEfficiency* b-tagging algorithm with a discriminator threshold of 2.03.

	Int. lum.	Jet multiplicity				
	(pb^{-1})	3	4	5	≥ 6	Rel. Uncert.
$N_{V\text{-}like}$	200	3531 ± 71	706 ± 36	120 ± 18	28 ± 11	$1.9\ \%$
	1000	17659 ± 159	3532 ± 79	601 ± 40	148 ± 21	$0.8\ \%$
$N_{t\bar{t}\text{-}like}$	200	1283 ± 52	982 ± 39	387 ± 24	160 ± 15	$2.5\ \%$
	1000	6418 ± 119	4908 ± 88	1937 ± 53	805 ± 33	$1.1\ \%$
ϵ_{mistag}	200	0.101 ± 0.014	0.095 ± 0.011	0.100 ± 0.015	0.108 ± 0.022	
	1000	0.101 ± 0.007	0.095 ± 0.006	0.100 ± 0.007	0.108 ± 0.010	

Table 7.4: Scenario I-ϵ_{mistag} : Estimated numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events and mis-tagging efficiencies for $t\bar{t}\text{-}like$ events with their associated statistical uncertainties for different jet multiplicities and integrated luminosities, using the *TrackCountingHighEfficiency* b-tagging algorithm with a discriminator threshold of 2.03.

Estimator	Int. lum. (pb^{-1})	Jet multiplicity 3	4	5	≥ 6	Rel. Uncert.
$N_{V\text{-}like}$	200	3532 ± 85	707 ± 57	122 ± 28	-18 ± 34	$2.5\ \%$
	1000	17658 ± 195	3534 ± 125	600 ± 72	149 ± 36	$1.1\ \%$
	2000	35315 ± 275	7065 ± 176	1209 ± 101	301 ± 60	$0.3\ \%$
$N_{t\bar{t}\text{-}like}$	200	1282 ± 72	980 ± 59	384 ± 35	160 ± 18	$3.6\ \%$
	1000	6419 ± 161	4906 ± 130	1932 ± 81	802 ± 50	$1.6\ \%$
	2000	12836 ± 225	9812 ± 185	3870 ± 115	1604 ± 73	$1.1\ \%$
ϵ'_{mistag}	200	0.062 ± 0.006	0.079 ± 0.016	0.084 ± 0.046	-0.019 ± 0.133	
	1000	0.062 ± 0.003	0.079 ± 0.007	0.087 ± 0.020	0.087 ± 0.044	
	2000	0.061 ± 0.002	0.079 ± 0.006	0.087 ± 0.014	0.090 ± 0.031	

Table 7.5: Scenario I-ϵ'_{mistag} : Estimated numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events and mis-tagging efficiencies for $V\text{-}like$ events with their associated statistical uncertainties for different jet multiplicities and integrated luminosities, using the *TrackCountingHighEfficiency* b-tagging algorithm with a discriminator threshold of 2.03. Gray-shaded cells correspond to a jet multiplicity and an integrated luminosity for which the method is not able to provide a robust estimation (cf. section *Study of the pull distributions*).

Scenario I-ϵ'_{mistag} :

Table 7.5 shows the estimated numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events, as well as the estimated mis-tagging efficiency for $V\text{-}like$ events and their associated statistical uncertainties as a function of the jet multiplicity and of the integrated luminosity. The last column also shows the overall relative statistical uncertainty associated to $N_{V\text{-}like}$ and $N_{t\bar{t}\text{-}like}$ at a given luminosity. Compared to the previous scenario, estimations on both $N_{V\text{-}like}$ and $N_{t\bar{t}\text{-}like}$ are affected by the introduction of an additional free parameter, ϵ'_{mistag}. Nevertheless, it remains possible with 2000 pb^{-1} of integrated luminosity to obtain overall relative uncertainties on $N_{V\text{-}like}$ and $N_{t\bar{t}\text{-}like}$ as low as those obtained in scenario 0 at 1000 pb^{-1} : $0.8\ \%$ for $N_{V\text{-}like}$ and $1.1\ \%$ for $N_{t\bar{t}\text{-}like}$. The relative uncertainty on ϵ'_{mistag} ranges from $3.2\ \%$ for events with 3 jets to $36\ \%$ for events with at least 6 jets.

Scenario II-ϵ_{btag}-ϵ_{mistag} :

Table 7.6 shows the estimated numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events, as well as the estimated b-tagging and mis-tagging efficiencies for $t\bar{t}\text{-}like$ events and their associated statistical uncertainties as a function of the jet multiplicity and of the

Estimator	Int. lum. (pb^{-1})	Jet multiplicity 3	4	5	≥ 6	Rel. Uncert.
$N_{V\text{-}like}$	1000	17658 ± 195	3531 ± 103	599 ± 55	146 ± 34	$1.0\ \%$
	2000	35315 ± 274	7062 ± 148	1201 ± 79	294 ± 48	$0.7\ \%$
$N_{t\bar{t}\text{-}like}$	1000	6418 ± 166	4909 ± 111	1940 ± 66	808 ± 43	$1.5\ \%$
	2000	12835 ± 232	9816 ± 158	3878 ± 94	1613 ± 60	$1.1\ \%$
ϵ_{btag}	1000	0.76 ± 0.02	0.76 ± 0.02	0.77 ± 0.03	0.77 ± 0.04	
	2000	0.76 ± 0.01	0.76 ± 0.01	0.77 ± 0.02	0.77 ± 0.03	
ϵ_{mistag}	1000	0.101 ± 0.008	0.095 ± 0.007	0.100 ± 0.009	0.108 ± 0.014	
	2000	0.101 ± 0.006	0.095 ± 0.005	0.100 ± 0.007	0.108 ± 0.011	

Table 7.6: Scenario II-ϵ_{btag}-ϵ_{mistag} : Estimated numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events and b-tagging efficiencies for $V\text{-}like$ and $t\bar{t}\text{-}like$ events with their associated statistical uncertainties for different jet multiplicities and integrated luminosities, using the *TrackCountingHighEfficiency* b-tagging algorithm with a discriminator threshold of 2.03. Gray-shaded cells correspond to a jet multiplicity and an integrated luminosity for which the method is not able to provide a robust estimation (cf. section *Study of the pull distributions*).

integrated luminosity. The last column also shows the overall relative statistical uncertainty associated to $N_{V\text{-}like}$ and $N_{t\bar{t}\text{-}like}$ at a given luminosity. In this scenario, with 1000 pb^{-1} of integrated luminosity, the method is able to estimate the numbers of $N_{V\text{-}like}$ and $N_{t\bar{t}\text{-}like}$ events with a relative uncertainty comparable to the one obtained in scenario I-ϵ_{btag}. In addition, the b-tagging and mis-tagging efficiencies are estimated for events with up to 5 jets with comparable relative uncertainties to those obtained in scenarios with only one free tagging efficiency parameter, scenario I-ϵ_{btag} and scenario I-ϵ'_{mistag}, respectively.

7.2.2 Performance summary

The performances of the method have been studied for various combinations of free parameters. It has bee found that as long as the mis-tagging efficiency for $V\text{-}like$ is used as fixed parameter, the method is able to provide robust estimates of the number of $V\text{-}like$ and $t\bar{t}\text{-}like$ events as well as robust estimates of the b-tagging and mis-tagging efficiencies for $t\bar{t}\text{-}like$ events. With 1000 pb^{-1} of integrated luminosity, the expected statistical precision of the estimated number of V-$like$ and $t\bar{t}$-$like$ events is of the order of the percent and remains under $10\ \%$ for the estimated b-tagging and mis-tagging efficiencies.

The main advantage of having the mis-tagging efficiency for $t\bar{t}$-$like$ events used as a free parameter is that so far, independent analyses in CMS have provided measurement of the mis-tagging efficiency using multi-jet events. Although suitable for being used as input for the mis-tagging efficiency for V-$like$ events, it is not expected to be compatible with the mis-tagging efficiency for $t\bar{t}$-$like$ events whose flavour composition is different from multi-jet events.

7.3 Improvement of the estimation

In order to improve the stability of the fitting procedure, another strategy has been investigated ; for a given b-tagging algorithm, the distribution of the number of b-tagged jets per event is a function of the b-discriminant threshold applied while the numbers of V-$like$ and $t\bar{t}$-$like$ events are constants. Therefore, it is possible to improve the method by simultaneously fitting the event samples obtained with different b-discriminant thresholds with probability density functions sharing common parameters. The estimators of the numbers of V-$like$ and $t\bar{t}$-$like$ events are used as common free parameters to all the probability density functions while estimators for the b-tagging and mis-tagging efficiency are introduced for each b-discriminant threshold.

For illustration purposes, three different b-discriminant thresholds have been considered. As previously, several scenarios have been defined and the method performances have been assessed by calculating the pull distributions as well as the distributions of the estimated parameters.

7.3.1 Statistical properties of the estimators

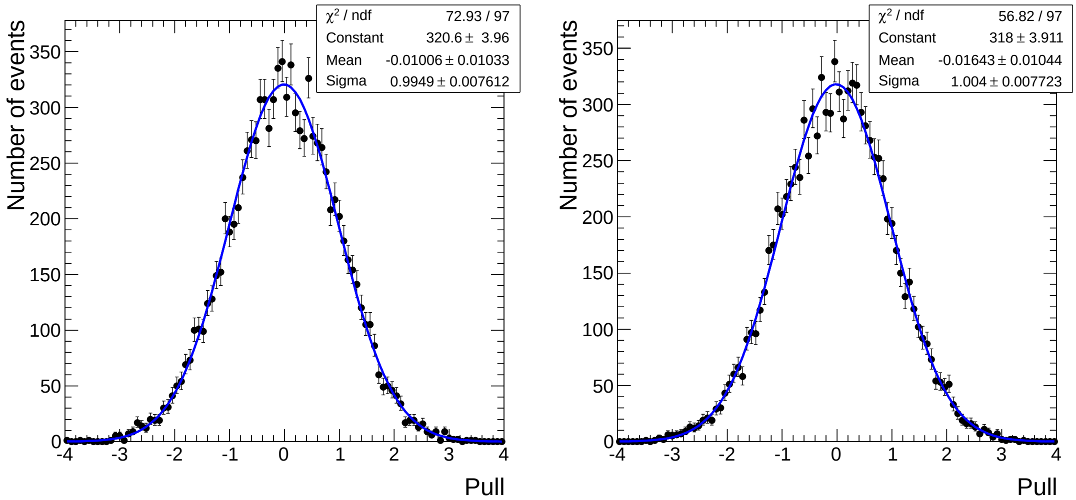
As previously, the pull and the relative statistical uncertainty on the estimated number of selected $t\bar{t}$-$like$ or V-$like$ events, as well as on the estimated b-tagging and mis-tagging efficiencies, have been evaluated using pseudo-experiments. Each pseudo-experiment consists in generating a sample of pseudo-data from the probability density function defined in Eq. (7.4) for each b-discriminant threshold. Then, the value of the parameter of interest is estimated by fitting simultaneously the three samples of pseudo-data, using an extended likelihood technique.

The procedure to generate these three samples of pseudo-data does not take into account the correlations between the distributions of the number of b-tagged jets per event obtained with the different b-discriminant thresholds. Indeed, an event with 0 b-tagged jet using a b-discriminant threshold of 2.03 cannot exhibit one or more b-tagged jets when using a higher b-discriminant threshold. In order to account for these correlations during the generation of pseudo-data, it would have been necessary to calculate the different conditional probabilities for an event to have Y b-tagged jets with a certain b-discriminant threshold, given it has X b-tagged jets with a lower b-discriminant threshold. Then, to generate pseudo-data with a b-tagging and mis-tagging efficiencies corresponding to the lowest b-discriminant threshold and generate the other pseudo-data by using a multinomial law whose parameters are the conditional probabilities previously defined. However, because of the limited number of simulated events, it has been found impossible to estimate reliably these conditional probabilities. Finally, it has been decided to generate uncorrelated pseudo-data. Therefore, the statistical uncertainties obtained for the estimated numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events are likely to underestimated. However, additional constraints could be provided to these estimations by imposing relation orders between the b-tagging (mis-taaging) estimators, resulting in reduced statistical uncertainties on the estimated numbers of events.

Study of the pull distributions

As in section 7.2.1, the pull distributions have been calculated using the *TrackCountingHighEfficiency* b-tagging algorithm with b-discriminant thresholds of 2.03, 3.2 and 4.38, for different jet multiplicities, from 3 jets to ≥ 6 jets, with pseudo-data corresponding to different amounts of integrated luminosity.

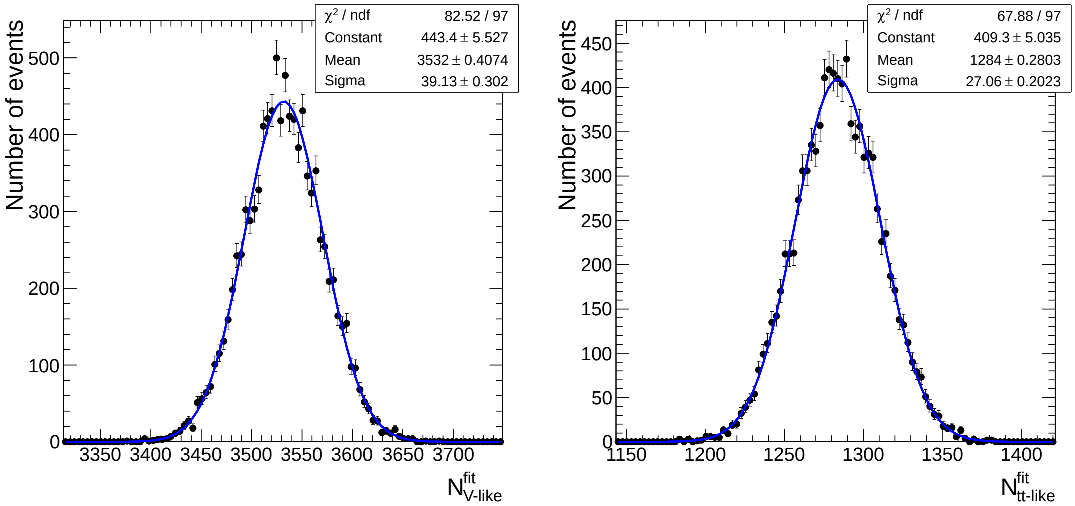
For the scenario 0+, the Gaussian fits to the pull distributions of the two free parameters, $N_{V\text{-}like}$ and $N_{t\bar{t}\text{-}like}$ (Fig 7.11), have a χ^2 per degree of freedom smaller than 1 for all the different jet multiplicities at 1000 pb^{-1} of integrated luminosity. No significant bias is observed at any jet multiplicity. The uncertainties on the estimated parameters seem to be calculated correctly as the standard deviations of all the Gaussian fits are compatible with unity.



(a) $N_{V\text{-}like}$, ≥ 6 jets **(b)** $N_{t\bar{t}\text{-}like}$, ≥ 6 jets

Figure 7.11: Scenario 0+ : Pull distributions of the estimated number of V-$like$ events for events with at least 6 jets at 1000 pb^{-1} of integrated luminosity, using 10000 pseudo-experiments and the *TrackCountingHighEfficiency* b-tagging algorithm with discriminator thresholds of 2.03, 3.2 and 4.38.

The pull distributions for the other scenarios can be found in Appendix A.1.2. It has been observed that the stability of the method is slightly improved when using simultaneously data sets with different b-discriminant thresholds. For the Scenario I+-ϵ_{btag} and I+-ϵ_{mistag}, the method remains able to provide stable estimations for all the parameters with 1000 pb^{-1} of integrated luminosity ; the pull distributions are compatible with Gaussian distributions whose standard deviations are equal to one. Except a ~ 4 % bias on ϵ_{btag} for at least 6 jets, no significant bias is observed for the other parameters at any jet multiplicity. For the scenario I+-ϵ'_{mistag}, an improvement is observed compared to the scenario I ; the ϵ'_{mistag} pull distributions with at least 6 jets become compatible with Gaussian distributions already at 2000 pb^{-1} of integrated luminosity. A $5-6$ % bias is observed, though. Finally, in scenario II+-ϵ_{btag}-ϵ_{mistag}, the compatibility of the pull distributions with Gaussian distributions for ϵ_{btag} and ϵ_{mistag} using events with at least 6 jets is improved but the χ^2 per degree of freedom of the Gaussian fit remain greater than one.



(a) **(b)**

Figure 7.12: Scenario 0+ : Distributions of the estimated numbers of $V\text{-}like$ (a) and $t\bar{t}\text{-}like$ (b) events with three reconstructed jets within the detector acceptance with a transverse momentum greater than 30 GeV$/c$, using 10000 pseudo-experiments and the *TrackCountingHighEfficiency* b-tagging algorithm with discriminator thresholds of 2.03, 3.2 and 4.38. Each pseudo-experiment uses pseudo-data representing 200 pb^{-1} of integrated luminosity.

Estimation of the statistical uncertainty

As in previous section, the statistical uncertainties on the estimations are considered as equal to the standard deviation of the distributions of the estimated values, obtained by fitting these distributions with a Gaussian function (Fig 7.12). These statistical uncertainties have been calculated at an integrated luminosity of 200 pb^{-1} and 1000 pb^{-1} for different jet multiplicities, using the *TrackCountingHighEfficiency* b-tagging algorithm with discriminator thresholds of 2.03, 3.2 and 4.38.

The estimated numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events, as well as the additional estimated parameters, with their statistical uncertainties have been calculated, for the different scenarios, as a function of the jet multiplicity and of the integrated luminosity. The last column also shows the overall relative statistical uncertainty at a given luminosity. By using several b-tagging working points simultaneously, the method has been able to constrain better the parameters, resulting in smaller relative statistical uncertainties than the ones obtained with only one working point. For scenario 0+, the uncertainty on $N_{V\text{-}like}$ amounts to 1.0 % with an integrated luminosity of 200 pb^{-1} and to 0.5 % with 1000 pb^{-1}. These numbers become 1.4 and 0.6 for $N_{t\bar{t}\text{-}like}$ (Tab. 7.7). For scenario I+-ϵ_{btag}, the overall relative

Estimator	Int. lum. (pb^{-1})	Jet multiplicity 3	4	5	≥ 6	Rel. Uncert.
$N_{V\text{-}like}$	200	3532 ± 39	707 ± 20	120 ± 10	30 ± 5	1.0 %
	1000	17661 ± 86	3532 ± 44	602 ± 22	148 ± 11	0.5 %
$N_{t\bar{t}\text{-}like}$	200	1284 ± 27	982 ± 22	387 ± 14	161 ± 8	1.4 %
	1000	6418 ± 61	4908 ± 49	1938 ± 30	805 ± 18	0.6 %

Table 7.7: Scenario 0+ : Estimated numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events with their associated statistical uncertainties for different jet multiplicities and integrated luminosities, using the *TrackCountingHighEfficiency* b-tagging algorithm with a discriminator threshold of 2.03, 3.20 and 4.38.

uncertainties on $N_{V\text{-}like}$ and $N_{t\bar{t}\text{-}like}$ have decreased by ~ 40 % with respect to scenario I, while the relative uncertainty on the estimated b-tagging efficiency remains identical for a b-discriminant threshold of 2.03 (Tab. 7.8). The relative uncertainties on the b-tagging efficiencies estimated with higher b-discriminant thresholds are below 3.5 %. For scenario I+-ϵ_{mistag}, the level of accuracy on the estimations of $N_{V\text{-}like}$ and $N_{t\bar{t}\text{-}like}$ at 1000 pb^{-1} of integrated luminosity is identical to the one obtained in scenario 0+ : 0.5 % and 0.6 % for $N_{V\text{-}like}$ and $N_{t\bar{t}\text{-}like}$ respectively. Furthermore, the relative uncertainty on the mis-tagging efficiency for a b-discriminant threshold of 2.03 is slightly smaller compared to scenario I-ϵ_{mistag} while it is possible to estimate this efficiency at higher b-discriminant thresholds with a relative uncertainty ranging from ~ 9 % to $\sim 15\%$ (Tab.7.9). As in the two previous scenarios, an improvement has been observed for scenario I+-ϵ'_{mistag}, with respect to its corresponding scenario using only one working point (Tab. 7.10) ; at 2000 pb^{-1} of integrated luminosity, the relative uncertainties on the parameters $N_{V\text{-}like}$, $N_{t\bar{t}\text{-}like}$ and ϵ'_{mistag} have decreased on average by ~ 40 %. Finally, Finally, in scenario II+-ϵ_{btag}-ϵ_{mistag}, the statistical uncertainties on the estimated numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events have been reduced by almost 50 % with 1000 pb^{-1} of integrated luminosity. To a lesser extent, an improvement of the statistical uncertainties on the estimated b-tagging and mis-tagging efficiencies for $t\bar{t}\text{-}like$ events has been observed (Tab. 7.11). However, it is worth noticing that the statistical uncertainties on the estimated parameters is comparable to these obtained with less free parameters, scenario I+-ϵ_{btag} or scenario I+-ϵ_{mistag}.

	Int. lum.	Jet multiplicity				Rel. Uncert.
	(pb^{-1})	3	4	5	≥ 6	
$N_{V\text{-}like}$	200	3530 ± 48.4	706 ± 27	120 ± 15	29 ± 9	1.3 %
	1000	17660 ± 106.8	3531 ± 59	601 ± 33	148 ± 16	0.6 %
$N_{t\bar{t}\text{-}like}$	200	1286 ± 40	983 ± 28	388 ± 17	162 ± 11	1.9 %
	1000	6419 ± 88	4910 ± 63	1939 ± 39	805 ± 22	0.8 %
$\epsilon_{btag}^{b\text{-}disc.>2.03}$	200	0.76 ± 0.02	0.76 ± 0.02	0.77 ± 0.04	0.77 ± 0.06	
	1000	0.76 ± 0.01	0.76 ± 0.01	0.77 ± 0.02	0.77 ± 0.02	
$\epsilon_{btag}^{b\text{-}disc.>3.20}$	200	0.66 ± 0.02	0.66 ± 0.02	0.66 ± 0.03	0.66 ± 0.05	
	1000	0.66 ± 0.01	0.66 ± 0.01	0.66 ± 0.01	0.66 ± 0.02	
$\epsilon_{btag}^{b\text{-}disc.>4.38}$	200	0.57 ± 0.02	0.58 ± 0.02	0.57 ± 0.03	0.58 ± 0.05	
	1000	0.57 ± 0.01	0.58 ± 0.01	0.58 ± 0.01	0.58 ± 0.02	

Table 7.8: Scenario I+-ϵ_{btag} : Estimated numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events and b-tagging efficiencies with their associated statistical uncertainties for different jet multiplicities and integrated luminosities, using the *TrackCountingHighEfficiency* b-tagging algorithm with a discriminator threshold of 2.03, 3.20 and 4.38.

	Int. lum. (pb^{-1})	Jet multiplicity 3	4	5	≥ 6	Rel. Uncert.
$N_{V\text{-}like}$	200	3532 ± 40	707 ± 21	120 ± 10	30 ± 5	$1.1\ \%$
	1000	17661 ± 89	3532 ± 45	602 ± 23	148 ± 11	$0.5\ \%$
$N_{t\bar{t}\text{-}like}$	200	1284 ± 29	982 ± 23	387 ± 14	161 ± 8	$1.4\ \%$
	1000	6418 ± 64	4908 ± 50	1938 ± 31	805 ± 19	$0.6\ \%$
$\epsilon_{mistag}^{b\text{-}disc.>2.03}$	200	0.101 ± 0.013	0.095 ± 0.010	0.100 ± 0.015	0.109 ± 0.021	
	1000	0.101 ± 0.006	0.095 ± 0.005	0.100 ± 0.006	0.108 ± 0.009	
$\epsilon_{mistag}^{b\text{-}disc.>3.20}$	200	0.053 ± 0.011	0.049 ± 0.008	0.049 ± 0.010	0.049 ± 0.014	
	1000	0.053 ± 0.005	0.049 ± 0.004	0.049 ± 0.004	0.050 ± 0.006	
$\epsilon_{mistag}^{b\text{-}disc.>4.38}$	200	0.035 ± 0.010	0.032 ± 0.008	0.032 ± 0.009	0.034 ± 0.012	
	1000	0.035 ± 0.005	0.032 ± 0.003	0.032 ± 0.004	0.034 ± 0.005	

Table 7.9: Scenario I+-ϵ_{mistag} : Estimated numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events and mis-tagging efficiencies for $t\bar{t}\text{-}like$ events with their associated statistical uncertainties for different jet multiplicities and integrated luminosities, using the *TrackCountingHighEfficiency* b-tagging algorithm with a discriminator threshold of 2.03, 3.20 and 4.38.

	Int. lum. (pb^{-1})	Jet multiplicity 3	4	5	≥ 6	Rel. Uncert.
$N_{V\text{-}like}$	200	3533 ± 53	708 ± 34	124 ± 18	32 ± 8	1.5 %
	1000	17662 ± 116	3535 ± 76	604 ± 45	152 ± 25	0.7 %
	2000	35317 ± 166	7067 ± 108	1207 ± 65	301 ± 39	0.5 %
$N_{t\bar{t}\text{-}like}$	200	1283 ± 44	980 ± 36	384 ± 21	157 ± 13	2.2 %
	1000	6417 ± 98	4906 ± 79	1935 ± 50	800 ± 31	1.0 %
	2000	12838 ± 142	9814 ± 111	3872 ± 71	1604 ± 46	0.7 %
$\epsilon_{mistag}^{\prime,b\text{-}disc.>2.03}$	200	0.062 ± 0.004	0.079 ± 0.012	0.090 ± 0.033	0.098 ± 0.095	
	1000	0.062 ± 0.002	0.079 ± 0.005	0.087 ± 0.015	0.094 ± 0.032	
	2000	0.062 ± 0.001	0.079 ± 0.004	0.087 ± 0.010	0.092 ± 0.023	
$\epsilon_{mistag}^{\prime,b\text{-}disc.>3.20}$	200	0.025 ± 0.004	0.032 ± 0.010	0.046 ± 0.034	0.046 ± 0.088	
	1000	0.025 ± 0.002	0.032 ± 0.005	0.046 ± 0.014	0.053 ± 0.032	
	2000	0.025 ± 0.001	0.032 ± 0.003	0.046 ± 0.010	0.053 ± 0.022	
$\epsilon_{mistag}^{\prime,b\text{-}disc.>4.38}$	200	0.015 ± 0.003	0.019 ± 0.010	0.019 ± 0.036	0.036 ± 0.080	
	1000	0.015 ± 0.002	0.019 ± 0.004	0.026 ± 0.013	0.023 ± 0.036	
	2000	0.015 ± 0.001	0.019 ± 0.003	0.026 ± 0.009	0.029 ± 0.022	

Table 7.10: Scenario I+-ϵ'_{mistag} : Estimated numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events as well as estimated mis-tagging efficiencies for $V\text{-}like$ events with their associated statistical uncertainties for different jet multiplicities and integrated luminosities, using the *Track-CountingHighEfficiency* b-tagging algorithm with a discriminator threshold of 2.03, 3.20 and 4.38. Gray-shaded cells correspond to a jet multiplicity and an integrated luminosity for which the method is not able to provide a robust estimation (cf. section *Study of the pull distributions*).

	Int. lum. (pb^{-1})	Jet multiplicity 3	4	5	≥ 6	Rel. Uncert.
$N_{V\text{-}like}$	1000	17659 ± 109	3530 ± 60	600 ± 34	147 ± 17	0.6 %
	2000	35319 ± 154	7064 ± 85	1202 ± 47	295 ± 24	0.4 %
$N_{t\bar{t}\text{-}like}$	1000	6419 ± 91	4910 ± 64	1940 ± 40	806 ± 22	0.9 %
	2000	12836 ± 129	9817 ± 90	3877 ± 56	1611 ± 32	0.6 %
$\epsilon_{btag}^{b\text{-}disc.>2.03}$	1000	0.76 ± 0.01	0.76 ± 0.01	0.77 ± 0.02	0.77 ± 0.03	
	2000	0.76 ± 0.01	0.76 ± 0.01	0.77 ± 0.01	0.77 ± 0.02	
$\epsilon_{btag}^{b\text{-}disc.>3.20}$	1000	0.66 ± 0.01	0.66 ± 0.01	0.66 ± 0.02	0.66 ± 0.03	
	2000	0.66 ± 0.01	0.66 ± 0.01	0.66 ± 0.01	0.66 ± 0.02	
$\epsilon_{btag}^{b\text{-}disc.>4.38}$	1000	0.57 ± 0.01	0.58 ± 0.01	0.57 ± 0.02	0.58 ± 0.03	
	2000	0.57 ± 0.01	0.58 ± 0.01	0.57 ± 0.01	0.58 ± 0.02	
$\epsilon_{mistag}^{b\text{-}disc.>2.03}$	1000	0.101 ± 0.008	0.095 ± 0.006	0.100 ± 0.009	0.108 ± 0.013	
	2000	0.101 ± 0.005	0.095 ± 0.004	0.100 ± 0.006	0.108 ± 0.010	
$\epsilon_{mistag}^{b\text{-}disc.>3.20}$	1000	0.053 ± 0.006	0.049 ± 0.004	0.049 ± 0.006	0.050 ± 0.008	
	2000	0.053 ± 0.004	0.049 ± 0.003	0.049 ± 0.004	0.050 ± 0.006	
$\epsilon_{mistag}^{b\text{-}disc.>4.38}$	1000	0.035 ± 0.006	0.032 ± 0.004	0.032 ± 0.005	0.034 ± 0.007	
	2000	0.035 ± 0.004	0.032 ± 0.003	0.032 ± 0.003	0.034 ± 0.005	

Table 7.11: Scenario II+-ϵ_{btag}-ϵ_{mistag} : Estimated numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events as well as estimated b-tagging and mis-tagging efficiencies for $t\bar{t}\text{-}like$ events with their associated statistical uncertainties for different jet multiplicities and integrated luminosities, using the *TrackCountingHighEfficiency* b-tagging algorithm with a discriminator threshold of 2.03, 3.20 and 4.38.

7.3.2 Performance summary

The performances of the improved method have been studied for various combinations of free parameters. Noticeable improvements of the stability of the method have been observed at high jet multiplicity when using simultaneously three different b-tagging working points ; with 1000 pb^{-1} of integrated luminosity, the method is now able to provide robust estimates of the number of V-$like$ and $t\bar{t}$-$like$ events as well as robust estimates of the b-tagging and mis-tagging efficiencies for $t\bar{t}$-$like$ events at any jet multiplicity. The statistical uncertainties on the estimations have been improved too ; with 1000 pb^{-1} of integrated luminosity, the expected statistical precision of the estimated number of V-$like$ and $t\bar{t}$-$like$ events is now below the percent. For the estimated b-tagging and mis-tagging efficiencies, the uncertainties are of the order of these obtained with 2000 pb^{-1} when using only one b-tagging working point.

7.4 Systematic uncertainties

In addition to the statistical uncertainties, estimated in the previous section, the systematic uncertainties on the estimations have been studied. These uncertainties are related to uncertainties on the various inputs to the method such as the b-tagging or mis-tagging efficiencies in scenarios where they are used as fixed parameters or the numbers of multi-jet and single top quark events to be subtracted. Another source of systematic uncertainties arises from the jet energy corrections. Furthermore, the results presented in the previous sections have all been obtained with event samples produced with the `MADGRAPH` generator with given values for the renormalization and factorization scales, the matching threshold for matrix-element calculations and parton showering, as well as for the parameters controlling the amount of initial and final state radiations. Systematic uncertainties have thus also been derived by using other event generators or by using samples generated with modified parameters.

For each free parameter, P, the relative systematic uncertainties, $\Delta P^{syst.}$, have been calculated using the following formula :

$$\Delta P^{Syst.} = \frac{\hat{P}^{Syst.} - P^{Syst.}}{P^{Syst.}} - \frac{\hat{P}^{Nom.} - P^{Nom.}}{P^{Nom.}} \tag{7.12}$$

where $\hat{P}^{Syst.}$ and $P^{Syst.}$ are the estimated and the expected values of the parameter P respectively, taking into account the source of systematic uncertainties while $\hat{P}^{Nom.}$ and $P^{Nom.}$ are the estimated and the expected nominal values.

7.4.1 Uncertainties on b-tagging and mis-tagging efficiencies

The b-tagging and mis-tagging efficiencies are measured from data by independent analyses using multi-jet events. Using collision data collected during the year 2011, representing 500 pb^{-1}, it has been shown these efficiencies can be measured with a relative uncertainty of $10 - 15$ % [199] for the *TrackCountingHighEfficiency* b-tagging algorithm. In order to evaluate the systematic uncertainties on the estimated parameters, the nominal b-tagging and mis-tagging efficiencies have been varied independently within their uncertainties. Efficiencies for the different b-discriminant thresholds have been considered as fully correlated. For each possible combination of shifted efficiencies, 10000 Monte-Carlo simulated samples of pseudo-data have been generated while the nominal efficiencies have been kept during the estimations of the free parameters. For each parameter, the mean of the distribution of the estimated values has then been compared with the mean of the distribution obtained with samples generated with the nominal efficiencies. The minimum and maximum relative differences between these means are considered as the relative systematic uncertainties.

Using 1000 pb^{-1} of integrated luminosity, the relative systematic uncertainties have been calculated for the different combinations of free parameters in order to find the best trade-off between statistical uncertainties, increasing with the number of free parameters and systematic uncertainties, increasing with the number of fixed parameters. It has been found that the total uncertainties are dominated by the systematic uncertainties and, therefore, that the best trade-off is found by using as many free parameters as possible. The systematic uncertainties reach $\sim \pm 4.5$ % for the estimated number of V-$like$ events and $\sim \pm 12$ % for the estimated number of $t\bar{t}$-$like$ events for the scenario 0+ (Tab 7.12). These uncertainties decrease to $\sim \pm 2.0$ % and $\sim \pm 5.5$ % respectively for the scenario II+-ϵ_{btag}-ϵ_{mistag} while the statistical uncertainties on $N_{V\text{-}like}$ and $N_{t\bar{t}\text{-}like}$ only increase by less than 0.3 % (cf. Tab. 7.7 and Tab. 7.11). As the systematic uncertainties on the b-tagging and mis-tagging efficiencies are expected to yield the largest systematic uncertainties of the estimated parameters, it has been decided to only consider this scenario in what follows, when calculating the systematic uncertainties related to other sources.

Estimator	$\epsilon_{btag}/\epsilon_{mistag}/\epsilon'_{mistag}$	Jet multiplicity			
		3	4	5	≥ 6
$N_{V\text{-}like}$	Nominal	17661	3532	602	148
	$+10\,\%/+10\,\%/+10\,\%$	$-4.4\,\%$	$-12.7\,\%$	$-27.4\,\%$	$-44.6\,\%$
	$+10\,\%/+10\,\%/-10\,\%$	$-2.2\,\%$	$-8.9\,\%$	$-21.4\,\%$	$-36.5\,\%$
	$+10\,\%/-10\,\%/+10\,\%$	$-4.0\,\%$	$-11.4\,\%$	$-23.9\,\%$	$-37.7\,\%$
	$+10\,\%/-10\,\%/-10\,\%$	$-1.8\,\%$	$-7.6\,\%$	$-17.8\,\%$	$-29.5\,\%$
	$-10\,\%/+10\,\%/+10\,\%$	$1.9\,\%$	$8.5\,\%$	$20.3\,\%$	$33.9\,\%$
	$-10\,\%/+10\,\%/-10\,\%$	$4.1\,\%$	$12.3\,\%$	$26.3\,\%$	$41.8\,\%$
	$-10\,\%/-10\,\%/+10\,\%$	$2.4\,\%$	$10.4\,\%$	$26.4\,\%$	$46.5\,\%$
	$-10\,\%/-10\,\%/-10\,\%$	$4.6\,\%$	$14.2\,\%$	$32.4\,\%$	$54.7\,\%$
$N_{t\bar{t}\text{-}like}$	Nominal	6418	4908	1938	805
	$+10\,\%/+10\,\%/+10\,\%$	$12.0\,\%$	$9.1\,\%$	$8.5\,\%$	$8.2\,\%$
	$+10\,\%/+10\,\%/-10\,\%$	$6.1\,\%$	$6.4\,\%$	$6.6\,\%$	$6.7\,\%$
	$+10\,\%/-10\,\%/+10\,\%$	$10.9\,\%$	$8.2\,\%$	$7.4\,\%$	$6.9\,\%$
	$+10\,\%/-10\,\%/-10\,\%$	$4.9\,\%$	$5.5\,\%$	$5.5\,\%$	$5.5\,\%$
	$-10\,\%/+10\,\%/+10\,\%$	$-5.3\,\%$	$-6.1\,\%$	$-6.3\,\%$	$-6.2\,\%$
	$-10\,\%/+10\,\%/-10\,\%$	$-11.2\,\%$	$-8.9\,\%$	$-8.2\,\%$	$-7.7\,\%$
	$-10\,\%/-10\,\%/+10\,\%$	$-6.8\,\%$	$-7.5\,\%$	$-8.2\,\%$	$-8.6\,\%$
	$-10\,\%/-10\,\%/-10\,\%$	$-12.7\,\%$	$-10.2\,\%$	$-10.0\,\%$	$-10.1\,\%$

Table 7.12: Scenario 0+ : Relative systematic uncertainties on the estimated numbers of V-$like$ and $t\bar{t}$-$like$ events as a function of the variations ($\pm 10\,\%$) on the b-tagging and mis-tagging efficiencies for different jet multiplicities. Results are presented for an integrated luminosity of 1000 pb^{-1}, using the *TrackCountingHighEfficiency* b-tagging algorithm with a discriminator threshold of 2.03, 3.20 and 4.38.

7.4.2 Number of final-state b quarks from top quark pair decays

As explained in section 7.1.3, different b-tagged jet multiplicity probability functions are used for $t\bar{t}$-$like$ events according to the number of final-state b-jets per event and the fraction of events with 0, 1 and 2 b-jets in the final state is calculated using Monte-Carlo simulations. In order to calculate the systematic uncertainty related to these fractions, pseudo-data have been generated with different fractions of events in the three categories while the nominal values of the fractions have been kept during the estimations. In order to vary these fractions coherently as their sum must add up to 1, it has been decided to parametrize them as a function of the probability for a $t\bar{t}$-$like$ event to have a b-jet in the final state, ϵ_{b-jet}. Assuming independent probabilities for both b-jets, this probability can simply be taken as equal to the square root of the fraction of events with 2 b-jets. The fractions of events with 1 and 0 b-jets are then taken as equal to $2 \times \epsilon_{b-jet} \times (1 - \epsilon_{b-jet})$ and to $(1 - \epsilon_{b-jet})^2$ respectively. The probability ϵ_{b-jet} has been varied and the method has been performed in each case. An associated systematic uncertainty has been calculated by comparing these results with the one obtained with the nominal value of ϵ_{b-jet}. It has checked that a $\pm 5\ \%$ variation on ϵ_{b-jet} is sufficient to cover the range of fractions of events with 0, 1 and 2 b-jets obtained with events generated with MADGRAPH and ALPGEN . The results are summarized in Tab. 7.14.

The variations of the fraction of $t\bar{t}$-$like$ events with 0, 1 and 2 b-jets in the final state affects mainly the estimated b-tagging efficiencies at all the jet multiplicities and the estimated mis-tagging efficiencies for events with at least 5 jets. The systematic uncertainties on the estimated b-tagging efficiencies are of the same order of magnitude as the variation, $\pm 5\ \%$, on ϵ_{b-jet}, indicating that the method compensates for higher (lower) fraction of events with 2 b-jets in the final-state by estimating a higher (lower) b-tagging efficiency than the nominal one.

The systematic uncertainty reaches $+12\ \%$ and $-9\ \%$ for the estimated mis-tagging efficiencies. The effect on the estimated number of V-$like$ and $t\bar{t}$-$like$ events remains below the percent level.

Estimator	ϵ'_{mistag}	Jet multiplicity			
		3	4	5	≥ 6
$N_{V\text{-}like}$	Nominal	17659	3530	600	147
	-10 %	-2.1 %	-3.1 %	-4.9 %	-6.8 %
	$+10$ %	2.0 %	3.2 %	5.0 %	6.3 %
$N_{t\bar{t}\text{-}like}$	Nominal	6419	4910	1940	806
	-10 %	5.7 %	2.2 %	1.5 %	1.2 %
	$+10$ %	-5.5 %	-2.3 %	-1.6 %	-1.1 %
ϵ_{btag} (b-disc th : 2.03)	Nominal	0.76	0.76	0.77	0.77
	-10 %	-2.0 %	-0.7 %	-0.4 %	-0.4 %
	$+10$ %	2.4 %	0.9 %	0.6 %	0.3 %
ϵ_{btag} (b-disc th : 3.20)	Nominal	0.66	0.66	0.66	0.66
	-10 %	-3.2 %	-1.2 %	-0.8 %	-0.7 %
	$+10$ %	3.3 %	1.3 %	0.9 %	0.6 %
ϵ_{btag} (b-disc th : 4.38)	Nominal	0.57	0.58	0.57	0.58
	-10 %	-3.8 %	-1.5 %	-1.0 %	-0.9 %
	$+10$ %	3.8 %	1.6 %	1.1 %	0.9 %
ϵ_{mistag} (b-disc th : 2.03)	Nominal	0.10	0.09	0.10	0.11
	-10 %	-2.7 %	-1.0 %	-0.7 %	-0.2 %
	$+10$ %	2.5 %	1.1 %	0.5 %	0.6 %
ϵ_{mistag} (b-disc th : 3.20)	Nominal	0.05	0.05	0.05	0.05
	-10 %	1.8 %	0.5 %	0.1 %	0.0 %
	$+10$ %	-1.6 %	-0.8 %	-0.4 %	-0.4 %
ϵ_{mistag} (b-disc th : 4.38)	Nominal	0.04	0.03	0.03	0.03
	-10 %	5.8 %	1.5 %	0.6 %	0.7 %
	$+10$ %	-4.0 %	-1.6 %	-1.1 %	-0.5 %

Table 7.13: Scenario II+-ϵ_{btag}-ϵ_{mistag} : Relative systematic uncertainties on the estimated numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events as well as on the estimated b-tagging and mis-tagging efficiencies for $t\bar{t}\text{-}like$ events as a function of the variations (± 10 %) on the mis-tagging efficiency for $V\text{-}like$ events for different jet multiplicities. Results are presented for an integrated luminosity of 1000 pb^{-1}, using the *TrackCountingHighEfficiency* b-tagging algorithm with a discriminator threshold of 2.03, 3.20 and 4.38.

Estimator	ϵ_{b-jet}	Jet multiplicity			
		3	4	5	≥ 6
$N_{V\text{-}like}$	Nominal	17847	3556	601	147
	-5 %	-0.0 %	-0.0 %	-0.9 %	-2.9 %
	$+5$ %	0.0 %	0.0 %	0.8 %	1.7 %
$N_{t\bar{t}\text{-}like}$	Nominal	6230	4884	1938	806
	-5 %	0.0 %	0.0 %	0.3 %	0.6 %
	$+5$ %	-0.0 %	-0.0 %	-0.2 %	-0.3 %
ϵ_{btag} (b-disc th : 2.03)	Nominal	0.77	0.76	0.76	0.76
	-5 %	-4.4 %	-4.4 %	-5.6 %	-6.0 %
	$+5$ %	4.3 %	4.4 %	5.2 %	5.3 %
ϵ_{btag} (b-disc th : 3.20)	Nominal	0.67	0.66	0.65	0.65
	-5 %	-4.6 %	-4.6 %	-5.7 %	-6.3 %
	$+5$ %	4.7 %	4.6 %	5.4 %	5.6 %
ϵ_{btag} (b-disc th : 4.38)	Nominal	0.58	0.58	0.57	0.57
	-5 %	-4.7 %	-4.7 %	-5.8 %	-6.2 %
	$+5$ %	4.8 %	4.7 %	5.5 %	5.7 %
ϵ_{mistag} (b-disc th : 2.03)	Nominal	0.10	0.09	0.10	0.11
	-5 %	-0.0 %	-0.0 %	10.2 %	10.0 %
	$+5$ %	0.2 %	0.1 %	-8.4 %	-8.0 %
ϵ_{mistag} (b-disc th : 3.20)	Nominal	0.05	0.05	0.05	0.05
	-5 %	0.2 %	0.1 %	11.1 %	11.4 %
	$+5$ %	0.1 %	0.0 %	-9.1 %	-9.6 %
ϵ_{mistag} (b-disc th : 4.38)	Nominal	0.04	0.03	0.03	0.03
	-5 %	-0.0 %	0.1 %	11.4 %	11.6 %
	$+5$ %	-0.3 %	0.2 %	-9.4 %	-9.4 %

Table 7.14: Scenario II+-ϵ_{btag}-ϵ_{mistag} : Estimated numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events as well as the estimated b-tagging and mis-tagging efficiencies for $t\bar{t}\text{-}like$ events with their relative systematic uncertainties ($\%$) associated to the variation of fractions of $t\bar{t}\text{-}like$ events with 0, 1 and 2 b-jets in the final state for different jet multiplicities. Results are presented for 1000 pb^{-1} of integrated luminosity, using the *TrackCountingHighEfficiency* b-tagging algorithm with a discriminator threshold of 2.03, 3.20 and 4.38.

7.4.3 Background subtraction

As already mentioned in section 7.2, the multi-jet and single top quark background have to be subtracted before the estimation is performed as they cannot be associated to one of the two event category considered, $V\text{-}like$ or $t\bar{t}\text{-}like$.

The fractions of multi-jet events with 0, 1 or 2 b-tagged jets have been calculated using the default Monte-Carlo simulated multi-jet sample and used to calculate the multi-jet probability density function of the number of b-tagged jets per event. The number of selected multi-jet background events to be subtracted has been calculated in a data-driven way according to the method presented in chapter 6. Single top quark events are treated in a similar way. Monte-Carlo simulated samples are used to calculate both the corresponding b-tagged jet multiplicity density probability function and the number of events to subtract. This number takes into account the $30\ \%$ uncertainty on the single top quark cross-section measured by CMS using the data collected in 2010 [160].

In order to assess the systematic uncertainty associated to this background subtraction, 10000 pseudo-experiments have been carried out. Each pseudo-experiment consists in sampling a data set, representing 1000 pb^{-1} of integrated luminosity, from the probability density function defined in Eq. (7.4), for $V\text{-}like$ and $t\bar{t}\text{-}like$ events but also from the multi-jet and single top quark probability density functions whose parameters have been set to their nominal values. The same functions are used to fit the whole data set. However, during the fit, Gaussian constraints are applied on the estimated numbers of multi-jet and single top quark events. These constraints are included in the model used to fit each pseudo-data set by multiplying the likelihood function with Gaussian functions whose means are set to the nominal numbers of multi-jet and single top quark events used during the generation. The standard deviations are set to the respective uncertainties on these numbers ; $30\ \%$ for the estimated single top quark events and $5\ \%/7\ \%/10\ \%$ for the estimated number of multi-jet events with 3, 4 and 5 jets. These numbers have been extrapolated from the results presented in Chapter 6 for 40 pb^{-1} of integrated luminosity. Indeed, with 1000 pb^{-1}, an estimation of the number of multi-jet events with 5 jets should be feasible with the ABCD method. However, due to the limited amount of simulated multi-jet events, it has not been possible to calculate the probability density function for multi-jet events with 6 jets. Nevertheless, as observed with Monte-Carlo simulations, the numbers of multi-jet events decrease by almost a factor 7 with each additional jet and therefore, the multi-jet background is not expected to contribute significantly to the total number of events with at least 6 jets.

Results have shown that, in this condition, the background subtraction has no significant influence on the accuracy of the estimation of the numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events. However, as shown in Chapter 6, a bias of few percent on the estimated number of multi-jet events might be possible. As the multi-jet background is as large as the $t\bar{t}+jets$ signal for events with three jets and represents $\sim 20\ \%$ of the total background for events with four jets, this bias is expected to have an impact on the estimated numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events. Biases of $2\ \%$ and $5\ \%$ have been applied on the mean of the multi-jet Gaussian function used to fit the pseudo-data sets. Results are presented in Table 7.14. The relative systematic uncertainties on the estimated total number of $V\text{-}like$ events reach $3.8\ \%$ and $9.4\ \%$ for a bias of $2\ \%$ and $5\ \%$ of the estimated number of multi-jet events. It reaches $3.2\ \%$ and $8.3\ \%$ for the estimated total number of $t\bar{t}\text{-}like$ events. For the estimated b-tagging efficiency, the relative systematic uncertainty related to the background subtraction amounts to $\sim 1\ \%$ for events with three jets for a bias of $5\ \%$. In the other cases, as well as for the estimated mis-tagging efficiency, it remains under $0.4\ \%$.

7.4.4 Jet energy scale

As explained in section 4.3.4, after reconstruction, jets need to be calibrated to obtain a flat jet response compatible with one. Systematic uncertainties on the estimations are associated to this calibration. In order to assess these uncertainties, the four-momenta of jets have been rescaled up and down with a constant factor $1+\alpha$, prior to the application of the event selection criteria used in this thesis. The results obtained with the nominal jet energy calibration have been compared with results obtained with $\alpha = \pm 10\ \%$. Modifying the jet energy scale affects obviously the number of selected $V\text{-}like$ and $t\bar{t}\text{-}like$ events but also the average b-tagging and mis-tagging efficiencies as this may modify the flavour composition of the sample of selected events. However, systematic uncertainties related to the b-tagging and mis-tagging efficiencies have already been considered in section 7.4.1. In order to distinguish systematic effects arising from the jet energy scale but not related to the variations of the b-tagging and mis-tagging efficiencies, these efficiencies have been recalculated for each value of α and used to generate 10000 samples of pseudo-data. It has been checked that the variations on the b-tagging and mis-tagging efficiencies induced by modifying the jet energy scale are smaller than $10\ \%$. For each parameter, the mean of the distributions of the estimated values has been compared with the mean of the distribution obtained with samples generated with the nominal parameter values.

Estimator	Bias on the estimation of the nb. of multi-jet events	Jet multiplicity 3	4	5	≥ 6
$N_{V\text{-}like}$	Nominal	17659	3530	600	147
	2 %	−0.5 %	−0.3 %	−0.3 %	0.0 %
	5 %	−1.2 %	−0.9 %	−0.9 %	0.0 %
$N_{t\bar{t}\text{-}like}$	Nominal	6419	4910	1940	806
	2 %	−0.7 %	−0.1 %	0.0 %	0.0 %
	5 %	−1.8 %	−0.3 %	−0.0 %	0.0 %
ϵ_{btag} (b-disc th : 2.03)	Nominal	0.76	0.76	0.77	0.77
	2 %	0.4 %	0.1 %	0.1 %	0.1 %
	5 %	1.0 %	0.2 %	0.1 %	0.1 %
ϵ_{btag} (b-disc th : 3.20)	Nominal	0.66	0.66	0.66	0.66
	2 %	0.4 %	0.1 %	0.0 %	0.0 %
	5 %	0.9 %	0.2 %	0.0 %	0.0 %
ϵ_{btag} (b-disc th : 4.38)	Nominal	0.57	0.58	0.57	0.58
	2 %	0.3 %	0.1 %	0.1 %	−0.1 %
	5 %	0.9 %	0.2 %	0.1 %	−0.1 %
ϵ_{mistag} (b-disc th : 2.03)	Nominal	0.10	0.09	0.10	0.11
	2 %	−0.3 %	0.0 %	−0.1 %	−0.0 %
	5 %	−0.4 %	0.0 %	−0.1 %	−0.0 %
ϵ_{mistag} (b-disc th : 3.20)	Nominal	0.05	0.05	0.05	0.05
	2 %	−0.0 %	−0.2 %	−0.1 %	−0.4 %
	5 %	0.0 %	−0.3 %	−0.1 %	−0.4 %
ϵ_{mistag} (b-disc th : 4.38)	Nominal	0.04	0.03	0.03	0.03
	2 %	0.2 %	−0.1 %	−0.3 %	0.3 %
	5 %	0.3 %	−0.1 %	−0.4 %	0.3 %

Table 7.15: Scenario IIϵ_{btag}-ϵ_{mistag} : Relative systematic uncertainties associated to the subtraction of the single top quark and multi-jet background events as a function of the jet multiplicity. Results are presented for 1000 pb^{-1} of integrated luminosity, using the *TrackCountingHighEfficiency* b-tagging algorithm with a discriminator threshold of 2.03, 3.20 and 4.38.

It has been found that, except for the variations on the b-tagging and mis-tagging efficiencies, other effects induced by a modified jet energy scale do not prevent the method to provide unbiased and robust estimates for the free parameters ; the calculated systematic uncertainties are all compatible with zero.

7.4.5 Uncertainties related to the modelling of $t\bar{t} + jets$ and $V + jets$ processes

As the only fixed parameters of the method are the mis-tagging efficiency for V-$like$ events and the fractions of $t\bar{t}$-$like$ events with 0, 1 and 2 b-jets in the final state, for which related systematic uncertainties have already been calculated, the method is not expected to be sensitive to any other sources of systematic uncertainties.

Monte-Carlo generators

In order to assess the estimation systematic uncertainties related to the modelling of the top quark pairs and of the vector boson production in association with jets, the results obtained with the default samples, produced with the `MADGRAPH` generator, have been compared to results obtained with samples produced with the `ALPGEN` event generator (see section 5.3). Results are all in perfect agreement, except for the estimated mis-tagging efficiency for which a systematic uncertainty of $0.1 - 0.3\ \%$ has been observed, depending the b-discriminant threshold.

Renormalization and factorization scales

As explained in section 2.1, the renormalization and factorization scales are just artefacts of the calculation of any observable at a given order in QCD corrections. As in principle, the corrections at all orders have to be calculated, the choice of these scales does not matter. In practice, any observable is calculated up to a certain order in QCD corrections and therefore, any result for this observable is affected by the choice of these scales. In order to evaluate the systematic uncertainties associated to the particular values of renormalization and factorization scales used produce the default $t\bar{t} + jets$ samples, generated with `MADGRAPH` , the method has been performed with two other $t\bar{t} + jets$ samples, still generated with `MADGRAPH` but with modified scales ; $\mu_R = \mu_F = \mu_0/2$ and $\mu_R = \mu_F = 2 * \mu_0$, where

$\mu_0^2 = 2 * m_{top}^2 + \sum_i p_T^2$ is the reference scale used to generate the default samples (cf. section 5.3). The associated systematic uncertainties on the estimated numbers of V-$like$ and $t\bar{t}$-$like$ events, as well as on the b-tagging efficiency have been found to be smaller than 0.1 % for all jet multiplicities, and smaller than 0.2 % for the estimated mis-tagging efficiency for $t\bar{t}$-$like$ events.

However, it has been shown in section 5.2.2 that not only the magnitude of the renormalization and factorization scales affects the shape and the normalization of differential cross-sections but also their functional form and it would have been therefore interesting to have other samples generated with different kinds of dynamic scales. Nevertheless, the variation of the magnitude of these scales provides already an idea of the sensitivity of the results.

Initial and final state radiations

The systematic uncertainties associated to the amount of initial or final state radiations in $t\bar{t} + jets$ events have been evaluated using two samples generated with modified settings, compared to the settings used to generate the default samples, in order to produce less and more radiations respectively (cf. section 5.3).

Besides the variations on the number of selected $t\bar{t} + jets$ events, it has checked that the modification of the amount of ISR or FSR only induce only small variations on the b-tagging and mis-tagging efficiencies as well as on the fractions of $t\bar{t} + jets$ events with 0, 1 and 2 b-jets in the final-state. As these effects have already been taken into account, the same procedure as for the calculation of the systematic uncertainties related to the jet energy scale has been applied. It has been found that the systematic uncertainties related to the amount of ISR/FSR are smaller than 0.1 %.

7.4.6 Summary of the systematic uncertainties

The list of overall relative statistical and systematic uncertainties are summarized in Table 7.16 for the estimations of the number of V-$like$ and $t\bar{t}$-$like$ events, as well as for the estimated b-tagging and mis-tagging efficiencies for $t\bar{t}$-$like$ events (scenario II+-ϵ_{btag}-ϵ_{mistag}). The total uncertainty on each of these parameters is calculated by adding in quadrature all the aforementioned uncertainties, assuming that they are statistically independent. For collisions data at 7 TeV, representing 1000 pb^{-1} of integrated luminosity, the method presented in this thesis yields an overall relative uncertainty of 2.1 % on the estimated number of V-$like$ events, 3.0 % for the estimated number of $t\bar{t}$-$like$ events, $5.1-5.7$ % for the estimated b-tagging efficiency and $6.6-12.6$ % for the estimated mis-tagging efficiency for $t\bar{t}$-$like$ events, depending on the b-discriminant thresholds.

One of the major contribution to the overall systematic uncertainty, the uncertainty of the mis-tagging efficiency, has been derived from independent analyses using data collected in year 2010 and therefore is likely to improve with 1000 pb^{-1} of integrated luminosity, decreasing significantly the overall systematic uncertainty. On the other hand, data collected by CMS in year 2011 will suffer from additional collisions occurring during the same proton bunch crossing, called pile-up events. To study such effects, dedicated Monte-Carlo simulated samples are needed that were not available at the moment the analysis was performed. Although this effect is expected to be a source of significant systematic uncertainty, it is also expected that a collective effort from the CMS collaboration will be performed to reduce this effect at its minimum as it affects all measurements using the data collected by the CMS detector.

7.5 Conclusions

This chapter presents an original method to estimate, from data, the number of background events with a weak boson associated with jets, passing the top quark pair event selection criteria. After multi-jet and single top quark background subtraction, events are classified according to their numbers of b-jets ; on one hand, events with 0 b-jets, called V-$like$ events, contain mainly $V+jets$ events and, on the other hand, events with 2 b-jets, called $t\bar{t}$-$like$ events, which contain mainly $t\bar{t}+jets$ events but also $V+b\bar{b}$ events. The number of remaining events with only one b-jets, $V+b$, is negligible. B-tagging algorithms are then used to identify these b-jets and the number of V-$like$ and $t\bar{t}$-$like$ are finally estimated from the observed numbers of events with 0, 1 or 2 b-tagged jets, using an extended maximum likelihood technique. For that purpose, the probability density function of the random variable whose outcome represents the number of b-tagged jet per event as a function of the number of b-quarks present in the event final state has been calculated ; it is a function of the b-tagging and mis-tagging efficiencies and an explicit mathematical formulation has been given. The flavour composition of the event final state being different for V-$like$ and $t\bar{t}$-$like$ events, a mis-tagging efficiency for both event categories has been introduced.

Relative uncertainties	$N_{V\text{-}like}$	$N_{t\bar{t}\text{-}like}$	ϵ_{btag}	ϵ_{mistag}
			(b-disc th : $2.03/3.2/4.38$)	
statistical (1000 pb^{-1})	0.6 %	0.9 %	$1.3/1.5/1.7$ %	$6.3/8.2/12.5$ %
systematic				
relative bias	≤ 0.1 %	≤ 0.1 %	$-1.9/-1.7/2.6$ %	$-1.7/0.2/0.7$
mis-tagging eff. ϵ'_{mistag}	1.8 %	2.7 %	$0.9/1.3/1.6$ %	$1.1/0.8/1.6$ %
ϵ_{b-jet}, ± 5 %	≤ 0.1 %	≤ 0.1 %	$4.4/4.6/4.7$ %	$0.1/0.1/0.2$ %
bckgd. subtraction (bias : 5 %)	0.9 %	0.8 %	$0.2/0.2/0.2$ %	$0.0/-0.3/-0.1$ %
Jet energy scale ($\alpha = \pm 10$ %)	≤ 0.1 %	≤ 0.1 %	≤ 0.1 %	≤ 0.1 %
Monte-Carlo generators	≤ 0.1 %	≤ 0.1 %	≤ 0.1 %	$-0.2/-0.1/-0.3$ %
μ_F/μ_R	≤ 0.1 %	≤ 0.1 %	≤ 0.1 %	$0.1/0.1/0.2$ %
ISR/FSR	≤ 0.1 %	≤ 0.1 %	≤ 0.1 %	≤ 0.1 %
Total	2.1 %	3.0 %	$5.1/5.3/5.9$ %	$6.6/8.2/12.6$ %

Table 7.16: Scenario II+-ϵ_{btag}-ϵ_{mistag} : Overview of the statistical and systematic uncertainties on the estimated numbers of $V\text{-}like$ and $t\bar{t}\text{-}like$ events as well as in the estimated b-tagging and mis-tagging efficiencies for $t\bar{t}\text{-}like$ events. The systematic uncertainties for ϵ_{btag} and ϵ_{mistag} have been calculated for events with 4 jets, using the *TrackCountingHighEfficiency* b-tagging algorithm with a discriminator threshold (b-disc th.) of 2.03, 3.20 and 4.38.

The possibility to either estimate these efficiencies along with the numbers of V-$like$ and $t\bar{t}$-$like$ or use them as fixed parameters has also been investigated. The performances of the method have been studied using pseudo-experiments with proton-proton collisions simulated at a centre-of-mass energy of 7 TeV. The distributions of the pulls, as well as the distributions of the estimated parameters have been calculated in order to check the presence of possible biases and assess the statistical uncertainties on the estimations. It has been found that the method is able to provide stable and non-biased estimations of the numbers of V-$like$ and $t\bar{t}$-$like$ events, separately for each jet multiplicity, up to 6 or more than 6, together with an estimation of the b-tagging efficiency and of the mis-tagging efficiency for $t\bar{t}$-$like$ events, using data representing 1000 pb^{-1} of integrated luminosity. The statistical uncertainties on these estimations amount to 1.0 % for the number of V-$like$ events and to 1.5 % for the number of $t\bar{t}$-$like$ events. The b-tagging efficiency can be estimated with a statistical uncertainty of 2.0 %. This uncertainty increases to 7.4 % for the mis-tagging efficiency for $t\bar{t}$-$like$ events. In order to improve these performances, the method has been modified to be able to fit simultaneously data sets obtained with three different b-tagging discriminant thresholds. Both the stability of the method at high jet multiplicity and the statistical uncertainties of the estimations have been improved ; with 1000 pb^{-1} of integrated luminosity, the statistical uncertainties of the estimated number of V-$like$ and $t\bar{t}$-$like$ events are now below the percent, 0.6 % and 0.9 % respectively. The lowest uncertainties on the estimated b-tagging efficiencies have been obtained for events with 4 jets and range from 1.3 % to 1.7 %, depending on the b-discriminant thresholds. These numbers become 6.3 % and 12.5 % for the estimated mis-tagging efficiency for $t\bar{t}$-$like$ events. Finally, systematic uncertainties on the estimations, from various sources, have been studied for the improved method, using three different b-discriminant thresholds ; it has been found that the main source of systematic uncertainties on the estimated numbers of events is related to the uncertainties on the b-tagging or mis-tagging efficiencies when they are used as fixed parameters. Therefore, the best trade-off between statistical and systematic uncertainties has been obtained by using both the b-tagging and the mis-tagging efficiencies as free parameters, despite the increase of statistical uncertainties. Combined statistical and systematic uncertainties have been calculated for each free parameter ; using 1000 pb^{-1} of integrated luminosity, the method is able to estimate the number of V-$like$ with a combined uncertainty of 2.1 % and the number of $t\bar{t}$-$like$ events with a combined uncertainty of 3.0 %. Depending on the b-discriminant threshold, the combined uncertainty ranges from 5.1 % to 5.9 % for the b-tagging efficiency and from 6.6 % to 12.6 % for the mis-tagging efficiency.

Smoke whirls
After the passage of a train.
Young foliage.
Shiki Masaoka (1867-1902)

8 Conclusions

The current theoretical framework in Particle Physics, called Standard Model, describes all the known particles and their interactions. Over the last decades, its predictions have been found in agreement with experimental results to an astonishing level of precision for interactions occurring at higher and higher energies. However, the Standard Model cannot be considered as the final theory in particle physics as it exhibits several shortcomings ; it does not include gravity and fails in providing satisfactory answers to some experimental observations. A possible way to overcome this issue is to assume that the Standard Model is a low-energy approximation of a more fundamental theory. Many extensions of the Standard Model have been proposed, including Super-Symmetry or theories involving extra spatial dimensions. The search for signs of such new physics is an essential part of the physics program of the new CERN accelerator, the Large Hadron Collider (LHC).

In several extensions of the Standard Model, new production or decay channels of top quark pairs are predicted for proton-proton collisions at the LHC, which would lead to an excess in the observed top quark production with respect to the Standard Model predictions. Unfortunately, several background processes also contribute to the total number of selected top quark pair event candidates. This thesis describes the research work done to address the issue of the estimation of the two main background processes, namely the multi-jet production and the vector boson production in association with jets. This analysis focuses on the top quark pair semi-muonic decay channel.

As the multi-jet cross-section is several orders of magnitude greater than the top quark pair production cross-section and suffers from large theoretical uncertainties, even a small selection efficiency leads to a significant amount of multi-jet events after the event selection criteria are applied. Furthermore, the selection efficiency is difficult to estimate accurately as it requires to generate a large data set of multi-jet events. Therefore, it is require to estimate the multi-jet background from data.

The possibility to estimate such a background, with the so-called ABCD method, has been studied using simulated data. This method uses two statistically independent variables and therefore pairs of candidate variables have been studied and a methodology to control their statistical dependence, possibly with data, has been introduced. A pair of suitable variables has been identified, namely the muon relative isolation and the muon transverse impact parameter significance. The performances of the method have been evaluated using simulated proton-proton collisions at a centre-of-mass energy of 7 TeV, first with simulated multi-jet events only and then, in realistic conditions, with events from all the processes : multi-jet production, top quark pair and single top quark productions, single vector boson and vector boson pair productions. It has been shown that the bias due to the presence of non-multi-jet events in the control regions can be reduced by defining non-contiguous signal and control regions. Finally, the statistical uncertainty on the estimated number of multi-jet events has been calculated as a function of the integrated luminosity ; for 1000 pb^{-1}, a relative statistical uncertainty of $3\ \%$ is obtained for events with three jets reconstructed within the detector acceptance with a transverse momentum higher than 30 GeV$/c$. For events with at least four jets, the uncertainty increases to $7\ \%$. However, the ratio of the number of selected multi-jet events to the number of selected top quark pair events decreases from ~ 1 for events with three jets to ~ 0.14 for events with at least four jets. Despite a multi-jet event selection efficiency of $\sim 2.5 \times 10^{-8}$ and thus the limited amount of simulated multi-jet events passing the top quark pair selection criteria, an attempt has been made to calculate the systematic uncertainties on the estimation, related to the calibration of the jet energy scale, to the uncertainties on the two variables used by the method and to the uncertainties on the cross-sections of the non-multi-jet processes. Emphasize has been put on the methodology rather than on the results as the calculated systematic uncertainties are smaller than the statistical uncertainties on the estimation, due to the limited amount of simulated multi-jet events.

This ABCD method is attractive as the only assumption made by this method is the statistical independence of the two selected variables, used to define the control regions. This assumption can be evaluated with the data, using the methodology described in this case. In case of a linear correlation between the two variables, it should be possible to fit the fractions of events with $X < X_0$ ($Y < Y_0$) over the range $Y > Y_0$ ($X > X_0$), which is supposed to be dominated by the multi-jet background and extrapolate the fit in the signal-dominated region. Only in case of non-linear correlation, other methods should be envisaged. Furthermore, with enough integrated luminosity, the ABCD method can be used to predict the distribution of different variables, which is particularly useful in the context of search for signs of new physics. The principle consists in dividing the range of the variable of interest in several intervals and to perform the ABCD method in each interval. However, a careful optimisation of the interval widths needs to be investigated as large intervals would reduce the statistical uncertainty on each estimation but would dilute the observability of potential signs of new physics.

After having applied the top quark pair selection criteria, the main source of remaining background events is the production of single vector boson associated with jets. Indeed, when the vector boson decays leptonically, it leads to an isolated lepton in the final state and therefore these background events cannot be rejected with the same efficiency as the multi-jet events and it prevents the use of the ABCD method to estimate this background. the theoretical uncertainties on the single boson production cross-section are potentially large at high jet multiplicity. An original method has been developed to estimate, from data, this particular background.

After multi-jet and single top quark background subtraction, events are classified according to their numbers of b-jets ; on one hand, events with 0 b-jets, called V-$like$ events, contain mainly $V + jets$ events and, on the other hand, events with 2 b-jets, called $t\bar{t}$-$like$ events, which contain mainly $t\bar{t} + jets$ events but also $V + b\bar{b}$ events. The number of remaining events with only one b-jets, $V + b$, is negligible. B-tagging algorithms are then used to identify these b-jets and the number of V-$like$ and $t\bar{t}$-$like$ are finally estimated from the observed numbers of events with 0, 1 or 2 b-tagged jets, using an extended maximum likelihood technique. For that purpose, the probability density function of the random variable whose outcome represents the number of b-tagged jet per event as a function of the number of b-quarks present in the final state has been calculated ; it is a function of the b-tagging and mis-tagging efficiencies and an explicit mathematical formulation has been given. The flavour composition of the event final state being different for V-$like$ and $t\bar{t}$-$like$ events, a different mis-tagging efficiency for both event categories has been introduced.

The possibility to either estimate these efficiencies along with the numbers of V-$like$ and $t\bar{t}$-$like$ or use them as fixed parameters has also been investigated. The performances of the method have been studied using pseudo-experiments with proton-proton collisions simulated at a centre-of-mass energy of 7 TeV. The distributions of the pulls, as well as the distributions of the estimated parameters have been calculated in order to check the presence of possible biases and assess the statistical uncertainties on the estimations. It has been found that the method is able to provide stable and non-biased estimations of the numbers of V-$like$ and $t\bar{t}$-$like$ events, separately for each jet multiplicity, up to 6 or more than 6, together with an estimation of the b-tagging efficiency and of the mis-tagging efficiency for $t\bar{t}$-$like$ events, using data representing 1000 pb^{-1} of integrated luminosity. The statistical uncertainties on these estimations amount to 1.0 % for the number of V-$like$ events and to 1.5 % for the number of $t\bar{t}$-$like$ events. The b-tagging efficiency can be estimated with a statistical uncertainty of 2.0 %. This uncertainty increases to 7.4 % for the mis-tagging efficiency for $t\bar{t}$-$like$ events. In order to improve these performances, the method has been modified to be able to fit simultaneously data sets obtained with three different b-tagging discriminant thresholds. Both the stability of the method at high jet multiplicity and the statistical uncertainties of the estimations have been improved ; with 1000 pb^{-1} of integrated luminosity, the statistical uncertainties of the estimated number of V-$like$ and $t\bar{t}$-$like$ events are now below the percent, 0.6 % and 0.9 % respectively. The lowest uncertainties on the estimated b-tagging efficiencies have been obtained for events with 4 jets and range from 1.3 % to 1.7 %, depending on the b-discriminant thresholds. These numbers become 6.3 % and 12.5 % for the estimated mis-tagging efficiency for $t\bar{t}$-$like$ events. Finally, systematic uncertainties on the estimations, from various sources, have been studied for the improved method, using three different b-discriminant thresholds ; it has been found that the main source of systematic uncertainties on the estimated numbers of events is related to the uncertainties on the b-tagging or mis-tagging efficiencies when they are used as fixed parameters. Therefore, the best trade-off between statistical and systematic uncertainties has been obtained by using both the b-tagging and the mis-tagging efficiencies for $t\bar{t}$-$like$ events as free parameters, despite the increase of the statistical uncertainties. It has been also found that the main uncertainty on the estimated b-tagging efficiency arises from the fractions of $t\bar{t}$-$like$ events with 0, 1 or 2 b-jets in the final state, calculated from Monte-Carlo simulations and used as inputs for the method. A 5 % uncertainty on these fractions is translated into a ~ 4.5 % systematic uncertainty on the estimated b-tagging efficiency. The uncertainties on the estimated mis-tagging efficiencies are by far dominated by the statistical uncertainties. Combined statistical and systematic uncertainties have been calculated for each free parameter ; using 1000 pb^{-1} of integrated luminosity, the method is able to estimate the number of $V\text{-}like$ with a combined uncertainty of 2.1 % and the number of $t\bar{t}\text{-}like$ events with a combined uncertainty of 3.0 %. Depending on the b-discriminant threshold, the combined uncertainty ranges from 5.1 % to 5.9 % for the b-tagging efficiency and from 6.6 % to 12.6 % for the mis-tagging efficiency.

Unlike methods based on distribution templates derived from Monte-Carlo simulations, this method is insensitive to uncertainties related to the modelling of the signal and background processes. As shown in this thesis, these uncertainties might be large for V+jets processes at high jet multiplicities. Furthermore, as the estimation is performed per jet multiplicity, the method is also insensitive the jet energy scale uncertainty, which is usually a source of sizeable systematic uncertainty.

Another advantage of this method consists in the possibility to simultaneously estimate the b-tagging and mis-tagging efficiencies. In the case of the b-tagging efficiency, the total uncertainty obtained with this method is competitive with the uncertainty of measurements performed with methods developed within the CMS b-tagging group, namely *PtRel* and *System8*. In addition, unlike these two methods, the present method does not rely on any template derived from Monte-Carlo simulations. Besides, these two methods measure the b-tagging efficiency using di-jet events, which present a different topology from the top quark pairs. Therefore, the present method could provide a valuable alternative measurement, especially for top quark pair cross-section measurements requiring b-tagged jets in the final state (+ clair). The mis-tagging efficiency is also measured using di-jet events and provided to the CMS collaboration by the same dedicated group. But the mis-tagging efficiency is dependent of the final-state flavour composition of the process of interest. Therefore, this measurement cannot be used directly in analyses concerning top quark pairs. As, up to now, it does not exist any method which does not rely on Monte-Carlo simulations in order to provide such measurement, the present method provides a valuable measurement for analyses where top quark pair final-state are selected using b-tagged jets.

As signs of new physics may appear as new production channels for top quark pairs, visible as a localized excess of selected events in distributions, for example, the top quark pair invariant mass distribution. The present method may also be used to estimate the V+jets background contribution to such distributions, as it can also provide the estimated number of V+jets events as a function of the b-tagged jet multiplicity. The distribution of a given variable for events with 0 b-tagged jet is dominated by the V+jets contribution and can therefore be subtracted to the same distribution but obtained after having required the events to have 2 b-tagged jets, which is thus dominated by $t\bar{t}$+jets events. Then, the newly obtained distribution can be used to subtract the $t\bar{t}$+jets contribution to the distribution with

0 b-tagged jet. It is possible to reiterate this procedure until the distributions are no longer modified. The final distribution with 2 b-tagged jets, corrected for the V+jets contribution, can be compared with the observed one and thus be used to search for potential deviations due to new physics.

Throughout this thesis, the background estimation methods have been developed and validated using Monte-Carlo simulated samples. Although attempts have been made to account for the systematic uncertainties originating from possible mis-modelling of the background and signal processes, the validation procedure needs to be carried out further with real proton collision data. As more than 1000 pb^{-1} of data have been recently collected by the CMS detector, it is now possible and it should be envisaged to use these methods to provide alternative estimations to the one already used within the CMS collaboration, as these methods are, to a large extent, data-driven and could be used not only to estimate the number of background events but also to estimate the shape of differential distributions in the context of searches for signs of new physics.

Le savant n'étudie pas la nature parce que cela est utile;
il l'étudie parce qu'il y prend plaisir,
et il y prend plaisir parce qu'elle est belle.

Henry Poincaré (1854-1912)

A

Appendices

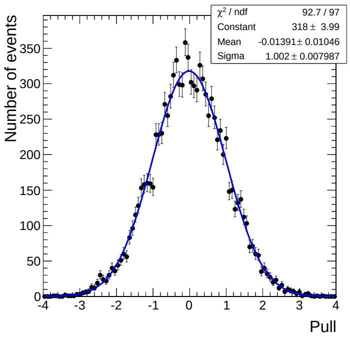
A.1 Estimation of the vector boson background for semi-muonic $t\bar{t}$ studies

A.1.1 B-tagged jet multiplicity probability equations

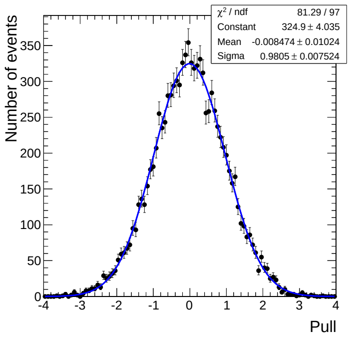
$$
\begin{aligned}
p(0\ b\text{-}tagged\ jet|t\bar{t}\text{-}like) =&(1-\epsilon_{btag})*(1-\epsilon_{btag})*(1-\epsilon_{mistag})^{n-2}*e^{2\ b\text{-}jets}\\
&+(1-\epsilon_{btag})*(1-\epsilon_{mistag})^{n-1}*e^{1\ b\text{-}jet}\\
&+(1-\epsilon_{mistag})^{n}*e^{0\ b\text{-}jet} \qquad \text{(A.1)}\\
p(1\ b\text{-}tagged\ jet|t\bar{t}\text{-}like) =&(2*\epsilon_{btag}*(1-\epsilon_{btag})*(1-\epsilon_{mistag})^{n-2}+(1-\epsilon_{btag})^{2}*(n-2)*\epsilon_{mistag}*(1-\epsilon_{mistag})^{n-3})*e^{2\ b\text{-}jets}\\
&+(\epsilon_{btag}*(1-\epsilon_{mistag})^{n-1}+(1-\epsilon_{btag})*(n-1)*\epsilon_{mistag}*(1-\epsilon_{mistag})^{n-2})*e^{1\ b\text{-}jet}\\
&+(n*\epsilon_{mistag}*(1-\epsilon_{mistag})^{n-1})*e^{0\ b\text{-}jet} \qquad \text{(A.2)}\\
p(2\ b\text{-}tagged\ jets|t\bar{t}\text{-}like) =&(\epsilon_{btag}^{2}*(1-\epsilon_{mistag})^{n-2}\\
&\quad+2*\epsilon_{btag}*(1-\epsilon_{btag})*(n-2)*\epsilon_{mistag}*(1-\epsilon_{mistag})^{n-3}\\
&\quad+(1-\epsilon_{btag})^{2}*((n-2)*(n-3)/2)*\epsilon_{mistag}^{2}*(1-\epsilon_{mistag})^{n-4})*e^{2\ b\text{-}jets}\\
&+(\epsilon_{btag}*(n-1)*\epsilon_{mistag}*(1-\epsilon_{mistag})^{n-2}\\
&\quad+(1-\epsilon_{btag})*((n-1)*(n-2)/2)*\epsilon_{mistag}^{2}*(1-\epsilon_{mistag})^{n-3})*e^{1\ b\text{-}jet}\\
&+((n*(n-1)/2)*\epsilon_{mistag}^{2}*(1-\epsilon_{mistag})^{n-2})*e^{0\ b\text{-}jet} \qquad \text{(A.3)}\\
p(3\ b\text{-}tagged\ jets|t\bar{t}\text{-}like) =&(\epsilon_{btag}^{2}*(n-2)*\epsilon_{mistag}*(1-\epsilon_{mistag})^{n-3}\\
&\quad+2*\epsilon_{btag}*(1-\epsilon_{btag})*((n-2)*(n-3)/2)*\epsilon_{mistag}^{2}*(1-\epsilon_{mistag})^{n-4}\\
&\quad+(1-\epsilon_{btag})^{2}*((n-2)*(n-3)*(n-4)/6)*\epsilon_{mistag}^{3}*(1-\epsilon_{mistag})^{n-5})*e^{2\ b\text{-}jets}\\
&+(\epsilon_{btag}*((n-1)*(n-2)/2)*\epsilon_{mistag}^{2}*(1-\epsilon_{mistag})^{n-3}\\
&\quad+(1-\epsilon_{btag})*((n-1)*(n-2)*(n-3)/6)*\epsilon_{mistag}^{3}*(1-\epsilon_{mistag})^{n-4})*e^{1\ b\text{-}jet}\\
&+((n*(n-1)*(n-2)/6)*\epsilon_{mistag}^{3}*(1-\epsilon_{mistag})^{n-3})*e^{0\ b\text{-}jet}\\
&\qquad \text{(A.4)}
\end{aligned}
$$

Hereby, $e^{x\ b\text{-}jets}$ denotes the probability for a $t\bar{t}$-$like$ events to have x b-jets in the final state.

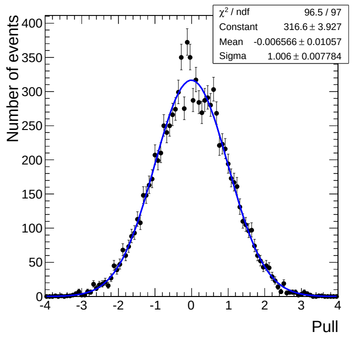
A.1.2 Pull distributions for the improved method



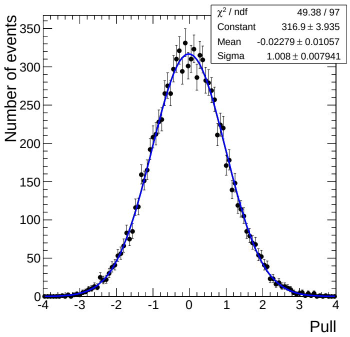
(a) b-th : 2.03, 3 jets



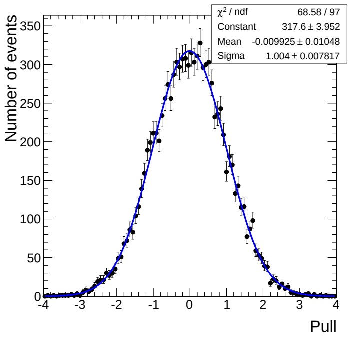
(b) b-th : 3.20, 3 jets



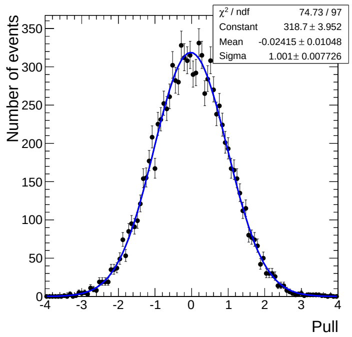
(c) b-th : 4.38, 3 jets



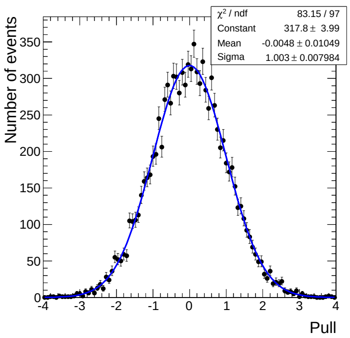
(d) b-th : 2.03, 4 jets



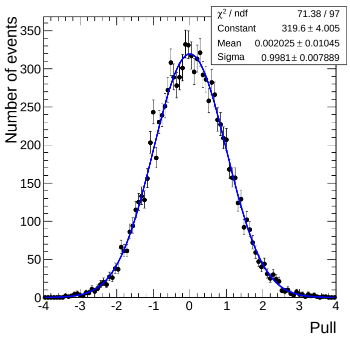
(e) b-th : 3.20, 4 jets



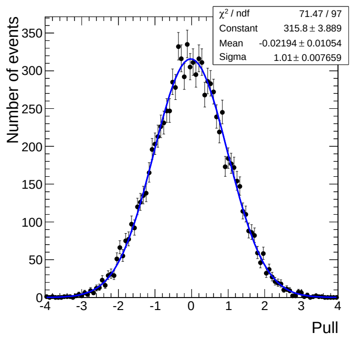
(f) b-th : 4.38, 4 jets



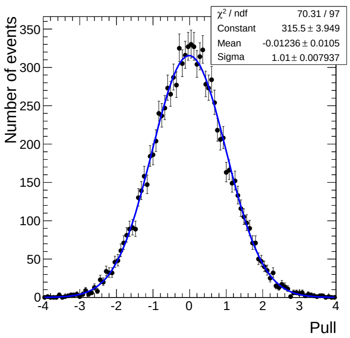
(g) b-th : 2.03, 5 jets



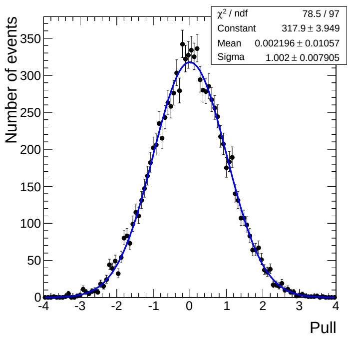
(h) b-th : 3.20, 5 jets



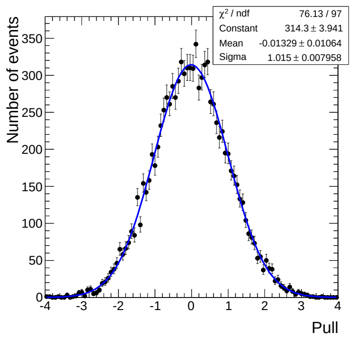
(i) b-th : 4.38, 5 jets



(j) b-th : 2.03, ≥ 6 jets

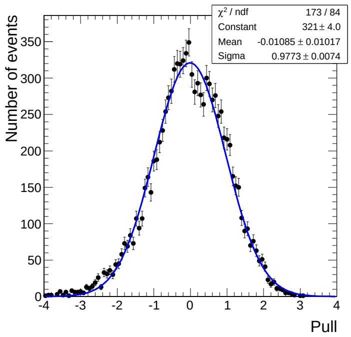


(k) b-th : 3.20, ≥ 6 jets

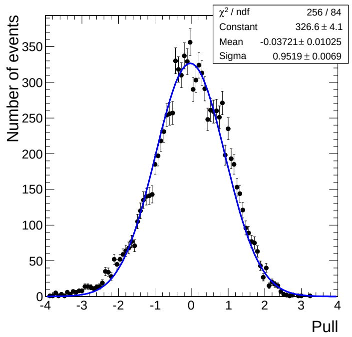


(l) b-th : 4.38, ≥ 6 jets

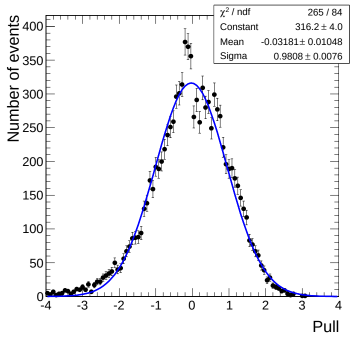
Figure A.1: Scenario I+-ϵ_{btag} : Pull distributions of the estimated b-tagging efficiency for different jet multiplicities at 1000 pb^{-1} of integrated luminosity, using 10000 pseudo-experiments and the *TrackCountingHighEfficiency* b-tagging algorithm with discriminator thresholds (b-th) of 2.03, 3.2 and 4.38.



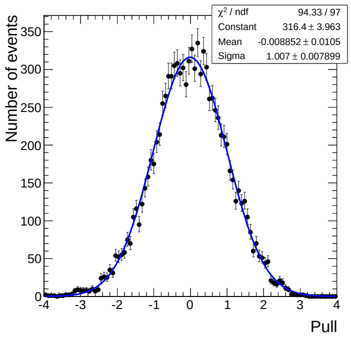
(a) b-th : 2.03, ≥ 6 jets



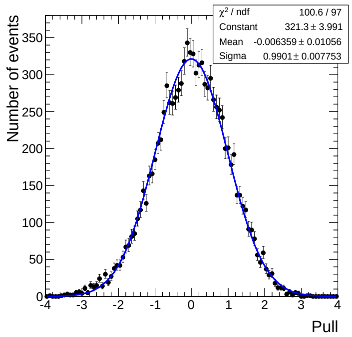
(b) b-th : 3.20, ≥ 6 jets



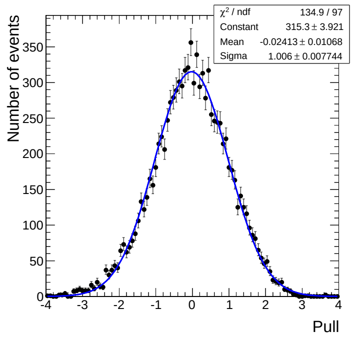
(c) b-th : 4.38, ≥ 6 jets



(d) b-th : 2.03, ≥ 6 jets

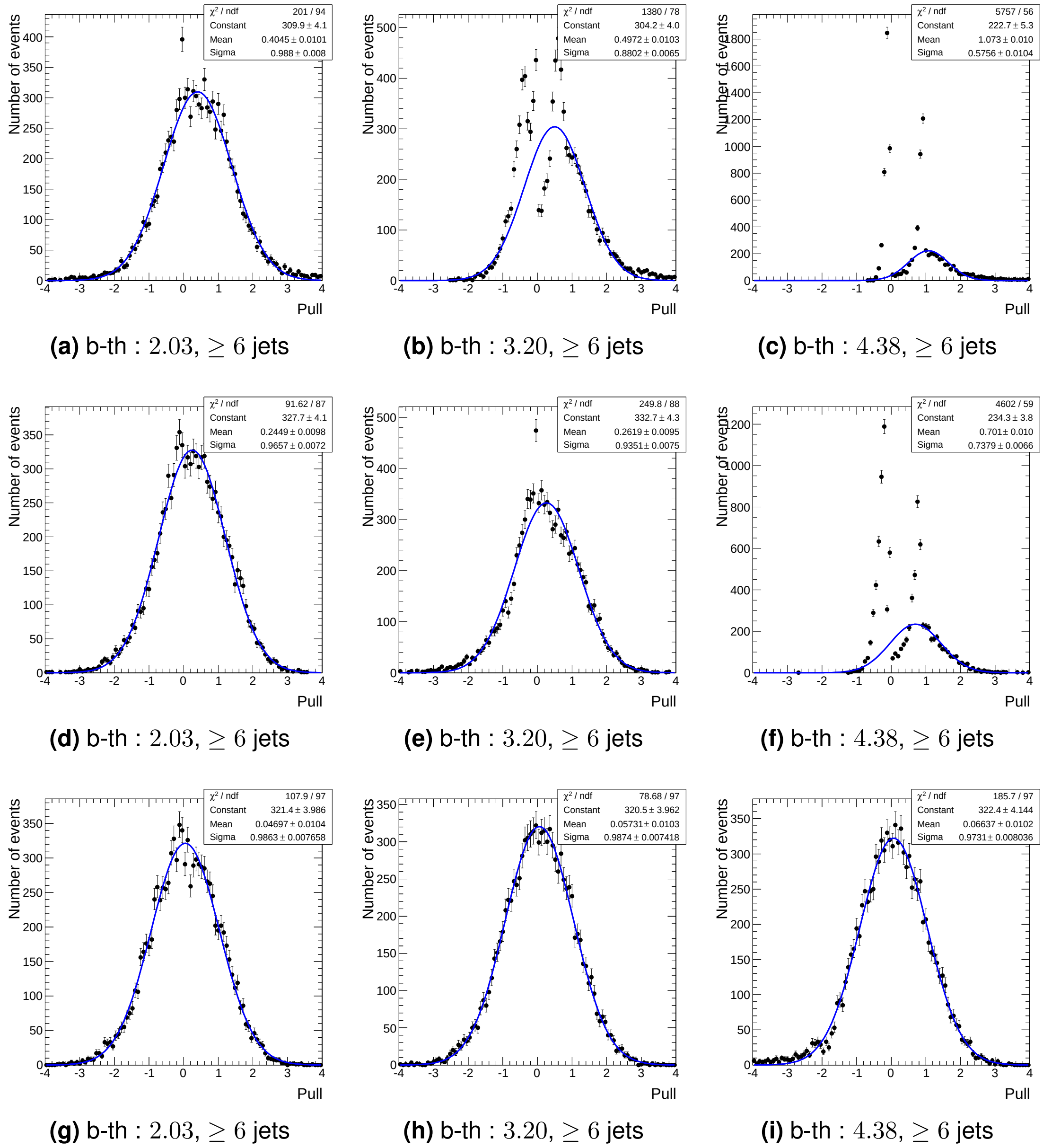


(e) b-th : 3.20, ≥ 6 jets



(f) b-th : 4.38, ≥ 6 jets

Figure A.2: Scenario l+-ϵ_{mistag} : Pull distributions of the estimated mis-tagging efficiency for $t\bar{t}$-$like$ events with at least 6 jets at 200 pb^{-1} (a,b,c) and at 1000 pb^{-1} (d,e,f) of integrated luminosity, using 10000 pseudo-experiments and the *TrackCountingHighEfficiency* b-tagging algorithm with discriminator thresholds (b-th) of 2.03, 3.2 and 4.38.



(a) b-th : 2.03, ≥ 6 jets

(b) b-th : 3.20, ≥ 6 jets

(c) b-th : 4.38, ≥ 6 jets

(d) b-th : 2.03, ≥ 6 jets

(e) b-th : 3.20, ≥ 6 jets

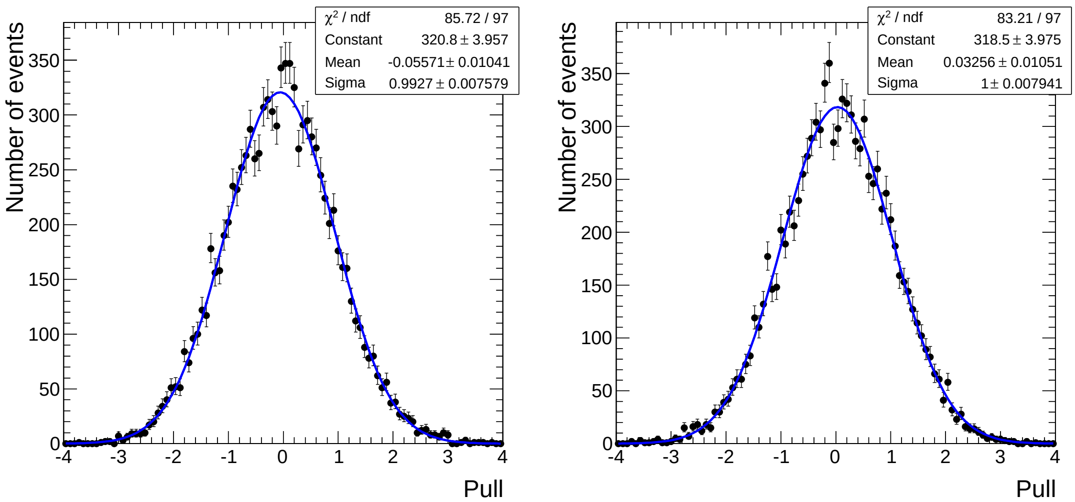
(f) b-th : 4.38, ≥ 6 jets

(g) b-th : 2.03, ≥ 6 jets

(h) b-th : 3.20, ≥ 6 jets

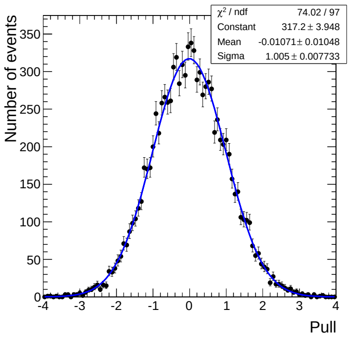
(i) b-th : 4.38, ≥ 6 jets

Figure A.3: Scenario I+-ϵ'_{mistag} : Pull distributions of the estimated mis-tagging efficiency for V-$like$ events with at least 6 jets at 200 pb^{-1} (a,b,c), 1000 pb^{-1} (d,e,f) and 2000 pb^{-1} (g,h,i) of integrated luminosity, using 10000 pseudo-experiments and the *TrackCountingHigh-Efficiency* b-tagging algorithm with discriminator thresholds (b-th) of 2.03, 3.2 and 4.38.

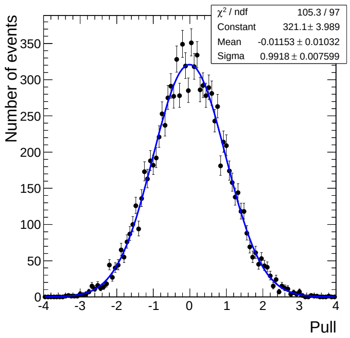


(a) **(b)**

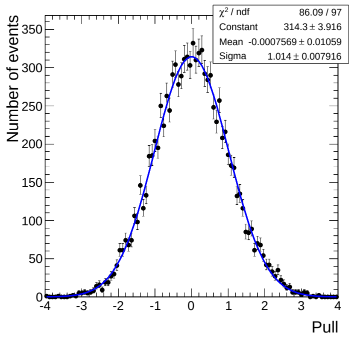
Figure A.4: Scenario II+-ϵ_{btag}-ϵ_{mistag} : Pull distributions of the estimated numbers of V-$like$ (a) and $t\bar{t}$-$like$ (b) events with at least 6 jets at 1000 pb^{-1} of integrated luminosity, using 10000 pseudo-experiments and the *TrackCountingHighEfficiency* b-tagging algorithm with discriminator thresholds (b-th) of 2.03, 3.2 and 4.38.



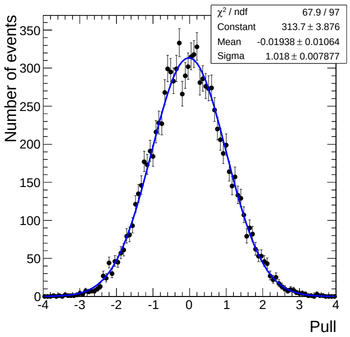
(a) b-th : 2.03, 3 jets



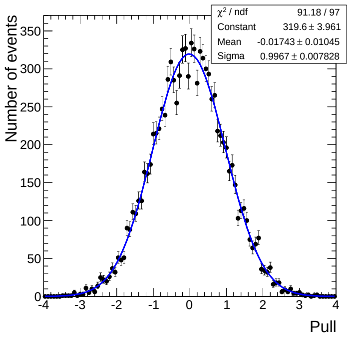
(b) b-th : 3.203, 3 jets



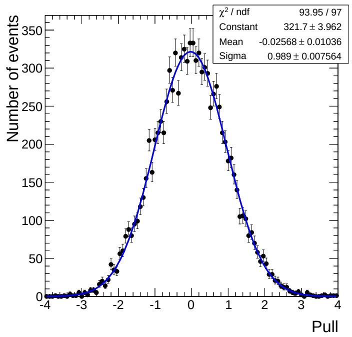
(c) b-th : 4.38, 3 jets



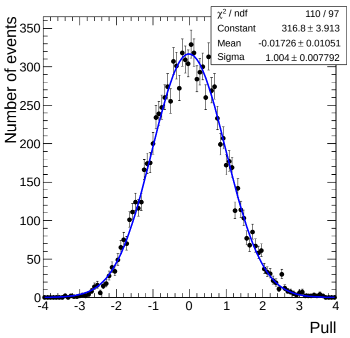
(d) b-th : 2.03, 4 jets



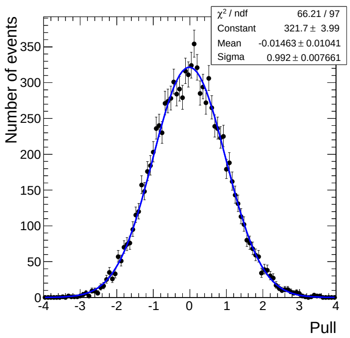
(e) b-th : 3.203, 4 jets



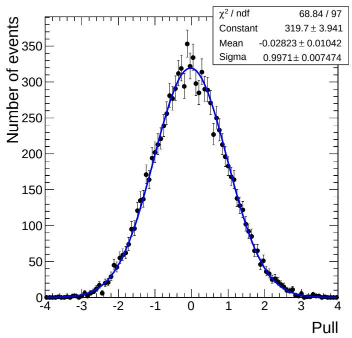
(f) b-th : 4.38, 4 jets



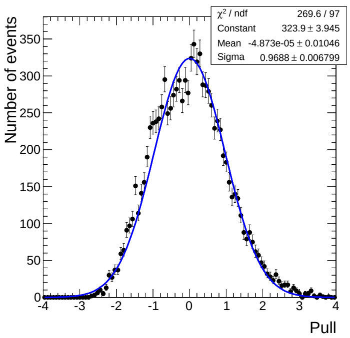
(g) b-th : 2.03, 5 jets



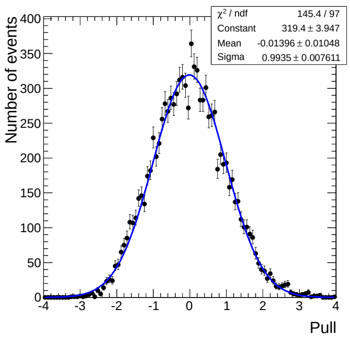
(h) b-th : 3.203, 5 jets



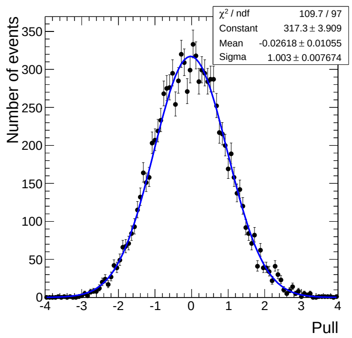
(i) b-th : 4.38, 5 jets



(j) b-th : 2.03, ≥ 6 jets

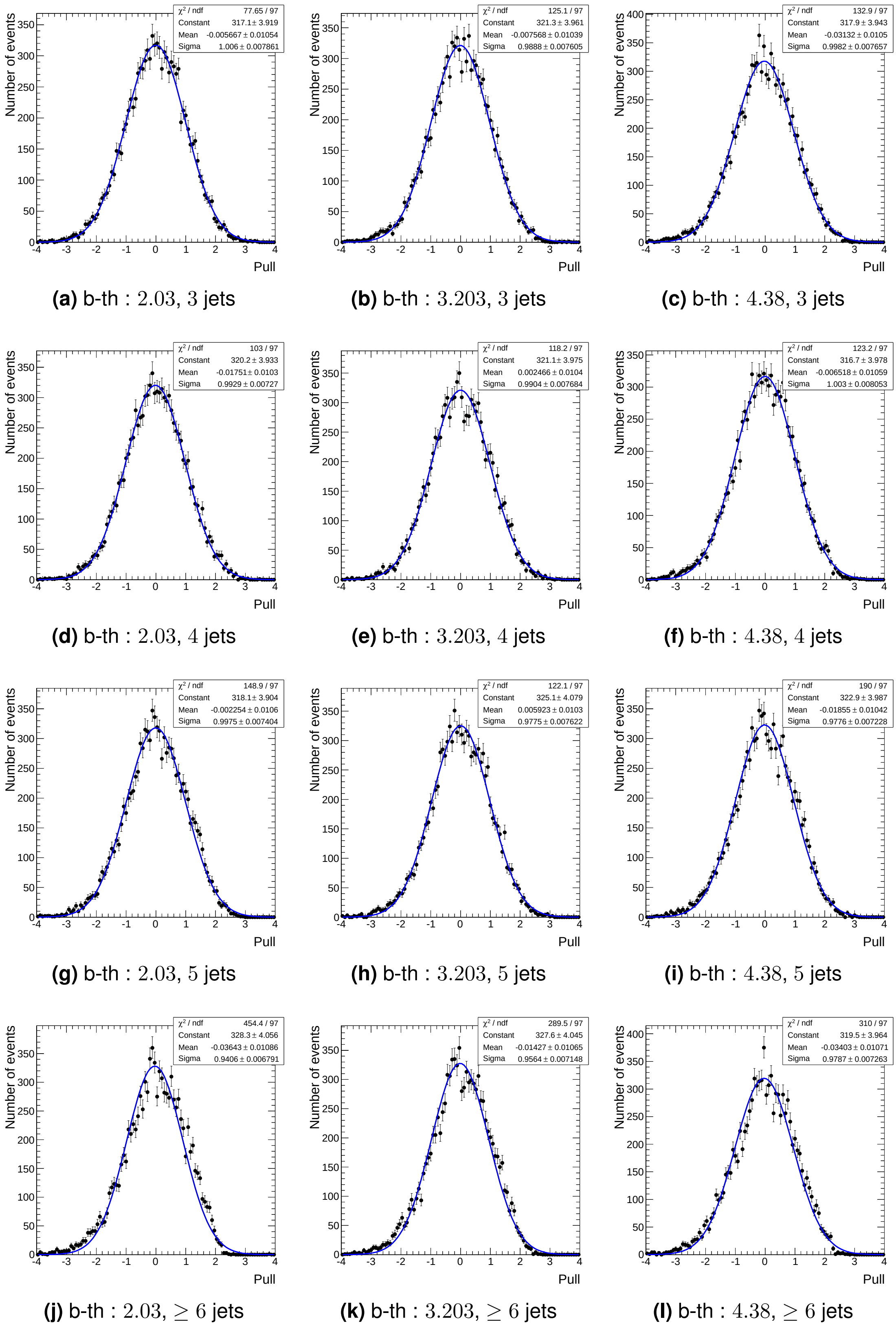


(k) b-th : 3.203, ≥ 6 jets



(l) b-th : 4.38, ≥ 6 jets

Figure A.5: Scenario II+-ϵ_{btag}-ϵ_{mistag} : Pull distributions of the estimated b-tagging efficiency at different jet multiplicities with 1000 pb^{-1} of integrated luminosity, using 10000 pseudo-experiments and the *TrackCountingHighEfficiency* b-tagging algorithm with discriminator thresholds (b-th) of 2.03, 3.2 and 4.38.



(a) b-th : 2.03, 3 jets **(b)** b-th : 3.203, 3 jets **(c)** b-th : 4.38, 3 jets

(d) b-th : 2.03, 4 jets **(e)** b-th : 3.203, 4 jets **(f)** b-th : 4.38, 4 jets

(g) b-th : 2.03, 5 jets **(h)** b-th : 3.203, 5 jets **(i)** b-th : 4.38, 5 jets

(j) b-th : 2.03, ≥ 6 jets **(k)** b-th : 3.203, ≥ 6 jets **(l)** b-th : 4.38, ≥ 6 jets

Figure A.6: Scenario II+-ϵ_{btag}-ϵ_{mistag} : Pull distributions of the estimated mis-tagging efficiency for $t\bar{t}$-$like$ events at different jet multiplicities with 1000 pb^{-1} of integrated luminosity, using 10000 pseudo-experiments and the *TrackCountingHighEfficiency* b-tagging algorithm with discriminator thresholds (b-th) of 2.03, 3.2 and 4.38.

Bibliography

[1] F. Halzen and A. D. Martin, *Quarks and Leptons : An Introductory Course in Modern Particle Physics*. John Wiley and Sons, New York, 1984.

[2] S. Weinberg, "A Model of Leptons," *Physical Review Letters* **19** no. 21, (Nov., 1967) 1264–1266. http://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[3] A. Salam, *Elementary Particle Theory*. Almqvist and Wiksells, Stockholm, 1969.

[4] S. L. Glashow, "Partial-symmetries of weak interactions," *Nuclear Physics* **22** no. 4, (Feb., 1961) 579–588. http://linkinghub.elsevier.com/retrieve/pii/0029558261904692.

[5] A. Salam and J. C. Ward, "Weak and electromagnetic interactions," *Il Nuovo Cimento* **11** no. 4, (Feb., 1959) 568–577. http://www.springerlink.com/index/10.1007/BF02726525.

[6] A. Salam and J. C. Ward, "Electromagnetic and weak interactions," *Physics Letters* **13** no. 2, (Nov., 1964) 168–171. http://linkinghub.elsevier.com/retrieve/pii/0031916364907115.

[7] G. 't Hooft and M. Veltman, "Regularization and renormalization of gauge fields," *Nuclear Physics B* **44** no. 1, (July, 1972) 189–213. http://linkinghub.elsevier.com/retrieve/pii/0550321372902799.

[8] N. Cabibbo, "Unitary Symmetry and Leptonic Decays," *Physical Review Letters* **10** no. 12, (June, 1963) 531–533. http://link.aps.org/doi/10.1103/PhysRevLett.10.531.

[9] M. Kobayashi and T. Maskawa, "CP-Violation in the Renormalizable Theory of Weak Interaction," *Progress of Theoretical Physics* **49** no. 2, (Feb., 1973) 652–657. http://ptp.ipap.jp/link?PTP/49/652/.

[10] M. Gell-Mann, "A schematic model of baryons and mesons," *Physics Letters* **8** no. 3, (Feb., 1964) 214–215. http://linkinghub.elsevier.com/retrieve/pii/S0031916364920013.

[11] G. Zweig, "An SU(3) model for strong interaction symmetry and its breaking I," *CERN Report No.8182/TH.401* (1964) 26.

[12] G. Zweig, "An SU(3) model for strong interaction symmetry and its breaking II," *CERN Report No.8419/TH.412* (1964) 80.

[13] R. Feynman, “Very high-energy collisions of hadrons,” *Physical Review Letters* **23** no. 24, (Dec., 1969) 1415–1417. http://link.aps.org/doi/10.1103/PhysRevLett.23.1415.

[14] Y. Nambu, “Nobel Lecture: Spontaneous symmetry breaking in particle physics: A case of cross fertilization,” *Reviews of Modern Physics* **81** no. 3, (July, 2009) 1015–1018. http://link.aps.org/doi/10.1103/RevModPhys.81.1015.

[15] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, “Theory of Superconductivity,” *Physical Review* **108** no. 5, (Dec., 1957) 1175–1204. http://link.aps.org/doi/10.1103/PhysRev.108.1175.

[16] P. Higgs, “Broken Symmetries and the Masses of Gauge Bosons,” *Physical Review Letters* **13** no. 16, (Oct., 1964) 508–509. http://link.aps.org/doi/10.1103/PhysRevLett.13.508.

[17] F. Englert and R. Brout, “Broken symmetry and the mass of gauge vector mesons,” *Physical Review Letters* **13** no. 9, (Aug., 1964) 321–323. http://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[18] R. K. Ellis, W. J. Stirling, and B. R. Webber, *QCD and Collider Physics*. Cambridge University Press, Cambridge, 1996. http://ebooks.cambridge.org/ref/id/CBO9780511628788.

[19] A. Djouadi, “The anatomy of electroweak symmetry breakingTome I: The Higgs boson in the Standard Model,” *Physics Reports* **457** no. 1-4, (Feb., 2008) 1–216, arXiv:0503172 [hep-ph]. http://linkinghub.elsevier.com/retrieve/pii/S0370157307004334.

[20] M. Roos, “Dark Matter: The evidence from astronomy, astrophysics and cosmology,” arXiv:1001.0316. http://arxiv.org/abs/1001.0316.

[21] LSND Collaboration, “Evidence for neutrino oscillations from the observation of $\bar{\nu}_e$ appearance in a $\bar{\nu}_\mu$ beam,” *Physical Review D* **64** no. 11, (Nov., 2001) 71, arXiv:0104049 [hep-ex]. http://link.aps.org/doi/10.1103/PhysRevD.64.112007.

[22] KamLAND Collaboration, “Measurement of neutrino oscillation with KamLAND : evidence of spectral distortion,” *Physical Review Letters* **94** no. 8, (Mar., 2005) 5, arXiv:0406035 [hep-ex]. http://link.aps.org/doi/10.1103/PhysRevLett.94.081801.

[23] K2K Collaboration, “Measurement of neutrino oscillation by the K2K experiment,” *Physical Review D* **74** no. 7, (Oct., 2006) 40, arXiv:0606032 [hep-ex]. http://link.aps.org/doi/10.1103/PhysRevD.74.072003.

[24] MINOS Collaboration, “Observation of muon neutrino disappearance with the MINOS detectors in the NuMI neutrino beam,” *Physical Review Letters* **97** no. 19, (Nov., 2006) 6, arXiv:0607088 [hep-ex]. http://link.aps.org/doi/10.1103/PhysRevLett.97.191801.

[25] MINOS Collaboration, “Measurement of neutrino oscillations with the MINOS detectors in the NuMI beam,” *Physical Review Letters* **101** no. 13, (Sept., 2008) 5,

arXiv:0806.2237.
http://link.aps.org/doi/10.1103/PhysRevLett.101.131802.

[26] S. P. Martin, "A Supersymmetry Primer," *Nature* no. December, (Sept., 1997) 128, arXiv:9709356 [hep-ph]. http://arxiv.org/abs/hep-ph/9709356.

[27] N. ArkaniHamed, S. Dimopoulos, and G. Dvali, "The hierarchy problem and new dimensions at a millimeter," *Physics Letters B* **429** no. 3-4, (June, 1998) 263–272, arXiv:9803315 [hep-ph].
http://linkinghub.elsevier.com/retrieve/pii/S0370269398004663.

[28] L. Randall and R. Sundrum, "Large Mass Hierarchy from a Small Extra Dimension," *Physical Review Letters* **83** no. 17, (Oct., 1999) 3370–3373, arXiv:9905221 [hep-ph]. http://link.aps.org/doi/10.1103/PhysRevLett.83.3370.

[29] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert, and J. Winter, "Event generation with SHERPA 1.1," *Journal of High Energy Physics* **2009** no. 02, (Feb., 2009) 007–007, arXiv:0811.4622.
http://stacks.iop.org/1126-6708/2009/i=02/a=007?key=crossref.de8c1359dc803c693a4176f8a4c92155.

[30] S. Drell and T. Yan, "Massive lepton-pair production in hadron-hadron collisions at high energies," *Physical Review Letters* **25** no. 5, (Aug., 1970) 316–320.
http://link.aps.org/doi/10.1103/PhysRevLett.25.316.

[31] S. Drell and T. Yan, "Partons and their applications at high energies," *Annals of Physics* **66** no. 2, (Aug., 1971) 578–623.
http://linkinghub.elsevier.com/retrieve/pii/0003491671900716.

[32] T. Lee and M. Nauenberg, "Degenerate Systems and Mass Singularities," *Physical Review* **133** no. 6B, (Mar., 1964) B1549–B1562.
http://link.aps.org/doi/10.1103/PhysRev.133.B1549.

[33] T. Kinoshita, "Mass Singularities of Feynman Amplitudes," *Journal of Mathematical Physics* **3** no. 4, (1962) 650.
http://link.aip.org/link/JMAPAQ/v3/i4/p650/s1&Agg=doi.

[34] J. C. Collins, D. E. Soper, and G. Sterman, "Factorization of Hard Processes in QCD," *Annual Review of Nuclear and Particle Science* **37** no. 1, (Sept., 2004) 100, arXiv:0409313 [hep-ph]. http://arxiv.org/abs/hep-ph/0409313.

[35] Y. L. Dokshitzer, "Calculation of the structure functions for deep inelastic scattering and e+ e- annihilation by perturbation theory in quantum chromodynamics," *Sov. Phys. JETP* **46** (1977) 641–653.

[36] V. N. Gribov and L. N. Lipatov, "Deep inelastic e p scattering in perturbation theory," *Sov. J. Nucl. Phys.* **15** (1972) 438–450.

[37] G. Altarelli, "Asymptotic freedom in parton language," *Nuclear Physics B* **126** no. 2, (Aug., 1977) 298–318.
http://linkinghub.elsevier.com/retrieve/pii/0550321377903844.

[38] J. Conrad, M. Shaevitz, and T. Bolton, "Precision measurements with high-energy neutrino beams," *Reviews of Modern Physics* **70** no. 4, (Oct., 1998) 1341–1392. http://link.aps.org/doi/10.1103/RevModPhys.70.1341.

[39] H1 Collaboration, "Measurement of neutral and charged current cross-sections in positron-proton collisions at large momentum transfer," *The European Physical Journal C* **13** no. 4, (Aug., 2000) 609, arXiv:9908059 [hep-ex]. http://link.springer.de/link/service/journals/10052/bibs/0013004/00130609.htm.

[40] ZEUS Collaboration, "Measurement of the neutral current cross section and $\mathbf{F_2}$ structure function for deep inelastic $\mathbf{e^+p}$ scattering at HERA," *The European Physical Journal C* **21** no. 3, (July, 2001) 443–471, arXiv:0105090 [hep-ex]. http://www.springerlink.com/index/10.1007/s100520100749.

[41] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, "Parton distributions for the LHC," *The European Physical Journal C* **63** no. 2, (July, 2009) 189–285, arXiv:0901.0002. http://www.springerlink.com/index/10.1140/epjc/s10052-009-1072-5.

[42] J. Pumplin, J. Huston, H.-L. Lai, P. Nadolsky, W.-K. Tung, and C.-P. Yuan, "Collider inclusive jet data and the gluon distribution," *Physical Review D* **80** no. 1, (July, 2009) 1–16. http://link.aps.org/doi/10.1103/PhysRevD.80.014019.

[43] CMS Collaboration, "Measurement of muon charge asymmetry in inclusive W production in pp collisions at $\sqrt{s}$ = 7 TeV," *CMS PAS EWK-11-005* (2011) .

[44] MSTW Collaboration, "Web page." http://projects.hepforge.org/mstwpdf/.

[45] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, "NNLO global parton analysis," *Physics Letters B* **531** no. 3-4, (Jan., 2002) 18, arXiv:0201127 [hep-ph]. http://arxiv.org/abs/hep-ph/0201127.

[46] CTEQ Collaboration, "Web page." http://hep.pa.msu.edu/cteq/public/cteq6.html.

[47] J. Pumplin, D. R. Stump, J. Huston, H.-L. Lai, P. Nadolsky, and W.-K. Tung, "New Generation of Parton Distributions with Uncertainties from Global QCD Analysis," *Journal of High Energy Physics* **2002** no. 07, (July, 2002) 012–012, arXiv:0201195 [hep-ph]. http://stacks.iop.org/1126-6708/2002/i=07/a=012?key=crossref.a5a0a565647d01b0ed9d71334bc7433f.

[48] P. Nadolsky, H.-L. Lai, Q.-H. Cao, J. Huston, J. Pumplin, D. R. Stump, W.-K. Tung, and C.-P. Yuan, "Implications of CTEQ global analysis for collider observables," *Physical Review D* **78** no. 1, (July, 2008) 32, arXiv:0802.0007. http://link.aps.org/doi/10.1103/PhysRevD.78.013004.

[49] G. Sterman, J. Smith, J. C. Collins, J. Whitmore, R. Brock, J. Huston, J. Pumplin, W.-K. Tung, H. Weerts, C.-P. Yuan, S. Kuhlmann, S. Mishra, J. Morfín, F. Olness, J. Owens, J. Qiu, and D. E. Soper, "Handbook of perturbative QCD," *Reviews of Modern Physics* **67** no. 1, (Jan., 1995) 157–248. http://link.aps.org/doi/10.1103/RevModPhys.67.157.

[50] R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, A. Piccione, J. Rojo, and M. Ubiali, "A determination of parton distributions with faithful uncertainty estimation," *Nuclear Physics B* **809** no. 1-2, (Aug., 2008) 73, arXiv:0808.1231. http://arxiv.org/abs/0808.1231.

[51] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, "Uncertainties of predictions from parton distributions I: Experimental errors," *The European Physical Journal C* **28** no. 4, (June, 2003) 455–473, arXiv:0211080 [hep-ph]. http://www.springerlink.com/Index/10.1140/epjc/s2003-01196-2.

[52] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, "Uncertainties of predictions from parton distributions II: theoretical errors," *The European Physical Journal C* **35** no. 3, (June, 2004) 325–348, arXiv:0308087 [hep-ph]. http://www.springerlink.com/index/10.1140/epjc/s2004-01825-2.

[53] J. Pumplin, D. R. Stump, R. Brock, D. Casey, J. Huston, J. Kalk, H.-L. Lai, and W.-K. Tung, "Uncertainties of predictions from parton distribution functions. II. The Hessian method," *Physical Review D* **65** no. 1, (Dec., 2001) 30, arXiv:0101032 [hep-ph]. http://link.aps.org/doi/10.1103/PhysRevD.65.014013.

[54] V. V. Sudakov, "No Title," *Sov. Phys. J.E.T.P.* **30** no. 65, (1956) .

[55] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, "Parton fragmentation and string dynamics," *Physics Reports* **97** no. 2-3, (July, 1983) 31–145. http://linkinghub.elsevier.com/retrieve/pii/0370157383900807.

[56] B. R. Webber, "A QCD model for jet fragmentation including soft gluon interference," *Nuclear Physics B* **238** no. 3, (June, 1984) 492–528. http://linkinghub.elsevier.com/retrieve/pii/055032138490333X.

[57] C. Peterson, D. Schlatter, I. Schmitt, and P. Zerwas, "Scaling violations in inclusive e+e- annihilation spectra," *Physical Review D* **27** no. 1, (Jan., 1983) 105–111. http://link.aps.org/doi/10.1103/PhysRevD.27.105.

[58] A. Mueller and P. Nason, "Heavy particle content in QCD jets," *Physics Letters B* **157** no. 2-3, (July, 1985) 226–228. http://linkinghub.elsevier.com/retrieve/pii/0370269385915515.

[59] A. Mueller and P. Nason, "Heavy particle content in QCD jets," *Nuclear Physics B* **266** no. 2, (Mar., 1986) 265–273. http://linkinghub.elsevier.com/retrieve/pii/055032138690091X.

[60] M. L. Mangano and P. Nason, "Heavy quark multiplicities in gluon jets," *Physics Letters B* **285** no. 1-2, (July, 1992) 160–166. http://linkinghub.elsevier.com/retrieve/pii/0370269392913162.

[61] T. Sjöstrand and M. van Zijl, "A multiple-interaction model for the event structure in hadron collisions," *Physical Review D* **36** no. 7, (Oct., 1987) 2019–2041. http://link.aps.org/doi/10.1103/PhysRevD.36.2019.

[62] T. Sjöstrand, "Monte Carlo Generators," arXiv:0611247 [hep-ph]. http://arxiv.org/abs/hep-ph/0611247.

[63] T. Sjöstrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual," *Journal of High Energy Physics* **2006** no. 05, (May, 2006) 026–026, arXiv:0603175 [hep-ph]. http://stacks.iop.org/1126-6708/2006/i=05/a=026?key=crossref.7fbc8fa1a47a48f7565bead655446685.

[64] S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert, and J. Winter, "Homepage of the SHERPA event generator." http://www.sherpa-mc.de/.

[65] F. Krauss, R. Kuhn, and G. Soff, "AMEGIC++ 1.0, A Matrix Element Generator In C++," *Journal of High Energy Physics* **2002** no. 02, (Feb., 2002) 044–044, arXiv:0109036 [hep-ph]. http://stacks.iop.org/1126-6708/2002/i=02/a=044?key=crossref.c2c5e6fc7179c78efd1a81b4cae6ee01.

[66] M. L. Mangano, F. Piccinini, A. D. Polosa, M. Moretti, and R. Pittau, "ALPGEN, a generator for hard multiparton processes in hadronic collisions," *Journal of High Energy Physics* **2003** no. 07, (July, 2003) 001–001. http://stacks.iop.org/1126-6708/2003/i=07/a=001?key=crossref.8d9456967c3fd10e94e494eb55c2e4d2.

[67] J. Alwall, P. Demin, S. De Visscher, R. Frederix, M. Herquet, F. Maltoni, T. Plehn, D. L. Rainwater, and T. Stelzer, "MadGraph/MadEvent v4: the new web generation," *Journal of High Energy Physics* **2007** no. 09, (Sept., 2007) 028–028, arXiv:0706.2334. http://stacks.iop.org/1126-6708/2007/i=09/a=028?key=crossref.117225e2f8529215c57d62b23454782e.

[68] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, "Matching matrix elements and shower evolution for top-quark production in hadronic collisions," *Journal of High Energy Physics* **2007** no. 01, (Nov., 2006) 22, arXiv:0611129 [hep-ph]. http://arxiv.org/abs/hep-ph/0611129.

[69] S. Catani, F. Krauss, B. R. Webber, and R. Kuhn, "QCD Matrix Elements + Parton Showers," *Journal of High Energy Physics* **2001** no. 11, (Nov., 2001) 063–063, arXiv:0109231 [hep-ph]. http://stacks.iop.org/1126-6708/2001/i=11/a=063?key=crossref.759f85ac1d89d6d197aad08310cc8e64.

[70] L. Evans and P. Bryant (editors), "LHC Machine," *Journal of Instrumentation* **3** no. 08, (Aug., 2008) S08001–S08001. http://stacks.iop.org/1748-0221/3/i=08/a=S08001?key=crossref.c8f0bc7a1a7c3a99bae6dd73f234f91b.

[71] P. Lefèvre and T. Petterson, "The Large Hadron Collider : conceptual design," *CERN/AC/95-05(LHC)* (1995) .

[72] CMS Collaboration, "The CMS experiment at the CERN LHC," *Journal of Instrumentation* **3** no. 08, (Aug., 2008) S08004–S08004. http://stacks.iop.org/1748-0221/3/i=08/a=S08004?key=crossref.fc7f4422618a075830e017d1d930a35c.

[73] ATLAS Collaboration, "The ATLAS experiment at the CERN Large Hadron Collider," *Journal of Instrumentation* **3** no. 08, (Aug., 2008) S08003–S08003. http://stacks.iop.org/1748-0221/3/i=08/a=S08003?key=crossref.b2ac868e8992413771c34191a8138368.

[74] LHCb Collaboration, “The LHCb detector at the LHC,” *Journal of Instrumentation* **3** no. 08, (Aug., 2008) S08005–S08005. http://stacks.iop.org/1748-0221/3/i=08/a=S08005?key=crossref.358ac80e1a6b6ba36f68c89dc0c4bed4.

[75] ALICE Collaboration, “The ALICE experiment at the CERN LHC,” *Journal of Instrumentation* **3** no. 08, (Aug., 2008) S08002–S08002. http://stacks.iop.org/1748-0221/3/i=08/a=S08002?key=crossref.4a430fa328e181a89b6c10b850640204.

[76] TOTEM Collaboration, “The TOTEM experiment at the CERN Large Hadron Collider,” *Journal of Instrumentation* **3** no. 08, (Aug., 2008) S08007–S08007. http://stacks.iop.org/1748-0221/3/i=08/a=S08007?key=crossref.53593e597df8991a1dbab233c46bb618.

[77] LHCf Collaboration, “The LHCf detector at the CERN Large Hadron Collider,” *Journal of Instrumentation* **3** no. 08, (Aug., 2008) S08006–S08006. http://stacks.iop.org/1748-0221/3/i=08/a=S08006?key=crossref.c0bd9c5d1d22c73b3abfe7ebf37a08f4.

[78] CMS Collaboration, “CMS, Tracker Technical Design Report,” *CERN/LHCC 98-6* (1998) .

[79] CMS Collaboration, “Addendum to the CMS Tracker TDR,” *CERN/LHCC 2000-016* (2000) .

[80] CMS (ECAL) Collaboration, “Studies of the CMS electromagnetic calorimeter performance in the electron test beam,” *Journal of Physics: Conference Series* **160** (Apr., 2009) 012048. http://stacks.iop.org/1742-6596/160/i=1/a=012048?key=crossref.748a37f8b1d24d9155570bf289a61fe6.

[81] CMS (ECAL) Collaboration, “Energy resolution of the barrel of the CMS electromagnetic calorimeter,” *Journal of Instrumentation* **2** no. 04, (Apr., 2007) P04004–P04004. http://stacks.iop.org/1748-0221/2/i=04/a=P04004?key=crossref.74779ff3599f7ff03082321c42aa0801.

[82] CMS Collaboration, “Electromagnetic calorimeter calibration with 7 TeV data,” *CMS PAS EGM-10-003* (2010) 20.

[83] CMS (HCAL) Collaboration, “Design, performance, and calibration of CMS hadron-barrel calorimeter wedges,” *The European Physical Journal C* **55** no. 1, (Apr., 2008) 159–171. http://www.springerlink.com/index/10.1140/epjc/s10052-008-0573-y.

[84] CMS Collaboration, “Muon High-level trigger in CMS,” *CMS Analysis Note AN-2010/234* (2010) 1–89.

[85] R. Frühwirth, “Application of Kalman filtering to track and vertex fitting,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **262** no. 2-3, (Dec., 1987) 444–450. http://linkinghub.elsevier.com/retrieve/pii/0168900287908874.

[86] CMS Collaboration, "CMS Physics Technical Design Report, Volume I: Detector Performance and Software," *CERN/LHCC 2006-001* (June, 2006) 547. http://stacks.iop.org/0954-3899/34/i=6/a=S01?key=crossref.af04fed7d570d4c3c6901a202af586fc.

[87] T. Speer, W. Adam, R. Frühwirth, A. Strandlie, T. Todorov, and M. Winkler, "Track reconstruction in the CMS tracker," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **559** no. 1, (Apr., 2006) 143–147. http://linkinghub.elsevier.com/retrieve/pii/S0168900205022576.

[88] CMS Collaboration, "Beam position determination using tracks," *CMS Note 2007/021* (2007) .

[89] S. Cucciarelli, M. Konecki, D. Kotlinski, and T. Todorov, "Track parameter evaluation and primary vertex finding with the pixel detector," *CMS Note 2003/026* (2003) .

[90] W. Waltenberger, R. Frühwirth, and P. Vanlaer, "Adaptive vertex fitting," *Journal of Physics G: Nuclear and Particle Physics* **34** no. 12, (Dec., 2007) N343–N356. http://stacks.iop.org/0954-3899/34/i=12/a=N01?key=crossref.1572e0c569f8431ab1d721e989956caa.

[91] R. Frühwirth, K. Prokofiev, T. Speer, P. Vanlaer, and W. Waltenberger, "Vertex fitting in the CMS tracker," *CMS Note 2006/032* (2006) .

[92] CMS Collaboration, "CMS tracking performance results from early LHC operation," *The European Physical Journal C* **70** no. 4, (Nov., 2010) 1165–1192, arXiv:1007.1988. http://www.springerlink.com/index/10.1140/epjc/s10052-010-1491-3.

[93] CMS Collaboration, "Measurement of tracking efficiency," *CMS PAS TRK-10-002* (2010) .

[94] CMS Collaboration, "Measurement of momentum scale and resolution using low-mass resonances and cosmic ray muons," *CMS PAS TRK-10-004* (2010) .

[95] CMS Collaboration, "Muon reconstruction in the CMS detector," *CMS Analysis Note AN-2008/097* (2009) 76.

[96] S. Baffioni, C. Charlot, F. Ferri, D. Futyan, P. Meridiani, I. Puljak, C. Rovelli, R. Salerno, and Y. Sirois, "Electron reconstruction in CMS," *The European Physical Journal C* **49** no. 4, (Jan., 2007) 1099–1116. http://www.springerlink.com/index/10.1140/epjc/s10052-006-0175-5.

[97] W. Adam, R. Frühwirth, A. Strandlie, and T. Todorov, "Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC," *Journal of Physics G: Nuclear and Particle Physics* **31** no. 9, (Sept., 2005) N9–N20. http://stacks.iop.org/0954-3899/31/i=9/a=N01?key=crossref.64c1ba0887a29aebb3c0197cd1b051c6.

[98] CMS Collaboration, "Electron reconstruction and identification at 7 TeV," *CMS PAS EGM-10-004* (2010) 19.

[99] CMS Collaboration, "Performance of muon reconstruction and identification in pp collisions at $\sqrt{s}$ = 7 TeV," *CMS PAS MUO-10-004* (2011) .

[100] CMS Collaboration, "Jet performance in pp collisions at 7 TeV," *CMS PAS JME-10-003* (2010) 32.

[101] CMS Collaboration, "Jet Plus Tracks algorithm for calorimeter jet energy corrections in CMS," *CMS PAS JME-09-002* (2009) 16.

[102] CMS Collaboration, "Particle-flow event reconstruction in CMS and performance for jets, taus and missing transverse energy," *CMS PAS PFT-09-001* (2009) 25.

[103] G. C. Blazey, J. R. Dittmann, S. D. Ellis, V. D. Elvira, K. Frame, S. Grinstein, R. Hirosky, R. Piegaia, H. Schellman, R. Snihur, V. Sorin, and D. Zeppenfeld, "Run II Jet Physics," *Proceedings of the Run II QCD and Weak Boson Physics Workshop* (May, 2000) 32, `arXiv:0005012 [hep-ex]`. `http://arxiv.org/abs/hep-ex/0005012`.

[104] G. Sterman and S. Weinberg, "Jets from Quantum Chromodynamics," *Physical Review Letters* **39** no. 23, (Dec., 1977) 1436–1439. `http://link.aps.org/doi/10.1103/PhysRevLett.39.1436`.

[105] S. D. Ellis and D. E. Soper, "Successive combination jet algorithm for hadron collisions," *Physical Review D* **48** no. 7, (May, 1993) 18, `arXiv:9305266 [hep-ph]`. `http://arxiv.org/abs/hep-ph/9305266`.

[106] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, "Longitudinally-invariant $k_\perp$-clustering algorithms for hadron-hadron collisions," *Nuclear Physics B* **406** no. 1-2, (Sept., 1993) 187–224. `http://linkinghub.elsevier.com/retrieve/pii/055032139390166M`.

[107] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," *Journal of High Energy Physics* **2008** no. 04, (Apr., 2008) 063–063. `http://stacks.iop.org/1126-6708/2008/i=04/a=063?key=crossref.2ffef838d94c2b899ee27950c18ee8d2`.

[108] Particle Data Group, "Review of Particle Physics," *Journal of Physics G: Nuclear and Particle Physics* **37** no. 7A, (July, 2010) 075021. `http://stacks.iop.org/0954-3899/37/i=7A/a=075021?key=crossref.de0390bfb70101fa23c6dbe588ea1324`.

[109] CMS Collaboration, "Algorithms for b jet identification in CMS," *CMS PAS BTV-09-001* (2009) 16.

[110] CMS Collaboration, "Plan for jet energy corrections at CMS," *CMS PAS JME-07-002* (2008) 14.

[111] CDF Collaboration, "Determination of the jet energy scale at the Collider Detector at Fermilab," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **566** no. 2, (Oct., 2005) 375–412, `arXiv:0510047 [hep-ex]`. `http://arxiv.org/abs/hep-ex/0510047`.

[112] D0 Collaboration, "Determination of the absolute jet energy scale in the DØcalorimeters," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **424** no. 2-3, (Mar., 1999) 352–394.
http://linkinghub.elsevier.com/retrieve/pii/S0168900298013680.

[113] CMS Collaboration, "Measurement of jet energy scale corrections using top quark events," *CMS PAS TOP-07-004* (2008) 10.

[114] CMS Collaboration, "Jet corrections to parent parton energy," *CMS PAS JME-08-002* (2009) 10.

[115] CMS Collaboration, "Determination of the jet energy scale in CMS with pp collisions at 7 TeV," *CMS PAS JME-10-010* (2010) 21.

[116] CMS Collaboration, "Performance of missing transverse energy using calorimeter and tracks in CMS," *CMS PAS JME-09-006* (2009) 8.

[117] CMS Collaboration, "Performance of track-corrected missing transverse energy in CMS," *CMS PAS JME-09-010* (2009) 7.

[118] CDF Collaboration, "Evidence for top quark production in $p\bar{p}$ collisions at $\sqrt{s}$ =1.8 TeV," *Physical Review D* **50** no. 5, (Sept., 1994) 2966–3026.
http://link.aps.org/doi/10.1103/PhysRevD.50.2966.

[119] DO Collaboration, "Observation of the top quark," *Physical Review Letters* **74** no. 14, (Apr., 1995) 2632–2637.
http://link.aps.org/doi/10.1103/PhysRevLett.74.2632.

[120] K. Lane, "An introduction to technicolor," arXiv:9401324 [hep-ph].
http://arxiv.org/abs/hep-ph/9401324.

[121] C. Hill, "Topcolor assisted technicolor," *Physics Letters B* **345** no. 4, (Feb., 1995) 483–489, arXiv:9411426 [hep-ph].
http://linkinghub.elsevier.com/retrieve/pii/0370269394016605.

[122] K. Lane, "A new model of topcolor-assisted technicolor," *Physics Letters B* **433** no. 1-2, (Aug., 1998) 96–101, arXiv:9805254 [hep-ph].
http://linkinghub.elsevier.com/retrieve/pii/S0370269398007084.

[123] T. G. Rizzo, "Z' Phenomenology and the LHC," arXiv:0610104 [hep-ph].
http://arxiv.org/abs/hep-ph/0610104.

[124] CDF Collaboration, "Combination of CDF top quark mass measurements," *CDF Note 10444* **63** no. Winter, (2011) .

[125] D0 Collaboration, "Precise measurement of the top quark mass from lepton+jets events at D0," tech. rep., May, 2011. arXiv:1105.6287.
http://arxiv.org/abs/1105.6287.

[126] D0 Collaboration, "Precise measurement of the top quark mass in the dilepton channel at D0," tech. rep., May, 2011. arXiv:1105.0320.
http://arxiv.org/abs/1105.0320.

[127] CDF Collaboration, "CDF Top Quark Physics Public Results." http://www-cdf.fnal.gov/physics/new/top/top.html.

[128] D0 Collaboration, "DØ's Top Quark Physics Results." http://www-d0.fnal.gov/Run2Physics/top/top_public_web_pages/top_public.html.

[129] ATLAS Collaboration, "Top quark pair production cross-section measurement in ATLAS in the single lepton+jets channel without b-tagging," *ATLAS-CONF-2011-023* (2011) .

[130] CMS Collaboration, "Measurement of the top quark mass in the lepton+jets channel," *CMS PAS TOP-10-009* **1** (2011) .

[131] CMS Collaboration, "First measurement of the top quark mass in pp collisions at $\sqrt{s}$ = 7 TeV," *CMS PAS TOP-10-006* (2011) .

[132] CDF Collaboration, "Measurement of B(tWb)/B(tWq) at the Collider Detector at Fermilab," *Physical Review Letters* **95** no. 10, (Aug., 2005) 7, arXiv:0505091 [hep-ex]. http://link.aps.org/doi/10.1103/PhysRevLett.95.102002.

[133] D0 Collaboration, "Measurement of B(tWb)/B(tWq) at $\sqrt{s}$ = 1.96 TeV," *Physics Letters B* **639** no. 6, (Aug., 2006) 616–622, arXiv:0603002 [hep-ex]. http://linkinghub.elsevier.com/retrieve/pii/S0370269306008872.

[134] J. Alwall, R. Frederix, J.-M. Gérard, A. Giammanco, M. Herquet, S. Kalinin, E. Kou, V. Lemaitre, and F. Maltoni, "Is Vtb1?," *The European Physical Journal C* **49** no. 3, (Dec., 2006) 791–801, arXiv:0607115v2 [arXiv:hep-ph]. http://www.springerlink.com/index/10.1140/epjc/s10052-006-0137-y.

[135] Tevatron Electroweak Working Group (D0 and CDF Collaborations), "Combination of CDF and D0 Measurements of the Single Top Production Cross Section," arXiv:0908.2171. http://arxiv.org/abs/0908.2171.

[136] M. Jeabek and J. Kühn, "QCD corrections to semileptonic decays of heavy quarks," *Nuclear Physics B* **314** no. 1, (Feb., 1989) 1–6. http://linkinghub.elsevier.com/retrieve/pii/0550321389901089.

[137] M. Jeabek and J. Kühn, "Top quark width: Theoretical update," *Physical Review D* **48** no. 5, (Sept., 1993) R1910–R1913. http://link.aps.org/doi/10.1103/PhysRevD.48.R1910.

[138] CDF Collaboration, "A measurement of top quark width using template method in lepton+jets channel with $4.3fb^1$," *CDF Note 10035* no. CI, (2010) 12.

[139] I. Bigi, Y. L. Dokshitzer, V. Khoze, J. Kühn, and P. Zerwas, "Production and decay properties of ultra-heavy quarks," *Physics Letters B* **181** no. 1-2, (Nov., 1986) 157–163. http://linkinghub.elsevier.com/retrieve/pii/037026938691275X.

[140] The Durham HepData Project, "Online PDF plotting and calculation." http://hepdata.cedar.ac.uk/pdf/pdf3.html.

[141] N. Kidonakis, "Higher-order corrections to top-antitop pair and single top quark production," arXiv:0909.0037. http://arxiv.org/abs/0909.0037.

[142] CDF Collaboration, "Combination of CDF top quark pair production cross-section measurements with up to $4.6\ fb^{-1}$.," *CDF Note 9913* (2009) 15.

[143] D0 Collaboration, "Measurement of the $t\bar{t}$ production cross-section using dilepton events in $p\bar{p}$ collisions," arXiv:1105.5384 [hep-ex]. http://arxiv.org/abs/1105.5384.

[144] CMS Collaboration, "First measurement of the cross section for top quark pair production in protonproton collisions at $\sqrt{s}$=7 TeV," *Physics Letters B* **695** no. 5, (Jan., 2011) 424–443, arXiv:1010.5994. http://linkinghub.elsevier.com/retrieve/pii/S037026931001333X.

[145] ATLAS Collaboration, "Measurement of the top quark-pair production cross section with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV," *The European Physical Journal C* **71** no. 3, (Mar., 2011) . http://www.springerlink.com/index/10.1140/epjc/s10052-011-1577-6.

[146] CMS Collaboration, "Combination of top pair production cross-section in pp collisions at $\sqrt{s}$ = 7 TeV and comparisons with theory," *CMS PAS TOP-11-001* (2011) .

[147] ATLAS Collaboration, "A combined measurement of the top quark pair production cross-section using dilepton and single-lepton final states," *ATLAS-CONF-2011-040* (2011) .

[148] CMS Collaboration, "CMS Top Physics Results." https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsTOP.

[149] M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer, and M. Wiedermann, "HATHOR HAdronic Top and Heavy quarks crOss section calculatoR," *Computer Physics Communications* **182** no. 4, (Apr., 2011) 1034–1046, arXiv:1007.1327. http://linkinghub.elsevier.com/retrieve/pii/S0010465510005333.

[150] B. Lillie, L. Randall, and L.-T. Wang, "The Bulk RS KK-gluon at the LHC," *Journal of High Energy Physics* **2007** no. 09, (Sept., 2007) 074–074. http://stacks.iop.org/1126-6708/2007/i=09/a=074?key=crossref.a08d4647737fe18833667f9ee21886db.

[151] R. Frederix and F. Maltoni, "Top pair invariant mass distribution: a window on new physics," *Journal of High Energy Physics* **2009** no. 01, (Jan., 2009) 047–047, arXiv:0712.2355. http://stacks.iop.org/1126-6708/2009/i=01/a=047?key=crossref.fd74d36953a80e4030601e9ffaa5b21d.

[152] CMS Collaboration, "Search for resonances in semi-leptonic top pair decays close to production threshold," *CMS PAS TOP-10-007* (2011) .

[153] J. M. Campbell, R. K. Ellis, and W. Ciaran, "MCFM - Monte Carlo for FeMtobarn processes." http://mcfm.fnal.gov/.

[154] J. M. Campbell and R. K. Ellis, "MCFM for the Tevatron and the LHC," *Nuclear Physics B - Proceedings Supplements* **205-206** (Aug., 2010) 10–15. http://linkinghub.elsevier.com/retrieve/pii/S0920563210001945.

[155] D0 Collaboration, "Observation of single top quark production," *Physical Review Letters* **103** no. 9, (Aug., 2009) 092001. http://link.aps.org/doi/10.1103/PhysRevLett.103.092001.

[156] CDF Collaboration, "Observation of electroweak single top quark production," *Physical Review Letters* **103** no. 9, (Aug., 2009) , arXiv:0903.0885. http://link.aps.org/doi/10.1103/PhysRevLett.103.092002.

[157] CDF Collaboration, "Observation of single top quark production and measurement of $|V_{tb}|$ with CDF," *Physical Review D* **82** no. 11, (Dec., 2010) , arXiv:1004.1181. http://link.aps.org/doi/10.1103/PhysRevD.82.112005.

[158] D0 Collaboration, "Model-independent measurement of t-channel single top quark production in $p\bar{p}$ collisions at $\sqrt{s}$ =1.96 TeV," arXiv:1105.2788. http://arxiv.org/abs/1105.2788.

[159] ATLAS Collaboration, "Searches for single top quark production with the ATLAS detector in pp collisions at $\sqrt{s}$ = 7 TeV," *ATLAS-CONF-2011-027* (2011) .

[160] CMS Collaboration, "Measurement of the t-channel single top quark production cross section in pp collisions at $\sqrt{s}$ = 7 TeV," *CMS PAS TOP-10-008* (June, 2011) , arXiv:1106.3052. http://arxiv.org/abs/1106.3052.

[161] CDF Collaboration, "Measurement of the cross-section for W boson production in association with jets in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96 TeV," *Physical Review D* **77** no. 1, (Jan., 2008) 9–12, arXiv:0711.4044. http://link.aps.org/doi/10.1103/PhysRevD.77.011108.

[162] C. F. Berger, Z. Bern, L. J. Dixon, F. Febres Cordero, D. Forde, T. Gleisberg, H. Ita, D. A. Kosower, and D. Maître, "Next-to-leading order QCD predictions for W+3-jet distributions at hadron colliders," *Physical Review D* **80** no. 7, (Oct., 2009) 1–31. http://link.aps.org/doi/10.1103/PhysRevD.80.074036.

[163] D0 Collaboration, "Measurements of inclusive W+jets production rates as a function of jet transverse momentum in ppbar collisions at $\sqrt{s}$ = 1.96 TeV," arXiv:1106.1457. http://arxiv.org/abs/1106.1457.

[164] CDF Collaboration, "Measurement of $Z/\gamma \rightarrow \mu^+\mu^-$+ jets production cross-section," *CDF Note 10216* (2010) . http://www.ncbi.nlm.nih.gov/pubmed/20366413.

[165] D0 Collaboration, "Measurement of differential Z/γ+jet+X cross sections in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96 TeV," *Physics Letters B* **669** no. 5, (Nov., 2008) 278–286. http://linkinghub.elsevier.com/retrieve/pii/S0370269308012148.

[166] ATLAS Collaboration, "Measurement of the production cross-section for W bosons in association with jets in pp collisions using $33 \sim pb^{-1}$ of data at $\sqrt{s}$ = 7 TeV with the ATLAS detector," *ATLAS-CONF-2011-060* (2011) .

[167] ATLAS Collaboration, "Measurement of the production cross-section for Z/γ^* in association with jets in pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector," *ATLAS-CONF-2011-042* (2011) .

[168] CMS Collaboration, "Rates of jets produced in association with W and Z bosons in pp collisions at $\sqrt{s}$=7TeV," *CMS PAS EWK-10-012* (2011) .

[169] C. F. Berger, Z. Bern, L. J. Dixon, F. Febres Cordero, D. Forde, T. Gleisberg, H. Ita, D. A. Kosower, and D. Maître, "Next-to-leading order QCD predictions for Z, γ^*+3-jet distributions at the Tevatron," *Physical Review D* **82** no. 7, (Oct., 2010) 50, arXiv:1004.1659. http://link.aps.org/doi/10.1103/PhysRevD.82.074002.

[170] C. F. Berger, Z. Bern, L. J. Dixon, F. Febres Cordero, D. Forde, T. Gleisberg, H. Ita, D. A. Kosower, and D. Maître, "Precise Predictions for W+3 Jet Production at Hadron Colliders," *Physical Review Letters* **102** no. 22, (June, 2009) 5, arXiv:0902.2760. http://link.aps.org/doi/10.1103/PhysRevLett.102.222001.

[171] R. K. Ellis, K. Melnikov, and G. Zanderighi, "W+3 jet production at the Tevatron," *Physical Review D* **80** no. 9, (Nov., 2009) 1–8. http://link.aps.org/doi/10.1103/PhysRevD.80.094002.

[172] C. F. Berger, Z. Bern, L. J. Dixon, F. Febres Cordero, D. Forde, T. Gleisberg, H. Ita, D. A. Kosower, and D. Maître, "Precise Predictions for W+4-Jet Production at the Large Hadron Collider," *Physical Review Letters* **106** no. 9, (Mar., 2011) 5, arXiv:1009.2338. http://link.aps.org/doi/10.1103/PhysRevLett.106.092001.

[173] C. F. Berger, Z. Bern, L. J. Dixon, F. Febres Cordero, D. Forde, T. Gleisberg, H. Ita, D. A. Kosower, and D. Maître, "Vector Boson + Jets with BlackHat and SHERPA," *Nuclear Physics B - Proceedings Supplements* **205-206** (Aug., 2010) 92–97. http://linkinghub.elsevier.com/retrieve/pii/S0920563210002082.

[174] K. Melnikov and G. Zanderighi, "W+3 jet production at the LHC as a signal or background," *Physical Review D* **81** no. 7, (Apr., 2010) 1–10. http://link.aps.org/doi/10.1103/PhysRevD.81.074025.

[175] S. Brensing, S. Dittmaier, M. Krämer, and A. Mück, "Radiative corrections to W-boson hadroproduction: Higher-order electroweak and supersymmetric effects," *Physical Review D* **77** no. 7, (Apr., 2008) 1–20. http://link.aps.org/doi/10.1103/PhysRevD.77.073006.

[176] A. Denner, S. Dittmaier, T. Kasprzik, and A. Mück, "Electroweak corrections to W + jet hadroproduction including leptonic W-boson decays," *Journal of High Energy Physics* **2009** no. 08, (Aug., 2009) 075–075, arXiv:0906.1656. http://stacks.iop.org/1126-6708/2009/i=08/a=075?key=crossref.db2f7b60f3b8a7b82087f2e81029aeca.

[177] CMS Collaboration, "Measurement of associated charm production in W final states at $\sqrt{s}$ = 7 TeV," *CMS PAS EWK-11-013* (2011) .

[178] F. Febres Cordero, L. Reina, and D. Wackeroth, "W- and Z-boson production with a massive bottom-quark pair at the Large Hadron Collider," *Physical Review D* **80** no. 3, (Aug., 2009) 1–13. http://link.aps.org/doi/10.1103/PhysRevD.80.034015.

[179] CDF Collaboration, "First measurement of the b-jet cross-section in events with a W boson in $p\bar{p}$ collisions at $\sqrt{s}$=1.96 TeV," *Physical Review Letters* **104** no. 13, (Apr., 2010) . http://link.aps.org/doi/10.1103/PhysRevLett.104.131801.

[180] CMS Collaboration, “Observation of $Z + b$, $Z \rightarrow ee, \mu\mu$ with CMS at $\sqrt{s}$ = 7 TeV,” *CMS PAS EWK-10-015* (2011) .

[181] J. M. Campbell and R. K. Ellis, “Update on vector boson pair production at hadron colliders,” *Physical Review D* **60** no. 11, (Nov., 1999) 1–9. http://link.aps.org/doi/10.1103/PhysRevD.60.113006.

[182] T. Binoth, M. Ciccolini, N. Kauer, and M. Krämer, “Gluon-induced W -boson pair production at the LHC,” *Journal of High Energy Physics* **2006** no. 12, (Dec., 2006) 046–046. http://stacks.iop.org/1126-6708/2006/i=12/a=046?key=crossref.a46e8ca9528854c05bebfabf2fdc3aac.

[183] CDF Collaboration, “Measurement of the WW+WZ Production Cross Section Using the lepton+jets Final State at CDF II,” *Physical Review Letters* **104** no. 10, (Mar., 2010) 101801. http://link.aps.org/doi/10.1103/PhysRevLett.104.101801.

[184] CMS Collaboration, “Measurement of the WW, WZ and ZZ cross-sections at CMS,” *CMS PAS EWK-11-010* (2011) .

[185] ATLAS Collaboration, “Measurement of the WW cross-section in $\sqrt{s}$ = 7 TeV pp collisions with ATLAS,” arXiv:1104.5225. http://arxiv.org/abs/1104.5225.

[186] D0 Collaboration, “Measurement of ratios of multi-jet cross-sections $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96 TeV,” *DO Note 6032-CONF* (2010) 1–5.

[187] CMS Collaboration, “Measurement of the Ratio of the 3-jet to 2-jet Cross Sections in pp Collisions at sqrt(s) = 7 TeV,” *CMS PAS QCD-10-012* (June, 2011) , arXiv:1106.0647. http://arxiv.org/abs/1106.0647.

[188] ATLAS Collaboration, “Measurement of multi-jet cross-sections in proton-proton collisions at 7 TeV center-of-mass energy,” *ATLAS-CONF-2011-043* (2011) .

[189] R. D. Field, “Early LHC underlying event data - Findings and surprises,” arXiv:1010.3558. http://arxiv.org/abs/1010.3558.

[190] R. Gavin, Y. Li, F. Petriello, and S. Quackenbush, “FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order,” arXiv:1011.3540. http://arxiv.org/abs/1011.3540.

[191] S. Rappoccio and S. Mrenna, “FlavorHistory : A heavy flavor overlap tool,” *CMS Analysis Note AN-2009/147* no. September, (2009) 110.

[192] CMS Collaboration, “Muon identification in CMS,” *CMS Analysis Note AN-2008/098* no. October, (2008) 26.

[193] CMS Collaboration, “Performance of muon identification in pp collisions at 7 TeV,” *CMS PAS MUO-10-002* **8** (2010) 27.

[194] CMS Collaboration, “Calorimeter jet quality criteria for the first CMS collision data,” *CMS PAS JME-09-008* (2009) 32.

[195] CMS Collaboration, “Tracking and primary vertex results in first 7 TeV collisions,” *CMS PAS TRK-10-005* (2010) .

[196] L. Lyons, W. Allison, and J. Panellacomellas, “Maximum likelihood or extended maximum likelihood? An example from high energy physics,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **245** no. 2-3, (May, 1986) 530–534. http://linkinghub.elsevier.com/retrieve/pii/0168900286912933.

[197] R. Barlow, “Extended maximum likelihood,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **297** no. 3, (Dec., 1990) 496–506. http://linkinghub.elsevier.com/retrieve/pii/0168900290913348.

[198] F. James, *Statistical Methods in Experimental Physics*. World Scientific Publishing Co. Pte. Ltd., Singapour, 2nd editio ed., 2006.

[199] CMS Collaboration, “Performance of the b-jet identification in CMS,” *CMS PAS BTV-11-001* (2011) .