Université catholique de Louvain
Secteur des Sciences et Technologies
Institut de Recherche en Mathématique et Physique
Center for Cosmology, Particle Physics and Phenomenology

First observation of the Top Quark at the Large Hadron Collider

Doctoral dissertation presented by

Julien Caudron

in fulfilment of the requirements for the degree of Doctor in Sciences

Prof. V. Lemaître (Advisor)	UCL, Belgium
Prof. J-M. Gérard (Chairman)	UCL, Belgium
Prof. F. Blekman	VUB, Belgium
Prof. C. Campagnari	UCSB, USA
Prof. C. Delaere	UCL, Belgium

September 2011

Remerciements

Il est difficile de faire dans l'originalité lorsqu'il s'agit de rédiger la section concernant les remerciements. Le texte ci-dessous ne fera pas exception à la règle. Mais si l'exercice peut paraître banal, l'intention d'exprimer ici une gratitude sincère est bien réelle. Je tiens donc à remercier l'ensemble des personnes ayant permis l'aboutissement de ma formation doctorale et la concrétisation de cette thèse grâce à leurs explications, leurs conseils ou leur soutien.

Je remercie tout d'abord Vincent Lemaître, mon promoteur, pour m'avoir offert la possibilité de réaliser ce doctorat, pour m'avoir encadré dans mon travail de recherche et pour s'être souvent montré disponible. Je remercie ensuite les professeurs Jean-Marc Gérard, Freya Blekman, Claudio Campagnari et Christophe Delaere, pour avoir accepté d'être membre du Comité de lecture de cette thèse. J'ai également beaucoup apprécié leurs nombreuses propositions pour en améliorer le contenu.

Je tiens aussi à remercier spécialement Andrea Giammanco. Non seulement il a été mon plus proche collaborateur et a contribué substantiellement au travail qui a permis l'élaboration de cette thèse, mais de plus, il m'a toujours spontanément offert son aide et prodigué de judicieux conseils. Merci beaucoup.

J'ai eu la chance de pouvoir travailler avec de nombreuses personnes à travers la collaboration CMS. La liste en est malheureusement trop longue pour pouvoir remercier chacune d'elles personnellement. Dans le cadre de la physique du quark top, je pense en particulier aux membres du sous-groupe "dileptonique" (Brown, IFCA d'Oviedo, IPHC de Strasbourg, LIP de Lisbonne et UCSB de Santa Barbara) ainsi qu'aux différents organisateurs en charge du groupe "top" durant la préparation de ma thèse, avec lesquels j'ai apprécié travailler et appris énormément. Dans le cadre de la reconstruction des TrackJets, je remercie tous mes coauteurs avec qui il a été très efficace et très enrichissant de collaborer. Merci également à tous ceux que j'ai

été amené à côtoyer sur divers sujets de manière plus ponctuelle (tâches de service, top commissionning, . . .).

Parmi les aides quotidiennes qui se sont avérées précieuses, je tiens à exprimer ma gratitude envers le secrétariat (en particulier Ginette Tabordon, Carine Baras, Luc Kruijfhooft) qui a toujours répondu sympathiquement lorsqu'il s'agissait de régler les problèmes administratifs et aussi envers l'équipe informatique (en particulier Pavel Demin, Thomas Keutgen, Jérôme De Favereau, Vincent Boucher) pour le support quotidien et l'excellente réactivité aux problèmes rencontrés.

Il a été également très agréable de pouvoir travailler dans un environnement aussi stimulant que le Centre CP3. J'y ai apprécié l'ambiance conviviale et constructive, résultat des qualités humaines de chacun de ses membres. Je souhaite une bonne continuation à ceux qui restent ou qui rejoignent cette équipe. À ceux qui ont ou qui vont quitter CP3, je souhaite qu'ils trouvent un environnement de travail tout aussi sympathique.

Ces remerciements ne seraient pas complets sans un mot pour mes amis et ma famille. Leur soutien et leur intérêt porté à mon travail ont contribué de manière substantielle à ma motivation, je les en remercie du fond du cœur.

Contents

Introduction

This thesis presents the first measurement of the top quark pair production cross section at the CERN Large Hadron Collider (LHC) and subsequently published in January, 2011 [1].

At LHC, the top quark is produced predominantly together with its antiparticle by means of the gluon fusion process governed by the strong interaction. The top and the antitop then decay almost instantaneously via the weak interaction into three possible final states. Amongst them, the dileptonic final state ($\sim 6.5\%$ of the branching ratio), composed of two opposite sign charged leptons (muons and electrons), two neutrinos and two jets, has been studied in this thesis. The results have been obtained from data harvested between March and September 2010 in proton-proton collisions at a centre-of-mass energy of 7 TeV, corresponding to an accumulated luminosity of $3.1\,\mathrm{pb}^{-1}$.

This study takes place in a long top quark physics history, which started with the first theoretical assumptions of its existence in 1970 [2], through indirect experimental evidences (existence of the third generation of leptons in 1975 [3], existence of the bottom quark in 1977 [4], measurement of the bottom quark properties in 80's, precision measurement on the Z boson decay in 1994 [5]) up to its first observation at the TEVATRON in 1995 [6][7], the recent observation of its single production at the TEVATRON in 2009 [8] and in its first observation at LHC in 2010.

The interest in the top quark physics is justified by its particular role with respect to Higgs boson physics and with processes beyond the Standard Model. Indeed, the Standard Model of fundamental interactions between elementary particles has shown spectacular successes by providing precise measurements and predictions, but is currently still partially undetermined due to the absence of experimental settlement of the mechanism breaking the electroweak symmetry and allowing the gauge bosons to have a mass. The fact that the top quark has a mass of the order of the electroweak

symmetry breaking energy scale suggests a special role in this mechanism. Moreover, its large mass has made of it an interesting participant in many models that go beyond the Standard Model. Finally, the top quark provides interesting experimental characteristics, which can also be affected by new physics.

Given the complexity of the topology and possible uncertainties related to the quality of early data, special care on the reconstruction techniques are applied that guaranty the robustness of the analysis. In this thesis, this aspect is emphasized by the design of an analysis which makes a minimal use of the calorimeters. Fortunately, the detector has rapidly appeared to behave outstandingly, and the analysis has provided remarkable result, consistent with the theoretical prediction, with a total uncertainty of about 40%. This first result, noticeable for its robustness, has opened the way to top quark physics at LHC and has already been overtaken by the rapid development of this field in CMS.

The structure of this document is based on three chapters: the top quark physics in general, the description of the detector used in this analysis and the analysis itself.
The first chapter is devoted to the theoretical and experimental context of the top quark physics, progressively oriented to the analysis discussed in the chapter three. In the first part of this chapter, the phenomenological aspect of the top quark is introduced, with its production and decay and its properties. A brief section is also dedicated to an overview of the role of the top quark physics in Beyond Standard Model scenarios. In the second part of this chapter, the indirect evidences of the top quark are briefly mentioned, the tools for its discovery, namely the high energy hadron colliders, are described, and the experimental methods and previous results are stated. This second part concludes with a description of the data that will be used in the analysis.
The second chapter describes the CMS detector and the reconstruction of the different objects that play a role in the dileptonic top quark pair production event identification: the tracks of charged particles, the muons, the electrons and the jets.
The third and last chapter details the analysis per se, with emphasis on the methods used to ensure the robustness of this first observation. The first part explains the selection of the dileptonic top quark pair production events. The second part describes the data-driven techniques used to estimate the backgrounds. The third part is dedicated to the systematic uncertainties study. The fourth part presents the final result: the measured cross section of the top quark pair production. Finally, the last part summarizes the latest results for this cross section obtained in CMS.
An appendix is also available and provides additional information on the reconstruction of jets from tracker information only (“TrackJets”) and its validation.

The work presented in this thesis is essentially based on the following publications and public results:

- "First Measurement of Top-Quark Pair Production Cross Section in Proton-Proton Collisions at $\sqrt{s} = 7\,\mathrm{TeV}$", *CMS Collaboration*, Physics Analysis Summary TOP-01-001, 2010. Phys.Lett.B695 424-443, arXiv:1010.5994
- "Early ttbar cross section in the dilepton channel at $\sqrt{s} = 10$ TeV", *CMS Collaboration*, Physics Analysis Summary TOP-09-002, 2009
- "Commissioning of TrackJets in pp Collisions at $\sqrt{s} = 7$ TeV", *CMS Collaboration*, Physics Analysis Summary JME-10-006, 2010

These public results have been each supported by CMS internal notes to which I have contributed [9][10][11][12]. Even if it was demonstrated in the cross-check exercises that our group was able to reproduce the results for the baseline scenario, the material specific to this scenario used in this dissertation has been provided by the dedicated note [13].

Chapter 1

Top Quark

In this chapter, I give an overview of top quark physics.

The phenomenological point of view is developed in the first section, where the top-quark production and decay, its properties and its implication in new physics researches are described. The experimental point of view is developed in the second section: the first indirect evidences of the top quark, the high energy hadron colliders where the top-quark observation is possible, the selection strategy for its observation and finally the observed and simulated events which are used in this thesis.

1.1 Theoretical framework

The top quark is the heaviest particle known in the Standard Model (SM) of particle physics. It is the weak-isospin partner of the bottom quark and it is therefore a spin-1/2 and charge-2/3 fermion. Even if it obeys the same interaction rules as the other quarks, the top quark plays a particular role in physics due to its large mass of the order of $170\,\mathrm{GeV}/c^2$, ~ 40 times heavier than the bottom quark, which is the second heaviest quark, and $\sim 20\,000$ times heavier than the first generation quarks. This characteristic has an impact on its late direct observation (due to the high energy needed of its production) and on its lifetime (shorter than the typical hadronization time). Its large mass also implies a special role in the Higgs sector (the Yukawa coupling with the Higgs boson is proportional to the fermion mass, which is of the order of unity for the top quark) and in Beyond Standard Model (BSM) physics (due to its possible coupling with exotic particles). In this section, the role of the top quark in the Standard Model

is described and its impact on Beyond Standard Model physics is briefly discussed.

1.1.1 Production and decay of top quarks

As all other quarks, the top quark is a charged particle that interacts strongly (and is therefore coloured) and via the electroweak interaction. The Standard Model Lagrangian parts concerning the quark interactions are:

$$\mathcal{L}^{\text{quark}}_{\text{QCD}} = i\overline{Q}^i_L\,\gamma^\mu(i\,g_S\,G_\mu)\,Q^i_L + i\overline{u}^i_R\,\gamma^\mu(i\,g_S\,G_\mu)\,u^i_R \tag{1.1}$$

$$\begin{aligned}\mathcal{L}^{\text{quark}}_{\text{EWK}} =& i\overline{Q}^i_L\,\gamma^\mu(\partial_\mu + i\,g\,W_\mu + i\,\frac{g'}{6}\,B_\mu)\,Q^i_L \\ &+ i\overline{u}^i_R\,\gamma^\mu(\partial_\mu + i\,\frac{2g'}{3}\,B_\mu)\,u^i_R + i\overline{d}^i_R\,\gamma^\mu(\partial_\mu - i\,\frac{g'}{3}\,B_\mu)\,d^i_R\end{aligned} \tag{1.2}$$

$$\mathcal{L}^{\text{quark}}_{\text{Yukawa}} = -\Gamma^{ij}_u\,\overline{Q}^i_L\,\epsilon\phi^*\,u^j_R - \Gamma^{ij}_d\,\overline{Q}^i_L\,\phi\,d^j_R + h.c. \tag{1.3}$$

where Eq. 1.1 describes the strong interaction, Eq. 1.2 the electroweak interaction and Eq. 1.3 the coupling to the Higgs boson. In these equations, the quarks are represented by Q_L, u_R and d_R, respectively the left-handed weak-isospin doublets, the right-handed up-type weak-isospin singlets and the right-handed down-type weak-isospin singlets. The three different generations are labelled by the lowercase Latin letters i and j and a summation over them is implied.

In the QCD Lagrangian Eq. 1.1, g_S is the strong coupling constant and G_μ the gluon gauge field ($G_\mu = \lambda_a G^a_\mu$ with λ_a the Gell-Mann matrices).

In the electroweak Lagrangian Eq. 1.2, g and g' are the electroweak coupling constants, W_μ is the SU(2) gauge field ($W_\mu = \tau_a W^a_\mu$ with τ_a the Pauli matrices) and B_μ is the U(1) gauge field singlet. The γ, $W^\pm$ and Z^0 particles can be formulated in terms of these gauge fields as respectively $\cos(\theta_W)B + \sin(\theta_W)W^3$, $1/\sqrt{2}(W^1 \mp iW^2)$ and $\sin(\theta_W)B + \cos(\theta_W)W^3$ with $\theta_W = g/\sqrt{g^2+g'^2} = M_W/M_Z$ the weak angle.

Finally, in the Yukawa interaction with the Higgs field (Eq. 1.3), ϵ is the total antisymmetric tensor in two dimensions, ϕ is the Higgs field and Γ are the complex matrices in generation space which describe the mixing between different generations.

Top quark pair production. The top quark pair production ($t\bar{t}$) at hadron colliders is a QCD process. Due to the confinement of the quarks, the computation of the cross

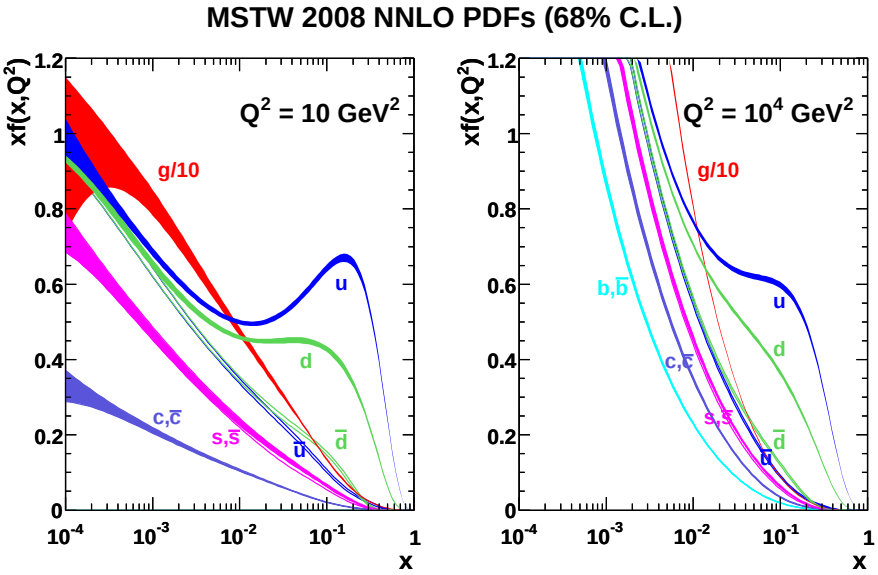


Figure 1.1: Parton distribution functions for the proton at $Q^2 = 10\,\text{GeV}^2$ and $Q^2 = 10\,000\,\text{GeV}^2$, from the MSTW 2008 NLO PDF set [14].

sections of such processes are not straightforward. However, in the case of hadrons at high energy, which is the case if the top quark mass threshold is reached, the QCD-improved parton model can be used: in this model, the process is factorized between a hard parton-parton cross section ($\hat{\sigma}$), which can be computed in perturbative QCD, and parton distribution functions (PDF), which absorb the divergences and which are determined experimentally from other deep inelastic experiments (e.g., MSTW [14] or CTEQ[15] collaboration). To compute the hard parton-parton cross section, the partons of the hadrons are considered as quasi-free with a momentum corresponding to a fraction x of the hadron momentum. This cross section is computed at a finite order in the perturbation series, which implies the use of the renormalization scale μ to regulate divergent terms. In the $t\bar{t}$ computation, this scale is chosen to be $\mu = m_t$. This hard parton-parton cross section is convoluted with the parton distribution functions $f(x)$ of the two partons to run over all the possible x values, $f(x)$ corresponding to the probability of finding parton of momentum fraction x. These PDFs also depend on a scale, called factorization scale, usually chosen to have the same value as the renormalization scale. One of the proton PDF parametrizations is depicted in Fig. 1.1. At this stage, the cross section for two colliding hadrons A and B is:

$$\sigma(AB \to t\bar{t}) = \sum_{i,j} \int dx_i dx_j f_{i,A}(x_i, \mu^2) f_{j,B}(x_j, \mu^2) \hat{\sigma}_{ij}(ij \to t\bar{t}; \hat{s}, \mu^2) \quad (1.4)$$

where f are the parton distribution functions and $\hat{\sigma}$ the hard parton-parton cross section. The sum runs over all pairs of possible partons. The parton distribution functions f depend of course of the nature of the parton and the nature of the hadrons used in the hadron collider.

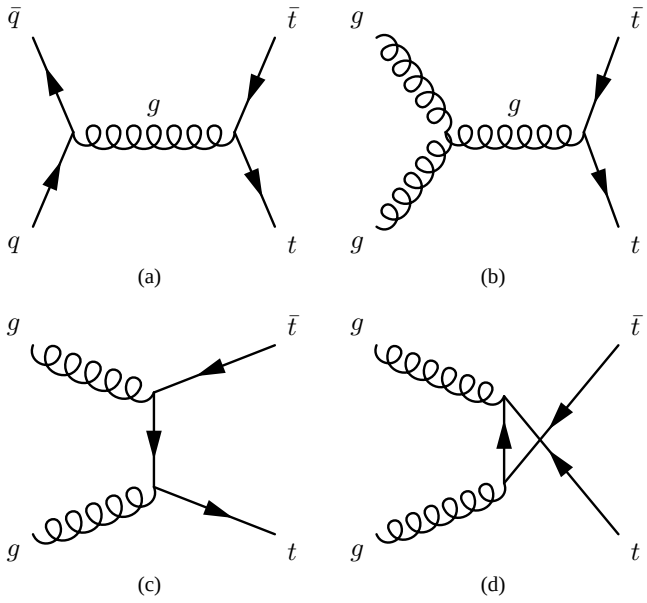
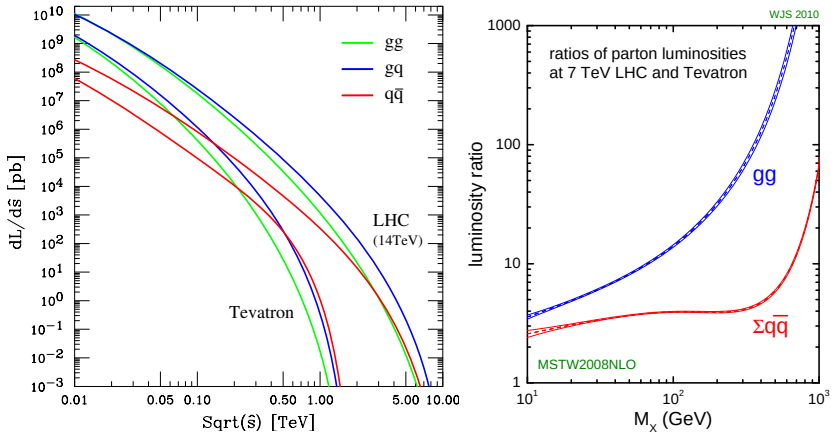


Figure 1.2: Main $t\bar{t}$ production diagrams: quark-antiquark annihilation (a) and gluon-gluon fusion (b, c, d).

The leading order (LO) of the hard parton-parton cross section is given by quark-antiquark annihilation and gluon-gluon fusion, as shown by Fig. 1.2. The cross section for quark-antiquark annihilation is given by:

$$\frac{d\hat{\sigma}}{d\hat{t}}(q\bar{q} \to t\bar{t}) = \frac{4\,\pi\,\alpha_s^2}{9\,\bar{s}^4}[(m_t^2 - \hat{t})^2 + (m_t^2 - \hat{u})^2 + 2m_t^2\hat{s}] \tag{1.5}$$

with $\hat{s} = (p_q + p_{\bar{q}})^2$, $\hat{t} = (p_q - p_t)^2$, $\hat{u} = (p_q - p_{\bar{t}})^2$ the Mandelstam variables of the partonic process. The next-to-leading order (NLO) calculation adds the virtual contributions to the LO processes, gluon bremsstrahlung processes ($q\bar{q} \to t\bar{t} + g$ and $gg \to t\bar{t} + g$) and flavour excitation processes (like $g + q(\bar{q}) \to t\bar{t} + q(\bar{q})$). Additional corrections can be provided at next-to-leading-logarithms using Sudakov resummation [16] which leads to approximate next-to-next-to-leading order (NNLO). The exact computation at NNLO is currently a work in progress.



(a) (b)

Figure 1.3: (a) Parton luminosity for TEVATRON and LHC at $\sqrt{s} = 14$ TeV for gluon-gluon fusion (gg), quark-antiquark annihilation ($q\bar{q}$) and quark excitation (gq), obtained from CTEQ6.1 PDFs [15]. The results for LHC at $\sqrt{s} = 7$ TeV are similar. (b) Ratio between the LHC at $\sqrt{s} = 7$ TeV parton luminosities and TEVATRON parton luminosities, for the gluon-gluon fusion and quark-antiquark annihilation, obtained from the MSTW 2008 NLO PDFs [14].

At the TEVATRON collider, protons and antiprotons are colliding at centre-of-mass energy of 1.96 TeV. For this energy, the mass of the top quark remains relatively large with respect to the centre-of-mass energy, and the $t\bar{t}$ production are mainly produced at large x region. For the LHC, lower x region are accessible while the colliding hadrons are only protons. In order to compare the processes at the different colliders, the parton luminosities can be computed:

$$L_{ij}(\hat{s}; s) = \frac{1}{s} \int_{\hat{s}}^{s} f_{i,A}(\frac{\tilde{s}}{s}) f_{j,B}(\frac{\hat{s}}{\tilde{s}}) \frac{1}{\tilde{s}} d\tilde{s} \tag{1.6}$$

As visible in Fig. 1.3 which compare this luminosity for the different parton interaction at the TEVATRON and at LHC, the LHC provides a significant larger luminosity than TEVATRON. At the TEVATRON, the dominant production process is the quark-antiquark annihilation ($q\bar{q}$) ($\sim 85\%$), while at LHC, it is dominated by quark excitation (gq) followed by gluon-gluon fusion (gg), which lead to a production process dominated by gluon-gluon fusion ($\sim 80\%$) (because quark excitation contribute to $t\bar{t}$ only at NLO).

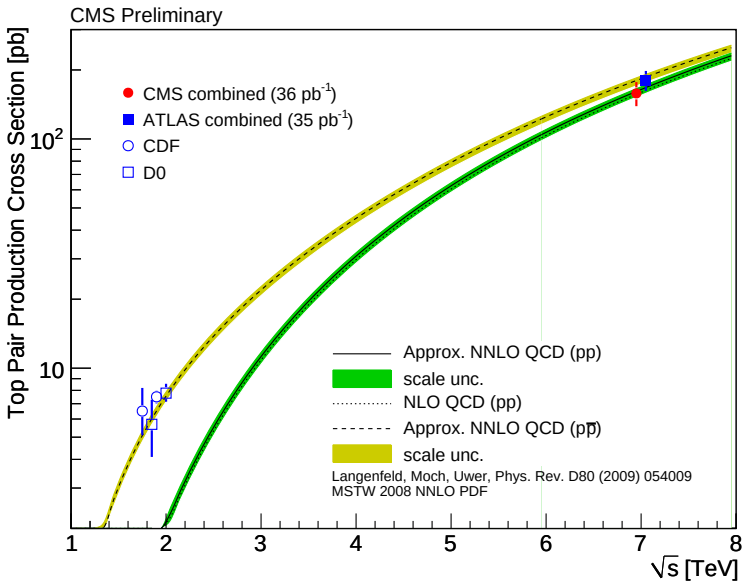


Figure 1.4: Latest results from TEVATRON ($p\bar{p}$) and LHC (pp) on top pair production cross section as a function of $\sqrt{s}$, from [17]. Data points are slightly displaced horizontally for better visibility. Both TEVATRON experiments have provided results at $\sqrt{s} = 1.8\,\text{TeV}$ and $\sqrt{s} = 1.96\,\text{TeV}$. Theory predictions at approximate NNLO are obtained using HATHOR [18]. The error band of the prediction corresponds to the scale uncertainty.

The theoretical cross sections for $t\bar{t}$ production at the TEVATRON and at LHC are detailed in Tab. 1.1. The latest results from TEVATRON and LHC agree with the theory, as illustrated in Fig. 1.4.

Single top quark production. The top quark can also appear singly when it is produced by electroweak interaction, as illustrated in Fig. 1.5. The production is called "single top" and is divided in three cases according to the virtuality of the W boson:

- The t-channel, where a bottom quark interacts with an up or antidown quark via a space-like virtual W boson. The up or antidown quark is usually a valence parton of one of the hadrons while the bottom quark is a sea quark of the hadron or produced from a gluon splitting;

Table 1.1: Theoretical cross sections for the $t\bar{t}$ production, for different incoming particles, at different centre-of-mass energies, using parton distribution functions derived from different physics collaborations, for a given top quark mass and at a given level of calculation. The centre-of-mass energy is given in TeV, the cross section in pb and the top mass in GeV/c^2.

Coll.	$\sqrt{s}$	Cross section	PDF	m_t	Order	Ref.
$p\bar{p}$	1.8	$5.19\ ^{+0.52}_{-0.68}$	CTEQ6M	175	NLO+NLL	[19]
$p\bar{p}$	1.96	$7.81\ ^{+0.45}_{-0.53}$	CTEQ6.6	171	app. NNLO	[20]
$p\bar{p}$	1.96	7.08 ± 0.36	MSTW 2008	173	app. NNLO	[21]
$p\bar{p}$	1.96	$7.48\ ^{+0.59}_{-0.65}$	CTEQ6.6	171	app. NNLO	[22]
pp	7	$163\ ^{+11}_{-10}$	MSTW 2008	173	app. NNLO	[21]
pp	7	$164\ ^{+10}_{-13}$	MSTW 2008	173	app. NNLO	[23]
pp	7	$158\ ^{+23}_{-24}$	PDF4LHC	172.5	NLO	[24]

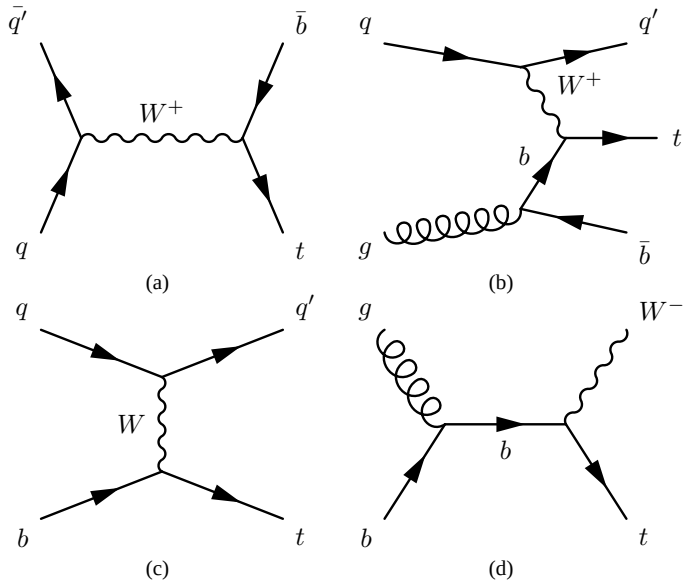


Figure 1.5: Main single top production diagrams: s-channel (a), t-channel (b, c) and W-associated production (d).

- The s-channel, where a top and an antibottom quarks are produced by a time-like W boson. The W boson is usually obtained from the fusion of two valence

quarks. The W boson has to have a virtually sufficiently higher to produce a top quark ($q^2 \geq (m_t + m_b)^2$);

- The W-associated production (tW), where the top quark is produced in association with a real W boson from a gluon and a bottom quark.

This single top processes have a lower cross section than $t\bar{t}$ production. The theoretical cross sections for single top production at the TEVATRON and at LHC are detailed in Tab. 1.2. The first evidence for single top has been obtained in 2007 at the TEVATRON [8] and in 2010 the first cross section at 7 TeV has been measured at LHC [25]. The latest results from TEVATRON and LHC are illustrated in Fig. 1.6. The LHC results agree within 1σ with the theory, while some less-than-2σ discrepancies are visible for the TEVATRON results.

Table 1.2: Theoretical cross sections for the single top production, for the different channels, for different incoming particles, at different centre-of-mass energies, using parton distribution functions derived from different physics collaborations, for a given top quark mass and at a given level of calculation. The centre-of-mass energy is given in TeV, the cross section in pb and the top mass in GeV/c^2. $4f$ and $5f$ correspond respectively to the diagrams b) and c) illustrated in Fig. 1.5.

Coll.	$\sqrt{s}$	Channel	Cross section	PDF	m_t	Order	Ref.
$p\bar{p}$	1.96	t-chan.	1.04 ± 0.06	MSTW 2008	173	app. NNLO	[21]
$p\bar{p}$	1.96	s-chan.	$0.523\ ^{+0.030}_{-0.028}$	MSTW 2008	173	app. NNLO	[21]
pp	7	t-chan. (t)	$41.7\ ^{+1.8}_{-0.8}$	MSTW 2008	173	app. NNLO	[21]
pp	7	t-chan. ($\bar{t}$)	$22.5\ ^{+0.9}_{-1.0}$	MSTW 2008	173	app. NNLO	[21]
pp	7	s-chan. (t)	$3.17\ ^{+0.14}_{-0.12}$	MSTW 2008	173	app. NNLO	[21]
pp	7	s-chan. ($\bar{t}$)	$1.42\ ^{+0.06}_{-0.07}$	MSTW 2008	173	app. NNLO	[21]
pp	7	tW	$7.8\ ^{+0.5}_{-0.6}$	MSTW 2008	173	app. NNLO	[21]
pp	7	t-chan. $4f$	$59.1\ ^{+3.0}_{-4.0}$	CTEQ6.6	172.5	NLO	[26]
pp	7	t-chan. $5f$	$62.3\ ^{+2.3}_{-2.4}$	CTEQ6.6	172.5	NLO	[26]

Top quark decay. In the Standard Model, the top quark is coupled with the 1/2 SU2-isospin quarks via the charged electroweak boson. However, these couplings are proportional to the CKM matrix element, and the tdW and tsW coupling are then strongly suppressed. The top quark decays then dominantly into a bottom quark and

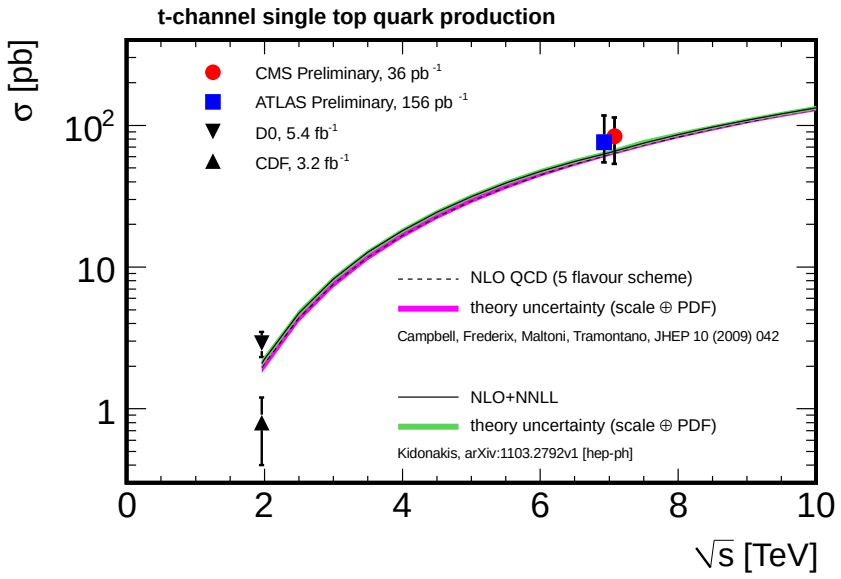


Figure 1.6: Latest results from TEVATRON and LHC on t-channel single top production cross section as a function of $\sqrt{s}$, from [27]. LHC data points are slightly displaced horizontally for better visibility.

a W boson. The top-quark width Γ_t is given by:

$$\Gamma_t = \frac{G_F m_t^3}{8\pi\sqrt{2}} |V_{tb}|^2 (1 - \frac{m_W^2}{m_t^2})^2 (1 + 2\frac{m_W^2}{m_t^2})[1 - \frac{2\alpha_s}{3\pi} f(\frac{m_W^2}{m_t^2})] \tag{1.7}$$

The $\frac{2\alpha_s}{3\pi} f(\frac{m_W^2}{m_t^2})$ contribution is obtained from QCD correction, where $f(y) = \frac{2}{3}\pi^2 - 2.5 - 3y + 4.5y^2 - 3y^2\ln(y)$, and corresponds to a 10% suppression of the leading order width. For $m_t = 172.6\,\text{GeV}/c^2$, this width is $1.34\,\text{GeV}$. Adding the electroweak correction increases Γ_t by 1.7%. But this is almost exactly compensated by taking into account the finite width of the W boson which reduces Γ_t by 1.5%. The experimental value obtained from TEVATRON data is $\Gamma_t = 2.0\,^{+0.7}_{-0.6}\,\text{GeV}$ [28].

This large Γ_t implies a small lifetime of $\sim 5 \times 10^{-25}\,\text{s}$. The hadronization time is of the order of $\sim 1\,\text{fm}/c \simeq 3 \times 10^{-24}\,\text{s}$. The top quark decays thus before any hadronization.

By conservation of angular momentum and because of the $V - A$ structure of the coupling, only left-handed particles are expected to couple to W bosons. Therefore, the top quark decays in a longitudinally polarized W^+ boson ($h^W = 0$) and a left-handed bottom quark (the mass of the bottom quark being negligible with respect to

the energy from the decay of the heavy top quark, the chirality conservation is directly expressed in the helicity) emitted in an antiparallel direction with respect to the top quark spin axis, or in a transversely polarized W^+ boson ($h^W = -1$) and a left-handed bottom quark emitted in a parallel direction with respect to the top quark spin axis. $h^W = 1$ is not possible (if $m_b = 0$). The ratio widths in the three W helicity states are given by:

$$\Gamma(h^W = -1) : \Gamma(h^W = 0) : \Gamma(h^W = +1) = 1 : \frac{m_t^2}{2m_W^2} : 0$$

For antitop, it is the negative helicity which is forbidden.

QCD corrections affect the width ratios by increasing the fraction of the left-handed W bosons by 2.2% and decreasing the fraction of the longitudinal W bosons by 1.1%. The electroweak and finite width of the W boson corrections have negligible effect on ratios (of the order of the per-mille).

If the mass of the bottom quark is not neglected, the ratios are not really affected, but a small positive helicity W boson fraction is possible: $\Gamma(h^W = +1)/\Gamma = 3.6 \times 10^{-4}$ at Born level.

The produced W boson decays leptonically ($32.40 \pm 0.27\%$) or hadronically ($67.60 \pm 0.27\%$), which leads to three different final states:

- $t\bar{t} \rightarrow W^+b\, W^-\bar{b} \rightarrow \ell\bar{\nu}_\ell\; b\; \bar{\ell}'\nu_{\ell'}\; \bar{b}$, called "dileptonic channel", corresponding to 10.3% of the $t\bar{t}$ decay;
- $t\bar{t} \rightarrow W^+b\, W^-\bar{b} \rightarrow q\bar{q}'\; b\; \ell\bar{\nu}_\ell\; \bar{b} + \bar{\ell}\nu_\ell\; b\; q\bar{q}'\; \bar{b}$, called "semileptonic channel" (or "lepton+jets channel"), corresponding to 43.5%;
- $t\bar{t} \rightarrow W^+b\, W^-\bar{b} \rightarrow q\bar{q}'\; b\; q''\bar{q}'''\; \bar{b}$, called "hadronic channel", corresponding to 46.2%.

Usually, the leptons used in the analysis are only electrons and muons because the tau lepton, which decays inside the detector, is more difficult to reconstruct and to identify. So, except when explicitly specified, the contribution from the tau semileptonic decay is not taken into account in the analyses. The leptonic decay of the tau can contribute, even if the selection is not optimized for its kinematic, because it appears as an electron or a muon final state. Taking into account the maximal possible contribution from the tau leptons, the dileptonic channel of the $t\bar{t}$ decay corresponds to $\sim 6.5\%$ and the semileptonic channel to $\sim 34\%$. In this thesis, I study three modes of the dileptonic channel: the ee channel, the $\mu\mu$ channel, and the $e\mu$ channel, where the final state is composed respectively of two electrons, two muons and one electron and one muon. In the semileptonic channel, the two most studied modes are the e+jets channel and

the μ+jets channel, where the final state is composed respectively of an electron and a muon.

1.1.2 Other top quark properties

$|\mathbf{V_{tb}}|$ CKM matrix element. The interaction eigenstates do not necessarily correspond to the mass eigenstates, depending of the Γ matrices in Eq. 1.3. In order to handle the particles as the mass eigenstates, a transformation of the Q, u and d interaction eigenstates where the Γ matrices are diagonalized can be done, which introduces the Cabibbo-Kobayashi-Maskawa (CKM) unitary rotation matrix in the Lagrangian. This transformation appears to be irrelevant in all quark interactions excepted when it couples with a $W^{\pm}$ boson. The $W^{\pm}$ interaction mixes the mass eigenstates of the quarks according to the elements of the CKM matrix expressed, in the Standard Model where three generations of quarks are defined, as:

$$\begin{aligned} V_{CKM} &= \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \\ &= \begin{pmatrix} 0.97428 \pm 0.00015 & 0.2253 \pm 0.0007 & 0.00347\ ^{+0.00016}_{-0.00012} \\ 0.2252 \pm 0.0007 & 0.97345\ ^{+0.00015}_{-0.00016} & 0.0410\ ^{+0.0011}_{-0.0007} \\ 0.00862\ ^{+0.00026}_{-0.00020} & 0.0403\ ^{+0.0011}_{-0.0007} & 0.999152\ ^{+0.000030}_{-0.000045} \end{pmatrix} \end{aligned} \tag{1.8}$$

with the numerical values from [28].The diagonal entries are close to the unity, when the off-diagonal are closer to zero, especially for heavy quarks. In top quark physics, the V_{tb} matrix element plays a central role because it defines the strength of the charged weak interaction of the top quark. Assuming unitarity of the 3×3 matrix, the V_{tb} matrix element can be estimated from the other measured matrix elements. In this case, the $|V_{tb}|$ value is bound between 0.9990 and 0.9992 at 90% C.L. [29]. The single top production cross section (cf. Section 1.1.1) is directly proportional to $|V_{tb}|^2$, and assuming a top decay dominated by the W boson and bottom quark production, the experimental limit on $|V_{tb}|$ obtained from TEVATRON is given by $|V_{tb}| > 0.78$ [30][31]. In the top quark pair production ($t\bar{t}$), the measured variable is $R_b = \frac{|V_{tb}|^2}{|V_{td}|^2+|V_{ts}|^2+|V_{tb}|^2}$, obtained by comparing events with and without jets identified as coming from a bottom quark decay. If the 3×3 matrix unitarity is assumed, it reduces to $R_b = |V_{tb}|^2$. If it is assumed that the top quark cannot decay to other quarks than the ones already observed, the limit is $\frac{|V_{tb}|^2}{|V_{td}|^2+|V_{ts}|^2} > 3.8$ at 95% C.L. [32]. All these measurements are compatible with the Standard Model, but will gain from improved studies of the top quark.

Mass. In the Standard Model, the fermions acquire mass due to the Brout-Englert-Higgs spontaneous symmetry breaking mechanism, which allows the apparition of $y\bar{\psi}\phi\psi$ Yukawa term, where y is the Yukawa coupling constant, ψ the fermions spinors and ϕ the Higgs scalar complex field, cf. Eq. 1.3. At the first order around the vacuum-expectation value of the Higgs boson, this term looks like a mass term $m\bar{\psi}\psi$ where the mass m is then given by $y\langle\phi\rangle$. The Yukawa coupling constant y being a free parameter, the Standard Model does not determine explicitly the fermions masses, which have to be measured.

The experimental determination of the quarks mass is complicated by the fact that they are coloured particles. Quantum chromodynamics (QCD) therefore implies that they are confined and cannot appear as asymptotic states. However, in the top quark case, this quark decays before any hadronization (cf. Section 1.1.1) and thus allows the extraction of its pole mass, which is the real part of the pole of its propagator: $\sqrt{p^2} = m_{\text{pole}} - \frac{i}{2}\Gamma$. Experimentally, a distribution closely related to this variable is obtained and is fitted to simulated distributions generated with different values of the top-quark mass as generator parameter. The observed top-quark mass is therefore the one which corresponds to the parameter of the generator.

There are two main concerns with this approach: Firstly, the pole mass can be shifted by non-perturbative QCD corrections and its theoretical expression depends on the renormalization scheme. This leads to an intrinsic ambiguity of order $\Lambda_{QCD} \sim 500\,\text{MeV}$ [33][34]. Secondly, the measured top-quark mass is in fact the parameter of the generator used for simulation. These generators are restricted to a fixed-order calculation and the measured mass does not corresponds to the pole mass because the latter absorbs some corrections that the former does not (due to parton shower cut offs and restriction in hadronization modelling).

Other methods avoiding such complication have been developed, e.g., the extraction of the running mass of the top quark from the top pair production cross section. However, because the low sensitivity of the total cross section with respect to the mass, it is limited by the theoretical uncertainty and requires a good precision. It is then not suited for early observation.

The more precise measurement of the top-quark pole mass available at the time of writing is given by combination of the several direct mass measurement from TEVATRON. Combining results from CDF and DØ experiments using up to 5.6 fb^{-1} of data, the average mass of the top quark has been measured at $m_{top}^{exp} = 173.3 \pm 1.1\,\text{GeV}/c^2$ [35].

Several Standard Model observables depend on the top-quark mass value via radiative corrections. For example, the quark loop corrections contribute to the W boson and Z^0 boson masses by a term proportional to the square of the quark mass. Because of

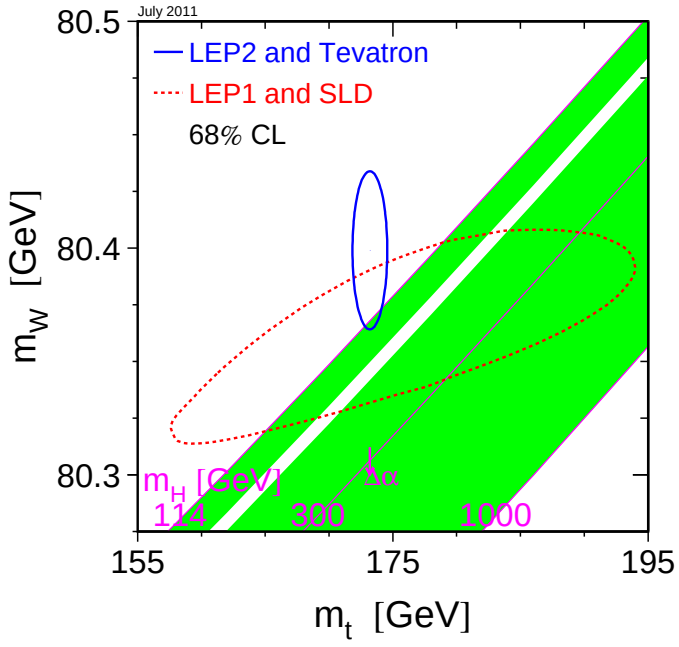


Figure 1.7: Comparison of the indirect constraints and direct measurements of m_W and m_t based on LEP and TEVATRON data. The contours correspond to 68% C.L. and the green bands shows the relationship for the masses as function of the SM Higgs boson mass for region not excluded. The $\Delta\alpha$ arrow shows the variation of this relation if $\alpha(m_Z^2)$ is varied by $\pm\sigma$. From [36].

the heaviness of the top quark, its contribution is largely dominant in these corrections. This has allowed an estimation of the top-quark mass before any direct measurement. For illustration, the extraction of the top-quark mass using Z-pole data, W boson mass and other electroweak quantities gives $m_{top} = 178.9\ ^{+11.7}_{-8.6}$ GeV/c^2 [36]. Similarly, the knowledge of the top-quark mass allows the extraction of information on other corrections. An improvement of the top-quark mass measurement leads then to a stronger constraint on the Higgs mass which also intervenes in corrections. The relation between the top-quark mass and the $W^{\pm}$ boson mass for different Higgs boson mass hypothesis is illustrated in Fig. 1.7.

Charge. The Standard Model top quark has a 2/3 charge. However, the usual analyses, which assume $t \to W^+b$ and $\bar{t} \to W^-\bar{b}$, are not taking into account the charge

of the bottom quark, allowing the possibility of $\pm 4/3$ charge exotic top quark decaying into $W^+\bar{b}$ or W^-b. The total production observed in the standard analyses can thus contain a contamination from an exotic top quark. However, the existence of 2/3 charge top quark has been confirmed by analyses associating the jets coming from the bottom quarks to the reconstructed W bosons in the way which corresponds to the best kinematical fit with constraints on m_t and m_W and measuring the charge associated to the jets directly correlated to the bottom quark charge. This jet charge determination can be done via a weighted sum of the charge of the tracks inside the jets: $Q_{jet} = \sum w_i q_i / \sum w_i$, where the weights have been defined according to the longitudinal or transverse component of the track momentum with respect to the jet momentum, or via the detection of soft leptons produced by a semileptonic decay of the bottom quark. Results from TEVATRON have limited the contamination from $4/3$ charge top-like particle to 80% at 90% C.L. [37] and have excluded the $-4/3$ charge exotic top quark model at the 99% C.L. [38].

V − A structure. The top quark decays into a bottom quark and a W boson with a branching ratio $\simeq 100\%$ (cf. Section 1.1.1). Due to the V − A structure of the weak decay, only the left-handed particles can couple to the W boson. So, the W boson from the decay is either left-handed ($-$, $\sim 30\%$) or longitudinal (0, $\sim 70\%$), and the right-handed contribution is expected to be negligible ($+$, $\sim 0\%$). The W boson helicity can be observed through the charged lepton in which it has decayed via several observables sensitive to the helicity: transverse momentum of the lepton, invariant mass of the lepton and the associated bottom quark, angle θ^* between the lepton and the associated bottom quark in the W boson rest frame. The latest studies at the TEVATRON have been performed using a kinematical fit method to select events from which $\cos\theta^*$ can be extracted [39] or using Matrix Element method [40]. Both methods show a good agreement with the SM expectations.

Forward-Backward asymmetry. Another interesting feature of the $t\bar{t}$ production is the forward-backward asymmetry between the top quark and antitop quark. This kinematical effect appears at NLO, due to radiative correction in quark-antiquark annihilation (in initial and final state, but also as interference with the box diagram) and interference of gluon-quark scattering ($g + q \rightarrow Q + \bar{Q} + q$) [41][42]. This asymmetry A_{FB} can be observed experimentally when the initial partons have a charge asymmetry by measuring the normalized difference between the number of events N_F in the forward region ($y \geq 0$) with respect to the number of events N_B in the backward region ($y \leq 0$): $A_{FB} = (N_F - N_B)/(N_F + N_B)$. The forward and backward regions can be defined in the laboratory frame using the rapidity of the top quark: $A_{FB}^{p\bar{p}}$ or in

the centre of mass of the top and antitop system using the difference between the top and the antitop rapidity: $A_{FB}^{t\bar{t}}$.

At the TEVATRON, where the initial state of the colliding hadrons is charge asymmetric, the forward-backward asymmetry amount at NLO is expected to be of the order of 8% (resp. 5%) for $A_{FB}^{t\bar{t}}$ (resp. $A_{FB}^{p\bar{p}}$). DØ using 5.4 fb^{-1} of data has found $A_{FB}^{t\bar{t}} = 0.196 \pm 0.065$ for the detector phase space which correspond to a theoretical asymmetry of 0.05[43]. The latest results at CDF using 5.6 fb^{-1} of data is $A_{FB}^{t\bar{t}} = 0.158 \pm 0.074$(syst. + stat.)[44] when a theoretical asymmetry of 0.058 ± 0.009 where expected. Even if these results are agreeing with the SM prediction within $\sim 2\sigma$, the results for large invariant top-antitop mass shows a deviation higher than 3σ.

At LHC, which is a proton-proton machine, the top-antitop production is symmetric in the laboratory frame, but small effect can be observed due to the $t\bar{t} + g$ contribution, where the antitop quark appears to be more central than the top quark. The measured variable is the charge asymmetry A_C, based on the difference between the absolute value of the (pseudo)rapidity of the top quark and the antitop quark. At CMS, the charge asymmetry A_C^y has been measured at -0.013 ± 0.026(stat.) $^{+0.026}_{-0.021}$(syst.) after 1.09 fb^{-1} while 0.011 ± 0.001 was expected from theory [45]. At ATLAS, the charge asymmetry A_C^y has been measured at -0.023 ± 0.015(stat.) ± 0.023(syst.) after 0.70 fb^{-1} while 0.005 ± 0.001 was expected from theory [46]. Both measurements show good agreement with respect to the Standard Model expectations, but are limited by the low amount of data.

Flavour-changing neutral current. Flavour-changing neutral currents (FCNC) are suppressed in the Standard Model: this kind of process is impossible at tree level while the quantum loops contributions are below 10^{-10}. The FCNC processes can however be enhanced by new physics. The presence of FCNC can be observed in two top quark decays channels: $t \to Zu(c)$ and $t \to \gamma u(c)$, and on single top quark production: $u(c) + g \to t$. The single top production has been studied at LEP, HERA and CDF and has led to an upper limit of 3.9×10^{-4} on $t \to u + g$ and 5.7×10^{-3} on $t \to c + g$ [47]. Branching ratios for $t \to Zu(c)$ and $t \to \gamma u(c)$ have been constrained at CDF to be lower than 3.7% [48] and 3.2% [49], respectively. Prospects are also available at LHC [50][51].

Spin correlation. Due to the fast decay of the top quark, its spin is preserved on average despite the depolarizing effect of gluon emission by strong interaction. This allows an observation of the possible top quark polarization via the angular distribution of its decay products. The spin of the two top quarks in the $t\bar{t}$ production are correlated,

depending on the production process: gluon-gluon fusion will dominantly yield a like-helicity $t\bar{t}$ pair, while quark annihilation will dominantly yield an opposite-helicity $t\bar{t}$ pair [52]. A commonly used observable is the asymmetry $\kappa = (N_o - N_s)/(N_o + N_s)$, where N_o and N_s are the opposite-sign helicity events and the same-sign helicity events respectively, measured in a given spin quantization axis. The measurement of the spin correlation in $t\bar{t}$ can be used to confirm the predicted $t\bar{t}$ production processes fraction, to provide an experimental upper limit to the top-quark lifetime and to indicate possible contributions from new physics which can impact the spin correlation. The current measurements performed at the TEVATRON are limited by uncertainties, but consistencies at at least 95% C.L. have been found [53][54][55].

1.1.3 Top quark and physics BSM

Due to its large mass of the order of the electroweak symmetry breaking scale, the top quark plays a special role in many theories beyond the Standard Model. The top-quark phenomenology also provides many observables which can be tested under the Standard Model hypothesis. A discrepancy between a predicted observable and its measurement can therefore indicate a new physics effect. This section briefly summarizes some of the main proposals of investigation of unobserved physics in top-quark phenomenology. More details, which are beyond the scope of this thesis, can be found in the references provided. I will notably include considerations about the Higgs boson in this section, even if this is sometimes considered as a Standard Model topic.

Numerous new physics can affect the way the top quark is produced or decays, or can enhance the measurement of the rate of a given channel by providing a similar final state.

If a model predicts a heavy neutral particle which can couple with the top quark [56], this particle can decay in a top quark pair and contributes to the standard model final state. This can also be the case through dynamic $t\bar{t}$ condensate formed by a new strong gauge force. In both case, the final state of these processes allows the observation of a resonance in the $t\bar{t}$ invariant mass. This is challenging for very high masses of the resonance, because of the boost of the top quarks produced by its decay, and the consequent small angular separation between the final-state objects; this leads to the development of specific algorithms, which reconstruct large jets containing substructures corresponding to the W boson decay products and the bottom quark hadronization. Non resonant top quark pair production is also possible. A first example is the associated Higgs boson production $t\bar{t}H$ where the Higgs boson decays in two bottom quarks, leading to four b-jets and two opposite charge W boson. This process is expected to be small in the Standard Model but can be used in combination

to increase the Higgs boson discovery. In other models, this process can be enhanced. A second example is the decay of a massive vector bosons (W') into top quarks and antibottom quarks. In a model independent approach, new operators introduced in a low energy (with respect to the new physics threshold) effective Lagrangian can interfere with the Standard Model Lagrangian and can affect the $t\bar{t}$ invariant mass in a non-resonant way [57]. In this approach, the contribution of the different operators can be constrained by the observation. The $t\bar{t}$ production being dominated by quark-antiquark annihilation (mainly up-quark annihilation) at the TEVATRON and by gluon fusion at LHC, the constraints on the different operator contributions are complementary in the two experiments. Single top production can also be affected through a W' boson or a charged Higgs boson (predicted by multi-Higgs models). Both channels are identifiable with the invariant mass of the decay products. The heavy W' boson can also affect the observed helicity if this boson couples to right-handed particles.

If a model predicts a new particle with a mass smaller than the top quark and couples with it, this particle can affect the top quark decay. It is the case in SUSY models, where the Higgs sector is composed of at least five Higgs particles, including two charged Higgs bosons which are expected to couple to top quark and mediate their decay. The top quark decay can then be mediated by $t \rightarrow H^+ b$ where $H^+ \rightarrow \bar{\tau}\nu$ or $H^+ \rightarrow c\bar{s}$ (or even $H^+ \rightarrow W^+ h_0$ where h_0 is the charge parity odd neutral Higgs boson which can decay in two tau leptons). This effect can be observed by the study of these new final states or by the observation of a lack of events with respect to the expectation due to the non optimization of the standard $t\bar{t}$ selection for these new final states. The top-quark lifetime can also be perturbed by new physics, such as fourth generation, non-standard top decay models or other extension, leading to a long-lived top quark. This effect can be observed by the presence of a displaced vertex for the W boson decay or by the spin decorrelation in the top quark pair production.

New particles can also replace the top quark to mimic the final state. A first example is the light stop particle $\tilde{t}_1$, which is the lighter of the two super partners of the top quark in supersymmetric models. This particle can be produced in pair and then decayed in a neutralino ($\tilde{\chi}_1^0$) and a top quark or in a chargino ($\tilde{\chi}_1^+$) and a bottom quark. In the latter case, the chargino decays in neutralino and W boson or in neutralino and leptons. In all the cases, assuming a stable neutralino, the final state is similar to the $t\bar{t}$ production expected possible high missing transverse energy. A second example is the additional heavy quark t', which appears in several models (fourth generation, Little Higgs, ... [58]). This process can be distinguished from the $t\bar{t}$ production by the mass reconstruction and by the scalar sum of of the transverse energy of reconstructed objects.

Finally, possible discrepancies observed between the measured observables and their expected value from the Standard Model can indicate the presence of an interaction

mediated by a new particle. For example, in the case of the forward-backward asymmetry, where a deviation of more than 3σ has been observed at the TEVATRON for high $t\bar{t}$ mass (cf. Section 1.1.2), this can be due to FCNC induced by a massive neutral vector boson. This hypothesis has been rejected due to the absence of same-sign top quark pair events [59]. Other observables (discussed in Section 1.1.2), such as the Wtb vertex coupling, the spin correlation, the flavour changing neutral current, the W boson polarization or the top quark charge can also potentially indicate the effect of new physics.

The analyses on these matters have currently established exclusion limit on the new particles mass or on the processes cross section according to the model chosen. For illustration in CMS studies, an exclusion at 95% C.L. has been reached for stop particles with a mass lower than $\sim 620\,\text{GeV}/c^2$ [60], for heavy t' particles with a mass lower than $\sim 450\,\text{GeV}/c^2$ [61] and for top-like particle T decaying totally in $T \to tZ$ with a mass lower than $417\,\text{GeV}/c^2$ [62].

1.2 Experimental observation of top quarks

This section first describes briefly the indirect evidence of the top quark and the experimental facilities where the top quark has been observed (with emphasis on the LHC), then summarizes the observation methods and the previous results. Finally, the experimental data harvesting and the simulation generation in the concrete case of my analysis are introduced.

1.2.1 Indirect evidences

The first direct observations of the top quark have been realized in 1995 at the TEVATRON and in 2010 at LHC. However, indirect evidences of its existence have been established before on a theoretical and experimental lines.

The possibility of a weak-isospin doublet of a third generation of quark has been first introduced in early 1970 in order to build a weak interaction model consistent with the CP violation observed in kaon decay [2].

The observation of the bottom quark in 1977 has re-enforced the assumptions on the existence of the top quark. Firstly, in a naive approach, it appeared natural to have a third generation of quark behaving as the two previous generations. Moreover, a third weak isospin doublet reestablished the symmetry between quarks and leptons. Secondly, the weak isospin doublet is also needed to apply the equivalent of the GIM

mechanism for the third generation and suppress the flavour-changing neutral current which is not experimentally observed. And thirdly, in a presence of third generation particles, the need of an isospin doublet was an elegant solution to cancel anomalous lepton loop diagrams. Indeed, these anomalous lepton loops are proportional to the product of the weak neutral current axial strength and the square of the electric charge, leading to a constraint on the properties of the fermions if an anomaly-free theory is wanted. Taking into account the three colour charges of the quarks, the minimal possible solution is the existence of a third generation weak-isospin doublet.

Several experimental observations have also indicated the presence of this quark. In 1983 in PETRA, the forward-backward asymmetry in bottom quark production [63], which is related to the axial coupling of the weak neutral current T_3 of the bottom quark, was favouring $T_{3b} = -\frac{1}{2}$ implying the existence of a weak isospin partner. In 1987, at ARGUS, DORIS and CESR experiments, the absence of $B_d^0 \bar{B}_d^0$ mixing has led to a lower mass limit on the top-quark mass $m_t \gtrsim 60\,\text{GeV}/c^2$ [64][65][66]. As stated before, the absence of observed flavour-changing neutral current, as observed in CDF and D0 in 1988-1993 [67], was also an indication in favour of the existence of the top quark. But the strongest indirect evidence has been obtained from the direct measurement of the bottom-antibottom production from Z^0 boson decay at LEP in 1994. In such analysis, the decay width of this process has been measured: $\Gamma(Z \to b\bar{b}) = 385 \pm 6\,\text{MeV}$ [5], while its theoretical expression is depending of T_{3b} [68]:

$$\Gamma(Z \to b\bar{b}) \propto (T_{3b} - Q_b \sin^2(\theta_W))^2 + (Q_b \sin^2(\theta_W))^2 \tag{1.9}$$

For a bottom quark singlet, $T_{3b} = 0$, this theoretical value is $24\,\text{MeV}$, when for a bottom quark doublet member, $T_{3b} = -\frac{1}{2}$, this value is $381\,\text{MeV}$. The observed value is clearly rejecting $T_{3b} = 0$ and compatible with $T_{3b} = -\frac{1}{2}$.

Additionally, as stated in Section 1.1.2, the top-quark mass parameter appears in precision electroweak measurements [36]. Combining these informations, a narrow top-quark mass region has been determined before its first observation. Such method can also be performed for the Higgs boson, but while the top quark mass dependence in the radiative correction is quadratic, the Higgs boson mass dependence is logarithmic, which leads to weaker constraints.

1.2.2 Hadron colliders

Due to its large mass, the top quark can only be observed in high energy collisions, currently only reached by two large synchrotron hadrons colliders: the Large Hadron Collider (LHC) and the TEVATRON. In these colliders, two beams composed of bunches of hadrons are accelerated in opposite direction in a circular facility containing notably

resonant cavities (for the acceleration) and multipoles (for beam curvature and beam collimation). The two beams cross at determined places in the accelerator ring where is located a detector.

The most relevant parameters in a particle collider are the type of particles used, the luminosity $\mathscr{L}$ and the beam energy. The type of particles and the beam energy determines the cross section of a given process, while the integrated luminosity $\mathcal{L} = \int \mathscr{L} dt$ allows the prediction of the number of expected occurrences of this process knowing its cross section. The instantaneous luminosity $\mathscr{L}$ is directly related to the beam intensities and densities, as expressed in the simplified equation Eq. 1.10 (neglecting the crossing angle effect and the variation of the σ and N parameters).

$$\mathscr{L} = f \frac{N_1 \, N_2}{4\pi \, \sigma_x \, \sigma_y} = f \frac{N_1 \, N_2}{4\sqrt{\epsilon_x \beta_x^* \epsilon_y \beta_y^*}} \tag{1.10}$$

where f is the frequency of the collision, N_i the number of particles in a bunch in the beam i, σ the transverse profile of the beam, ϵ the emittance (defined by $\epsilon = \pi \frac{\sigma^2}{\beta}$) and β^* the amplitude envelope function at interaction point.

Processes produced in hadron colliders are obtained from interactions of coloured partons carrying a fraction of their initial hadron momentum. With respect to lepton colliders, the amount of produced particles during the collision is larger, due to the coloured nature of the partons which leads to hadronization. Another complication with respect to lepton colliders is the unknown momentum of the parton, which cannot be determined due to the impossibility to measure the kinematics of the proton remnants. The total interaction cross section in hadron colliders, illustrated on Fig. 1.8, is dominated by elastic scattering, diffractive events and non-diffractive processes dominated by multijet QCD processes (respectively $\sim 35\%$, $\sim 10\%$ and $\sim 55\%$ at LHC). The more interesting processes have a much lower rate, e.g., $\sim 10^{-9}$ of the total cross section for $t\bar{t}$ at LHC. Beside the parton-parton interaction of a process of interest, the hadronization of the soft proton remnants can also interact and produce additional particles. There is also a possibility that more than one proton-proton interaction occur during the same bunch crossing or that a particle produced in a previous collision impacts the current event. These two effects, respectively called “underlying events” and “pile-up”, add background activities in the detector and should then be kept under control (multiple parton interaction can also happen). The high interaction rate is also unmanageable directly by the computing resources (in terms of processing time per event, and of storage). Therefore, an online selection is performed in order to process and store only the events showing interesting characteristics.

The **TEVATRON** is a circular proton-antiproton collider of 6.3 kilometres of circumference located at the Fermilab Laboratory, in Illinois, USA. Two general purpose

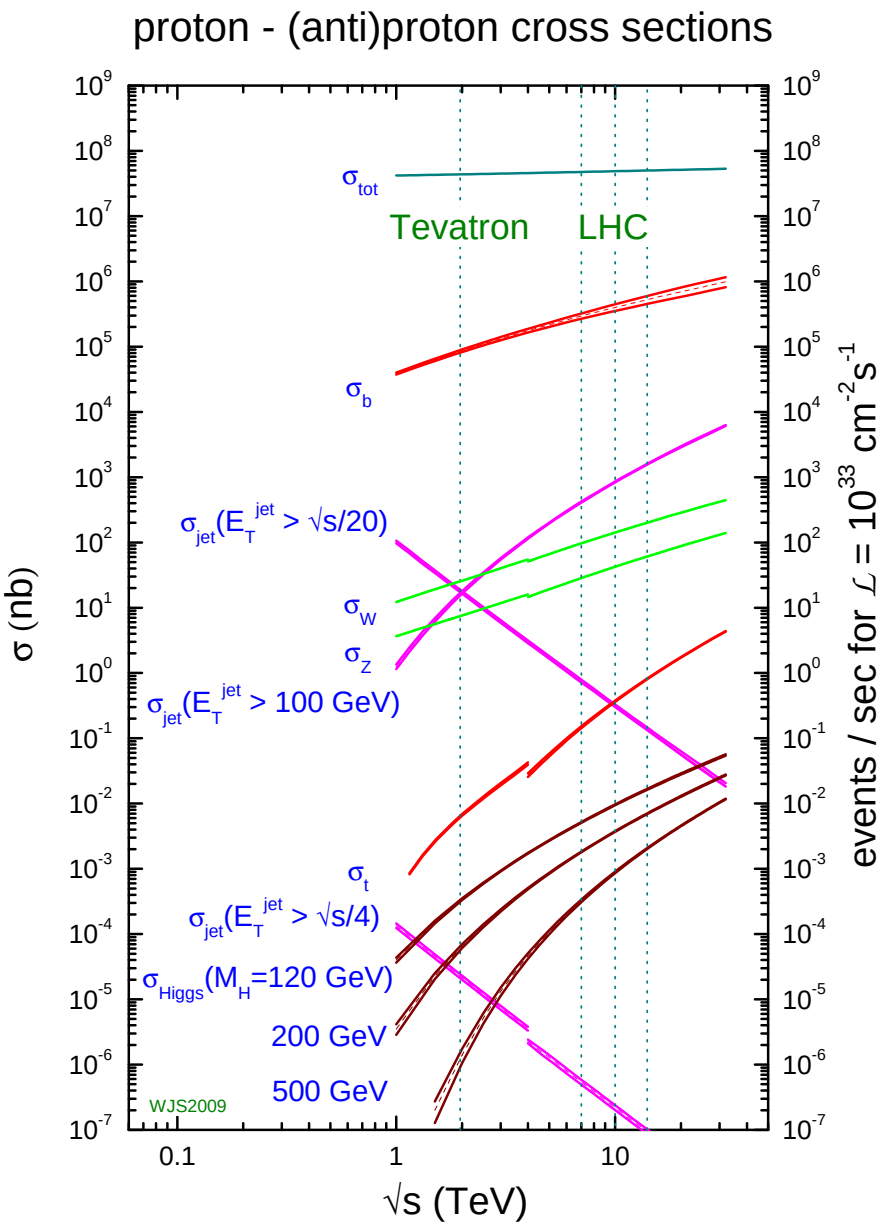


Figure 1.8: Production cross sections of several processes in proton-(anti)proton collisions at the TEVATRON and at LHC. From John Stirling.

detectors have been installed at the TEVATRON: CDF [69][70][71] and DØ [72][73]. A first phase of operation, called "Run I", has taken place between 1992 and 1996, colliding protons and antiprotons at a centre-of-mass energy of 1.8 TeV and providing 160 pb^{-1} of integrated luminosity (the maximum instantaneous luminosity reached during Run I is 3×10^{31} $\text{cm}^{-2}\text{s}^{-1}$, with $\sim 3.5\,\mu$s between each bunch of particles). After an upgrade of the accelerator and detectors, a second phase, called "Run II", has been ongoing from 2001 and is planned until September 2011, with a centre-of-mass energy of 1.96 TeV and of instantaneous luminosity (the maximum instantaneous luminosity reached during Run II is 5×10^{32} $\text{cm}^{-2}\text{s}^{-1}$). An upgrade of the tracking system of the DØ detector has been realized during a shutdown in 2006, and the data harvesting phases before and after this upgrade are respectively denoted as Run IIa and Run IIb. The Run II beams are composed of 36 bunches, containing 276×10^{9} particles in each proton bunch and 80×10^{9} particles in each antiproton bunch, and separated by 396 ns with a maximum interaction per crossing of ~ 3.5. By the end of September 2011, the TEVATRON will have provided approximately 12 fb^{-1} of integrated luminosity to each experiments.

The **Large Hadron Collider** (LHC) [74] is the particle accelerator built by the European Organization for Nuclear Research (CERN) near Geneva. It consists of a synchrotron of 26.7 kilometres of circumference under the border between France and Switzerland, in the tunnel built for the previous CERN collider, the Large Electron-Positron Collider (LEP). The construction started in 2000, the first beam has been injected the 10$^{\text{th}}$ of September 2008, the proton-proton collisions at $\sqrt{s} = 900$ GeV have started the 23$^{\text{rd}}$ of November 2009 and the proton-proton collisions at $\sqrt{s} = 7$ TeV have started the 30$^{\text{th}}$ of March 2010. During 2010, 47.03 pb^{-1} of integrated luminosity have been delivered and 43.17 pb^{-1} recorded by CMS (with 4% of uncertainty on these numbers), with 3.1 pb^{-1} certified before the September break (which corresponds to the data used in the analysis presented in this thesis). The LHC is currently (August 12$^{\text{th}}$ 2011) running at instantaneous luminosity of 2×10^{33} $\text{cm}^{-2}\text{s}^{-1}$, with 1380 bunches per beam, containing 1.25×10^{11} protons separated by 50 ns with an average number of interactions per crossing of ~ 7.

The collider, illustrated in Fig. 1.9, is divided in eight arc sections and eight straight sections constituting eight octants. The straight sections contain experiments and systems for the machine operation. Two proton beams are circulating in opposite direction into two vacuum chambers in multipole magnets at low temperature (1.9 K) and cross at small angle at four points of the collider, where are located the four big LHC experiments: ATLAS [75] (in the octant 1), ALICE [76] (in the octant 2), CMS [77] (in the octant 5) and LHCb [78] (in the octant 8). The other octants contain the two

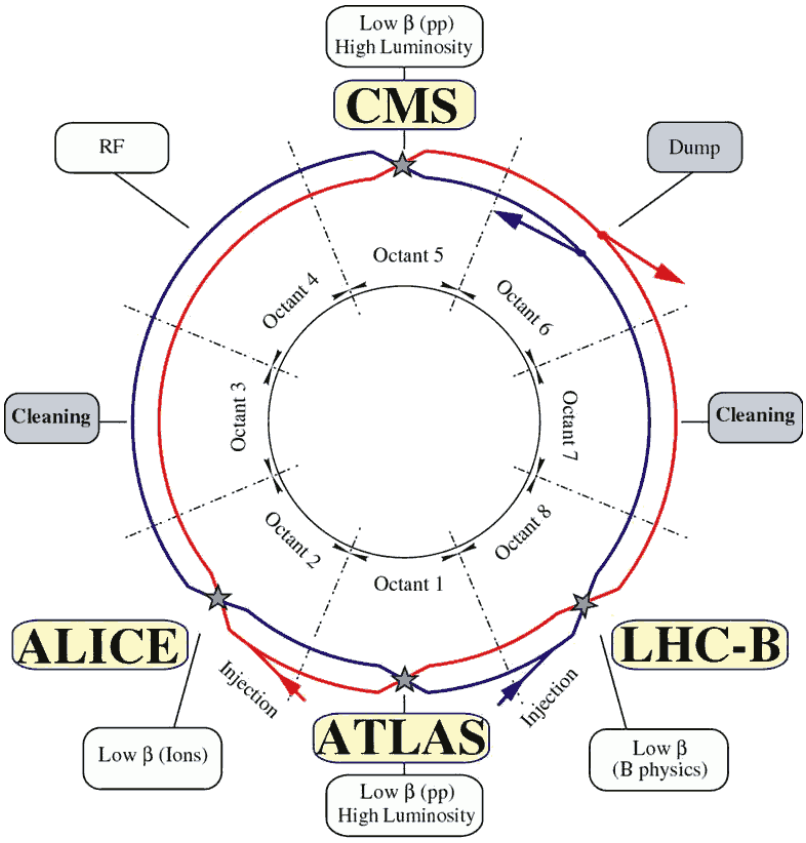


Figure 1.9: Schematic view of the Large Hadron Collider. From [74]

beam collimating systems (octant 3 and 7), the radio-frequency acceleration system (octant 4) and the beam dump system (octant 6).

1.2.3 Selection strategies and previous results

The first step in the selection of events of a given process is to ensure that these events will be kept by the online selection performed by the triggers. The outputs of each subdetector have then to be interpreted in order to reconstruct global objects such as the primary vertex of the interaction, the electrons, the muons, the jets and the missing transverse energy. The second step consists into applying a further selection based on these reconstructed objects knowing the topology of the studied process. At

this stage, usually called “pre-selection”, the selected sample contains events from the studied process, called “signal”, but also other events which are mimicking the studied process behaviour (due to processes producing a similar final states or due to the reconstruction limitations), called “background”. The pre-selection can be tuned to reject regions where the signal rate is negligible with respect to the background rate. The third step is the final analysis, which consists in applying specific methods to discriminate further the signal from the backgrounds or to extract information from the signal.

Among the top-quark production processes in hadron collision described in Section 1.1.1, the most visible channels appear to be the dileptonic $t\bar{t}$ channel ($\sim 10\,\text{pb}$) and the semileptonic $t\bar{t}$ channel ($\sim 54\,\text{pb}$). The single top processes are indeed penalized by their low cross sections, while the hadronic $t\bar{t}$ channel is overwhelmed by similar final state events produced by multijet QCD processes. Despite its relatively small branching ratio, the dileptonic channel constitutes a very clean subsample for the study of $t\bar{t}$ production. Indeed, the backgrounds producing energetic isolated lepton are dominantly coming from weak decays involving real or virtual intermediate vector bosons (W, Z), with negligible contamination from multijet QCD events. The event selection allows directly the measurement of the signal cross section. Indeed, the number of selected events after the rejection of the background processes corresponds to the product of the integrated luminosity, the cross section and the efficiency of the selection.

In the **dileptonic $t\bar{t}$ channel** case, the selection is first based on the presence of two energetic isolated leptons. The main backgrounds at this stage can be divided into two categories according to the origin of the leptons. They can be produced directly in the background process from weak interaction, in which case the momentum and the isolation (i.e., the absence of other particles produced colinearly) are similar to the signal leptons, or they can be resulting from a jet, a heavy quark decay or a particle incorrectly identified as a lepton, which are referred as “non-prompt lepton” in the following. The background processes of the first kind are the diboson processes, the single top tW and the Z^0 and continuum Drell-Yan production.

The diboson processes (noted VV), regrouping the $W^{\pm}W^{\mp}$, $W^{\pm}Z^0$ and the Z^0Z^0 processes and illustrated by Fig. 1.10, are dominated by the W^+W^- process. At the TEVATRON, it constitutes a non negligible background, due to its cross section similar to the $t\bar{t}$ ($\sigma_{W^+W^-} \sim 12\,\text{pb}$ at $\sqrt{s} = 1.96\,\text{TeV}$ [79], for $\sigma_{t\bar{t}} \sim 7.5\,\text{pb}$). At LHC, due to the increase of energy in the centre-of-mass system and the use of proton-proton collision, this background appears to be less dominant ($\sigma_{W^+W^-} \sim 43\,\text{pb}$ at $\sqrt{s} = 7\,\text{TeV}$ [80], for $\sigma_{t\bar{t}} \sim 160\,\text{pb}$). The WW process is usually discriminated with

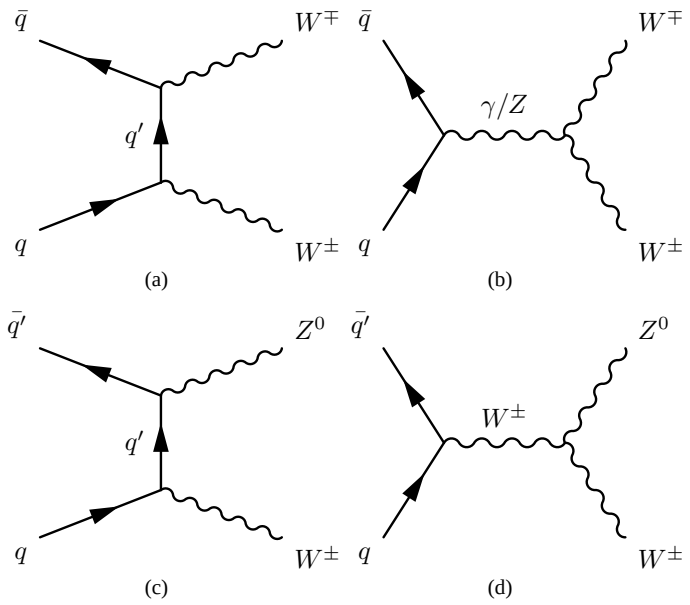


Figure 1.10: Main diboson process diagrams: t-channel WW (a), s-channel WW (b), t-channel WZ (c) and s-channel WZ (d).

respect to the $t\bar{t}$ by the absence of energetic jets. Other diboson processes producing two leptons from a Z^0 boson decay also can be rejected by the invariant mass of the two leptons and a low missing transverse energy.

A same flavour lepton-antilepton pair can be produced by electroweak interaction from a quark-antiquark annihilation via a virtual photon or a real or virtual Z^0. This process is called "Drell-Yan" and, for $m_{\ell\ell} > 50\,\text{GeV}/c^2$, is of the order of 300 pb at the TEVATRON and 3 000 pb at LHC ($\sqrt{s} = 7\,\text{TeV}$). This process is dominated by the real Z^0 boson decay (after a basic lepton selection in isolation and momentum, the real Z^0 boson peak can be observed in the dileptonic invariant mass, and $> 90\%$ of the events lies in a $\pm 15\,\text{GeV}/c^2$ window around the Z^0 boson mass). There are possibilities of jets production via flavour excitation or gluon splitting, but this is reduced (The ratio σ_{Z+2j}/σ_Z with a typical jet threshold is of the order of 2% at LHC [81]). Due to the leptonic family number conservation, a distinction is made between the Drell-Yan to two electrons or two muons ($DY \to \ell\ell$) and the Drell-Yan to two tau leptons ($DY \to \tau\tau$). The former contributes directly to the backgrounds of the ee and $\mu\mu$ dileptonic channels, while the later is the only possible Drell-Yan background contribution to the $e\mu$ channel via leptonic decays of one tau lepton in an electron and of the other tau lepton in a muon. The $\tau^+\tau^- \to e^{\pm}\mu^{\mp}$ branching ratio is $\sim 6\%$,

and the absence of large Drell-Yan background is the main reason why the $e\mu$ channel provides a better signal-over-background ratio than the ee and $\mu\mu$ channels. The Drell-Yan process can be discriminated from the signal by the absence of energetic jets and transverse energy imbalance. In the $DY \rightarrow \ell\ell$ case, the dileptonic invariant mass shows a peak corresponding to the Z^0 boson mass which can be used to discriminate this background. The Drell-Yan background is the dominant background in the ee and $\mu\mu$ channels and one of the dominant background in the $e\mu$ channel. For this reason, the contamination rate of this background is in general estimated from data.

The single top production can also contribute to the dileptonic background from the tW process. As described in Section 1.1.1, this process is negligible at the TEVATRON, but not at LHC. With respect to the $t\bar{t}$ production, the tW process produces only one jet from a bottom quark decay and has a lower cross section. Other single top channels are negligible in the dileptonic channel.

The rate of non-prompt leptons passing the lepton selection criteria is low (of the order of 10^{-4} for electrons and negligible for the muons). Background processes which need one (resp. two) non-prompt lepton to be selected have a selection efficiency of the order of the non-prompt lepton rate (resp. of the order of the non-prompt lepton rate squared), and can thus affect the analysis only in case of sufficiently high cross section. The contamination rate of these backgrounds depends strongly on the non-prompt lepton rate, which corresponds to tails of distribution particularly sensitive in simulation. Therefore, the contribution of these backgrounds is estimated from data. The background processes contributing via one or two non-prompt leptons are the semileptonic $t\bar{t}$, the W+Jets process and the multijet QCD processes.

The semileptonic part of the $t\bar{t}$ process is five time more abundant than the dileptonic part. This process produces one energetic isolated lepton, missing transverse energy and a high jet multiplicity, and one of the two bottom quarks can decay leptonically producing a second lepton. The probability that a lepton passing the momentum, isolation and identification criteria is a lepton from a bottom quark decay is however small ($\sim 10^{-3}$ for $t\bar{t}$ at LHC), which reduces significantly the contribution of this background.

The W+Jets process is the production of a real W boson associated with quarks or gluons, as illustrated in Fig. 1.11. This process, including the leptonic decay of the W boson, is of the order of 3 000 pb at the TEVATRON and 30 000 pb at LHC ($\sqrt{s} = 7$ TeV). The W+Jets process can be a background when the W boson decay leptonically accompanied by an additional non-prompt lepton. As in the case of the Drell-Yan, the cross section decreases when the jet multiplicity increases. Moreover, heavy flavour quarks can be produced by a gluon splitting (this case is called "W+ heavy flavour"), but again the cross section is much lower [82].

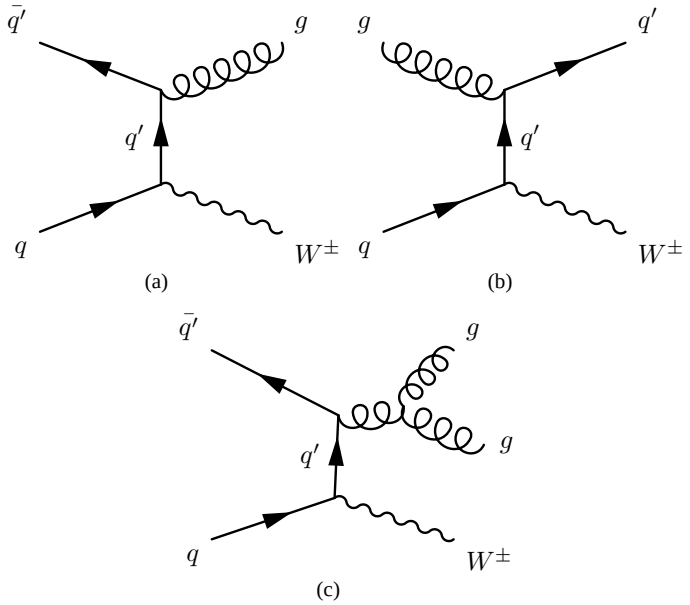


Figure 1.11: Three examples of W+Jets process diagrams.

The multijet QCD process, due to its large cross section, can also affect the dileptonic selection when two non-prompt leptons are selected.

After the leptonic selection, the dileptonic selection is based on the invariant mass between the two leptons (to reject Drell-Yan backgrounds), presence of missing transverse energy, presence of energetic jets and, if needed, presence of jets identified as produced by the decay of a bottom quark.

In the **semileptonic $t\bar{t}$ channel** case, the selection is based on the presence of one lepton, large jet multiplicity and missing transverse energy. The main backgrounds are the W+Jets and the multijet QCD processes described previously. In this case, the W+Jets process is not reduced by the need of one non-prompt lepton which leads to a lower signal over background ratio. Additional tools, like topological variables (such as jets aplanarity which describes the uniformity of the jets spatial distribution, and jets centrality which describes the tendency to produce low eta jets), kinematical variables (such the scalar sum of the transverse energy) and identification of jet produced by the decay of a bottom quark (called “b-tagging”), have to be used to discriminate the signal from the background.

At the **TEVATRON**, the first results at the CDF and the DØ experiments for the $t\bar{t}$ observation has been realized after $\sim 20\,\text{pb}^{-1}$ of integrated luminosity, using the dileptonic and the semileptonic channels. Due to the small luminosity, the statistical significance showed a clear preference for the existence of the $t\bar{t}$ process with 2.8σ statistical significance at CDF [83][84] and a 2.7% probability that a background fluctuation only can explain the observed excess in the events at DØ [85][86]. The 2nd March, 1995, TEVATRON announced [87] the discovery of the top quark, supported by a 4.8σ statistical significance at CDF [6] and a 4.6σ statistical significance at DØ [7]. At CDF, this limit has been reached with e+jets and μ+jets semileptonic channels using two b-tagging methods based on the presence of a displaced secondary vertex (SVX) or on the presence of a soft lepton from the bottom-quark decay (SLT) and with ee, $e\mu$ and $\mu\mu$ dileptonic channels. 6 dileptonic events have been selected for an expected background of 1.3 ± 0.3 and 37 semileptonic events have been selected, with 27 b-tags using the first method (SVX) for an expected background of 6.7 ± 2.1 and 23 b-tags using the second method (SLT) for an expected background of 15.4 ± 2.0. At DØ, this limit has been reached with ee, $e\mu$ and $\mu\mu$ dileptonic channels, with e+jets and μ+jets non b-tagged semileptonic channels using an aplanarity cut and with e+jets and μ+jets b-tagged semileptonic channels using a b-tagging method based on the presence of a soft muon from the bottom-quark decay. 17 events have been observed with an expected background of 3.8 ± 0.6.

Since 1995, DØ and CDF have provided many analyses concerning top quark physics (public top physics results for DØ: [88] and for CDF: [89]). The current main results for the top-quark properties and top-quark production were already discussed.

At **LHC**, the observation strategy is slightly different from the one used at the TEVATRON. The top quark production rate at early stage at LHC is about two order of magnitude higher than at the TEVATRON (due to the luminosity of the accelerator and the centre-of-mass energy). This implies that the $t\bar{t}$ process is visible during the early stage of the experiment, when the reliability of complex reconstruction methods, of subdetectors calibration and of data quality provided by the LHC have not been tested in real condition yet. A simple and robust selection is then preferred in order to understand the possible discrepancies from the expectations. Moreover, the $t\bar{t}$ events are complex events involving leptons, jets, secondary vertices and missing transverse energy, constituting an important step in the physics commissioning of the experiment. In addition, the top quark having already been observed, to reach a high statistical significance is less important than to provide a first milestone in the experiment start-up.

In this context, the first observations of $t\bar{t}$ at LHC have been performed by simple and robust selections. Different selection scenarios have been developed to allow observation even in case of malfunctioning of one specific subdetector or unavailability of more complex reconstruction method. This thesis describes the analyses proposed

for early data at CMS, which has led to the first publication of the evidence of the top quark at LHC the 28th October, 2010 in arXiv and the 17th January, 2011 in Physics Letter B [1].

1.2.4 Observed and simulated events at CMS

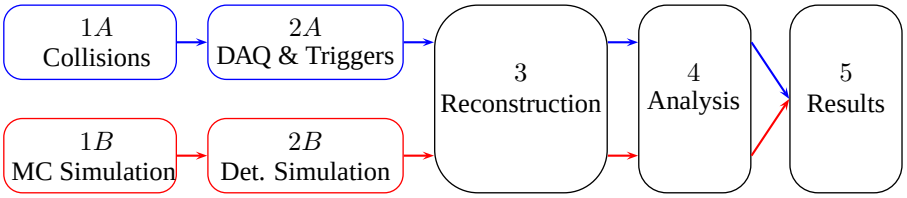


Figure 1.12: Workflow of the processing of the observed data and the simulation.

As described in Fig. 1.12, the infrastructure for the data treatment starts with the data acquisition during the collision. This is performed by the data acquisition system (DAQ) of the detector and the trigger system. The data are then saved on disk for offline reconstruction and further analysis. The results are finally compared with expectation of a given theoretical hypothesis to provide conclusions. This comparison with theory is based on simulated data. This simulation starts with the Monte Carlo (MC) generator, which generates the kinematics of the produced particles based on a given theoretical model. The process is usually computed at leading order and corrections are added using parton shower models. A matching procedure is performed to avoid a double counting between, for example, the $n + 1$ final partons at LO and the n final partons at LO + one radiated parton added by the parton shower procedure. Finally, the hadronization is also simulated to obtain the final colour-singlet particles which will interact with the detector. The interaction with the detector, the simulated response of the sensors and the hardware trigger decision are performed in the detector simulation step. At the end of this step, the simulated data format is totally similar to the format of the data from collision, with the exception of the additional generator information which can be used for performance and optimization studies. The same reconstruction algorithms are performed on simulated data and observed data in the reconstruction step, where the physical objects are deduced from the subdetector responses. An analysis is then performed on these reconstructed objects, optimized to extract information from the signal.

The **experimental data** (step $1A$ in Fig. 1.12) used in this analysis have been collected during the proton-proton collisions at 7 TeV energy in the centre-of-mass, when

the solenoidal magnetic field inside the detector was at the nominal value of 3.8 T. The recorded data are divided in runs (continuous period of data taking), and each runs is divided in luminosity blocks (section corresponding to a fixed number of beam orbiting period). In order to ensure that these data have been harvested during a proper configuration of all subdetector, only the runs and the luminosity blocks of data validated by the Data Quality Monitoring (DQM) group and the Physics Validation Team (PVT) have been used. This excludes events recorded during any commissioning run, when at least one subdetector was misbehaving or when trigger or calibration constants where not optimal. The integrated luminosity is based on information from the forward hadronic calorimeter [90], which leads to a total of $3.1\,\mathrm{pb}^{-1}$ of integrated luminosity. Additionally, the events recorded in bad conditions (events containing large amount of deposits due to beam halo particles, primary vertex reconstructed with a low confidence level, presence of significant noise in the hadronic calorimeter) are not proceeded to the analysis. More details on the data selection are available in [10].

The LHC is designed to collide proton bunches at a rate of 40 MHz. Moreover, a bunch crossing can provide several proton-proton interactions depending on the luminosity. Knowing that a typical event produces about 0.5 megabyte of data, and that relatively uninteresting interactions are dominant, the most interesting events are sorted and stored only if the event has been selected by the trigger system [91][92] (step $2A$ in Fig. 1.12). Due to the high rate of events, the decision has to be as fast as possible. This is achieved by several refinement levels applied from the less time consuming to the more complex reconstruction. The first trigger is the Level-1 Trigger (L1), which is based on programmable hardware assigned to the fast response subdetectors (calorimeters and muons chambers). At each collision, a basic and fast reconstruction is performed and a decision is made in less than 3.2 μs, rejecting about 99.75% of the events which reduces the event rate to less than 100 kHz. This decision is based on programmable thresholds on the reconstructed objects, but also on the readiness of the subdetectors and of the data acquisition system. The second trigger is the High Level Trigger (HLT), and is purely based on software on a computing farm. Thanks to the reduced frequency, a more complete topology reconstruction of the event can be performed by more sophisticated algorithms. The High Level Trigger is also internally structured with several layers, each layer refining the objects reconstructed in the previous layer. At the end of the High Level Trigger, an event is accepted if it has successfully satisfied the successive filters of at least one trigger path. The High Level Trigger reduces the event rate to a few 100 Hz which are stored definitively on tape for an offline reconstruction. Collected data samples are organized in skimmed datasets (“primary datasets”), regrouping events passing a given set of triggers aggregated according to the type of objects that fired the trigger. The two skims used in this analysis are based on the presence of a muon object or an electron/photon object, with

thresholds looser than the selection which is applied on the analysis selection. The events contained in both skims are treated consistently.

The trigger efficiency has to be taken into account to know the proportion of signal events that escape the selection. For the leptonic trigger, which are the ones used in this analysis, the efficiencies can be evaluated directly in the simulation and in data by the "tag and probe" method [93] in real data. The principle of the tag and probe method consists in a comparison of two selection efficiencies in a sample dominated by real dilepton events (the selection of this samples should of course not be biased with respect to the evaluated trigger). In this study, this sample is constituted of the events passing lepton identification and isolation requirements defined in the analysis (cf. Section 2.3.3 and 2.4.3) and leading to a dileptonic invariant mass close to the Z^0 mass peak ($76\,\text{GeV}/c^2 < M_{\ell\ell} < 106\,\text{GeV}/c^2$), which is dominated by Drell-Yan events. Three selections are defined: a strong tag selection, a looser or identical probe selection and a looser base selection. Requiring at least one lepton passing the tag selection, there are three categories of events: the second lepton passes the tag selection (T events), the second lepton passes the base selection and the probe selection but not the tag selection (P event), and the second lepton passes the base selection, but not the probe condition (F event). In this case, the efficiency ϵ of the probe selection is evaluated by:

$$\epsilon = \frac{2N_T + N_P}{2N_T + N_P + N_F} \tag{1.11}$$

where N_x is the number of events in the x category. Finally, from the single lepton efficiency ϵ, the selection efficiency for events containing two leptons can be obtained from $(1-(1-\epsilon)^2)$ under the reasonable assumption that the two lepton efficiencies are uncorrelated. This method can be validated at simulation level, with different process to ensure that the kinematic of the process has a negligible effect.

In this study, the trigger efficiency for events with two muons has been evaluated to $98.0 \pm 0.5\%$ in simulation and $97.7 \pm 0.3\%$ in data, mainly reduced by a low efficiency for muons with $|\eta| > 2.1$. Based on comparison between $t\bar{t}$ and Drell-Yan simulations, it appears that the trigger efficiency has to be corrected by an increase of 1%. The single electron efficiency appears to be really high, leading to a trigger efficiency for dielectron events higher than 99.9%. The single electron trigger efficiency dominates also for the $e\mu$ channel, where the dileptonic events efficiency is higher than 99.7%.

The CMS experiment uses dedicated **software** to perform the simulation of the detector (step $2B$ in Fig. 1.12), the reconstruction (step 3 in Fig. 1.12) and analysis of events (step 4 in Fig. 1.12).The software framework of CMS is called CMSSW. This framework uses GEANT4 [94] in the detector simulation. An additional abstraction

layer, The Physics Analysis Toolkit (PAT) [95], has been used for this analysis. The details of the version of CMSSW, of the additional packages used, and of the physical tuning based on observed condition of the detector is available in [10].

Concerning the **simulated samples** (referred as "Monte Carlo samples" or "MC samples"), several processes have been considered (step $1B$ in Fig. 1.12). These samples have been simulated with MADGRAPH [96] (version 4.4.12) or PYTHIA [97] (version 6.420). The list of these samples with their associated cross sections are listed in Tab. 1.3. The technical details are available in [10]. In order to compare the simulation and the data, each dataset is scaled to correspond to the amount of predicted events for the given integrated luminosity. The number of available events in each dataset is large enough to provide a small statistical uncertainty for the simulation with respect to the data, except for multijet QCD processes, that are constrained by data-driven estimation techniques anyway (Section 3.2). Minor corrections, as flavour content of the W/Z+jets samples and overlap between tW and $t\bar{t}$, have been found negligible at this stage and are then neglected.

Table 1.3: Monte Carlo samples used in this study.

Process	Description	Generator	$\sigma \times BR$
`TTbarJets`	$t\bar{t}$	MadGraph	157.5 pb
`WJets`	$W + jets\,(W \to l\nu)$	MadGraph	31 314 pb
`ZJets`	$Z + jets\,(\gamma^*/Z \to ll, M_{ll} > 50$ GeV)	MadGraph	3 048 pb
`DYee_M10to20`	$Z/\gamma* \to e^+e^-, (M_{ll} < 20$ GeV)	Pythia	2 659 pb
`DYmumu_M10to20`	$Z/\gamma* \to \mu^+\mu^-, (M_{ll} < 20$ GeV)	Pythia	2 659 pb
`Zee`	$Z/\gamma* \to e^+e^-, (M_{ll} > 20$ GeV)	Pythia	1 665 pb
`Zmumu`	$Z/\gamma* \to \mu^+\mu^-, (M_{ll} > 20$ GeV)	Pythia	1 665 pb
`Ztautau`	$Z/\gamma* \to \tau^+\tau^-, (M_{ll} > 20$ GeV)	Pythia	1 665 pb
`ST_tWChannel`	tW	MadGraph	10.6 pb
`VVJets`	$VV + jets\,(V = W, Z)$	MadGraph	4.8 pb
`QCD_Pt15`	QCD ($15 < \hat{p}_T < 30$ GeV)	Pythia	811.8 μb
`QCD_Pt30`	QCD ($30 < \hat{p}_T < 80$ GeV)	Pythia	59.49 μb
`QCD_Pt80`	QCD ($80 < \hat{p}_T < 170$ GeV)	Pythia	989.3 nb
`QCD_Pt170`	QCD ($\hat{p}_T > 170$ GeV)	Pythia	25.47 nb

Chapter 2

Event Detection and Reconstruction in CMS

This chapter describes the characteristics of the CMS detector, the methods used to obtain a reconstructed physical object from raw data output from subdetectors and performances predicted and observed with the first data.

This chapter starts with an overall description of the CMS detector and then describes each physics object used in the analysis: the tracks, the muons, the electrons and the jets. Each of these sections contains a description of the main subdetector dedicated to the object reconstruction, the reconstruction procedure and its performances observed in simulation and data (especially for early data, $\mathcal{L} \leq 3.1\,\text{pb}^{-1}$, corresponding to the knowledge at the time of the first $t\bar{t}$ observation).

2.1 CMS overview

The Compact Muon Solenoid [77] is a cylindrical detector of 12 500 tonnes, 21.5 m in length and 15 m of diameter centred on the interaction point (the origin of the system of coordinates) and oriented along the beam axis. To describe the position and direction of a particle inside the detector, several systems of coordinates can be used:

- The Cartesian system is composed of the x axis oriented in the direction of the centre of the LHC ring, the y axis pointing to the top and the z axis completing the right-handed oriented coordinate system;

- The cylindrical coordinates are based on the z axis, on the r coordinate which is the distance in the x-y plane and the ϕ angle with respect to the x axis in the x-y plane ($\pi/2$ corresponding to the y axis);

- Finally, a last system of coordinates can be defined based on the pseudorapidity η of a particle. This direction can be geometrically defined by the formula $\eta = -\ln[\tan(\theta/2)]$ where θ is the angle with respect to the z axis. The pseudorapidity can also be interpreted physically by the formula $\eta = 1/2[(|\vec{p}| + p_z)/(|\vec{p}| - p_z)]$ where $\vec{p}$ is the momentum of the particle and p_z is its z component. This coordinate is related to the rapidity y of the particle, of which the pseudorapidity is a good approximation in the relativistic limit. In this framework, the position of a point is given by its R coordinate (the distance with respect to the origin of the system), its ϕ coordinate and its pseudorapidity η. The distance ΔR in the η-ϕ plane is then defined by $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

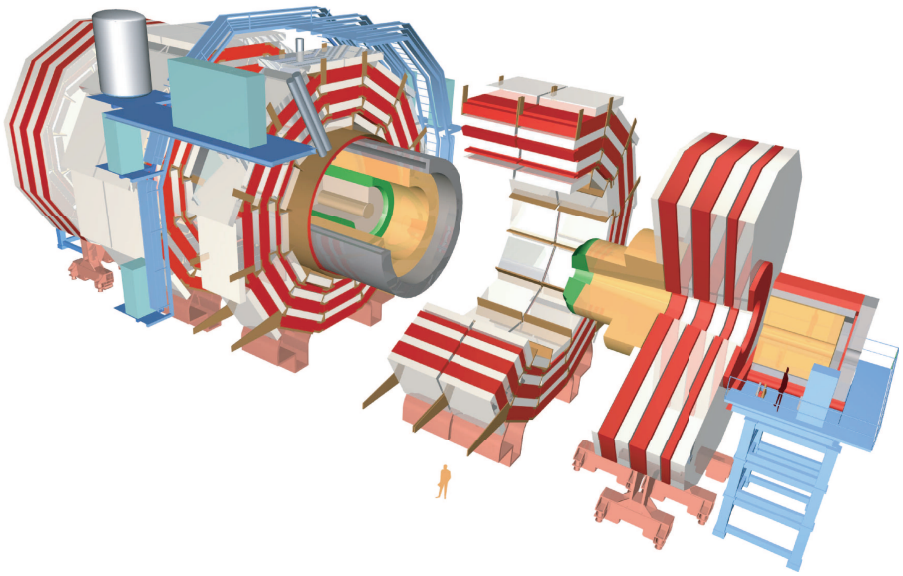


Figure 2.1: Overview of the CMS detector [77]. From inside to outside: the pixel detector (light brown), the tracker (light pink), the electromagnetic calorimeter (green), the hadronic calorimeter (orange), the magnetic solenoid (grey) and the muon chambers (white) and the magnet yoke structure (red).

The CMS detector is composed of multiple subdetectors, as illustrated by Fig. 2.1. They are disposed in successive layers on the cylindrical structure of the detector. In

this context, a subdetector is in general divided in a “Barrel”, the cylindrical part centred on the beam axis, and in two “Endcaps”, the two disks perpendicular to the beam axis closing the barrel. From the inner subdetector to the outer, there are the tracker (cf. Section 2.2.1), the electromagnetic calorimeter (cf. Section 2.4.1), the hadronic calorimeter (cf. Section 2.5.1) and the muon chambers (cf. Section 2.3.1) separated from the calorimeters by the solenoid. Other apparatus, like the hadron forward detector ($|\eta| < 5$), are also placed in the forward region to recover data close to the beam.

The CMS magnet system induces a 3.8 Tesla magnetic field parallel to the beam axis. The charged particles are then bent in the transverse plane, and the transverse momentum of a charged particle can be obtained by the sagitta measurement: $p_T \simeq 0.3BL^2/8s$ where p_T is the transverse momentum in GeV/c, B the magnetic field in Tesla, L the distance between the two extremities of the arc in metre and s the sagitta in metre. This magnetic field is generated by a superconducting coil of Niobium-Titanium, 13 m long and of 5.9 m of diameter. The coil is cooled to 1.9 K by an external cryogenic system using liquid helium. The magnetic flux returns in a saturated iron yoke that is also used as supporting structure for the muon chambers.

2.2 Reconstruction of charged particle tracks

When particles are created, they first pass through the tracker subdetector. The charged particles are detected by the microship sensors. The trajectories of the charged particles can therefore be extrapolated from the set of hits, and the reconstructed tracks can be used as a basis for more complex reconstructions. This section first describes the tracker subdetector, then the track reconstruction algorithm and finally the performance of the tracker and the reconstruction with the first data.

2.2.1 Tracker subdetector

The conception of the tracker subdetector [99][100][98] was very challenging. Firstly, this subdetector is close to the beam pipe and should be effective during the operating time of the LHC. Secondly, a high precision on the position of the hits in the tracker is essential to a precise reconstruction of the trajectories (and then the transverse momenta) and for the determination of the vertex or impact parameters, which implies a high granularity. Additionally, fast response is needed to avoid overlap between signals from consecutive events. But a high density of high power detectors needs complex cooling, increasing the density of material, which increases the probability

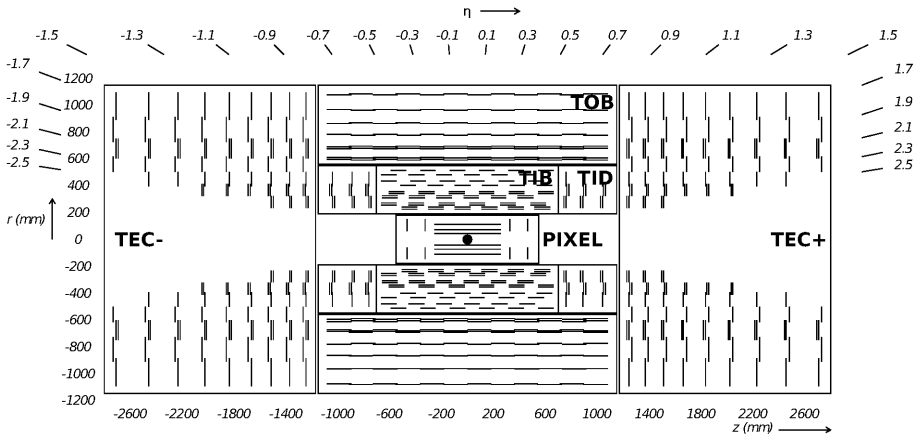


Figure 2.2: Schematic view of the CMS tracker [98], in the z-y plane.

of interaction between the particles from the event and the material (bremsstrahlung radiations, photon conversions, nuclear reactions). Two systems were chosen to perform the tracking. The first one is a pixel detector, the closest detector to the interaction point, composed of pixel sensors to enhance the precision in the reconstruction of vertices. The second one is a silicon strip detector, composed of silicon microstrip sensors on several layers to provide a good track reconstruction. The total tracker is 560 cm in length and has a diameter of 220 cm, Fig. 2.2.

The pixel detector barrel (BPix) is composed of three cylindrical layers of 98 cm in length with radii of 4.3, 7.3 and 10.2 cm and is covered with 768 hybrid pixel detector modules (48 millions of pixels). The pixel detector endcaps (or Forward Pixel) (FPix) are composed of two disks of pixel modules from 4.8 to 14.4 cm in radius, at 35.5 and 48.5 cm from the central point, covered with 672 pixel detector modules (18 millions of pixels). This allows a pseudorapidity coverage reaching $|\eta| = 2.5$, and a spatial resolution of 15-20 μm (the track reconstruction enhance the resolution up to $\sim 10\,\mu$m).

The silicon strip detector is composed of a tracker inner barrel (TIB) and two tracker inner endcaps (TID) surrounding the pixel detector, a tracker outer barrel (TOB) surrounding the TIB and two tracker endcaps (TEC) closing the cylinder.

The TIB extends radially from 20 cm to 55 cm and is 140 cm long. It contains 4 layers of silicon microstrip sensors with strips parallel to the beam axis. The r-ϕ resolution is 23 μm for the two first layers and 35 μm for the two last layers. The two first layers

are double sided with microstrip detectors tilted by 100 mrad, allowing a resolution of 230 μm in the z direction.

The two TID close the TIB cylinder from 80 cm to 90 cm in the z axis. Their inner and outer radii are respectively 20 cm and 50 cm and they contain 3 disks of silicon microstrip sensors with strips radial to the beam axis. The first two disks are also double sided with shifted silicon microstrip. Thanks to the TID, the acceptance reaches $|\eta| = 2.5$.

The TOB surrounds the TIB and the TID. It radially extends from 55.5 cm to 116 cm and is 218 cm long (236 cm with the cooling system). It is composed of 6 layers of silicon microstrip sensors parallel to the beam axis. The r-ϕ resolution is 53 μm in the four first layers and 35 μm in the two last layers. The two first layers are double sided with microstrip detectors tilted by 100 mrad, allowing a resolution of 530 μm in the z axis.

The two TEC finally close the tracker from 124 cm to 280 cm on the z axis. They are each composed of 9 disks which extend radially up to 113.5 cm. The minimal inner radius of these disks is 22.9 cm. Each disk is divided in 16 petals and can contain up to seven rings of silicon microstrip detectors (the ring 1 is present in the three first disks only, the ring 2 disappears after the sixth disk, and the last disk only contains rings 4 to 7). The rings 1, 2 and 5 are double sided with additional modules tilted by 100 mrad.

2.2.2 Track reconstruction

Each tracker module yields a signal when crossed by a charged particle. This signal is clustered in "hits", i.e., mainly a charge, a position and its uncertainty. The particle trajectory can then be reconstructed as a "track", i.e., a sequence of hits. These trajectories point to the interaction point where these particles have been produced. Therefore, the vertex of an interaction can be reconstructed as the point regrouping several tracks origins.

The track reconstruction process [101] is iterative: a first run is applied using stringent parameters on the different reconstruction steps, leading to primary tracks. The hits composing these tracks are then excluded from the hits collection taken into account during the reconstruction and a second run is applied with different parameters, dedicated to secondary tracks, tracks from conversions, etc. For the standard track reconstruction process described in this section, called "Combinatorial Track Finder" (CTF), a total of six iterations is applied (the first iteration leads to tracks flagged as "high purity"). Each iteration of the CTF can be divided in four steps. The first step

is the seed generation, where the initial position, direction and momentum (and uncertainties on these values) are defined. The second step is the pattern recognition, trying to extrapolate iteratively the position of the next hit according to an helix trajectory based on the previous trajectory state starting with the seed. When the track candidates are created, a final track fit is applied. Finally, a track selection is applied to select only the reliable tracks and the avoid hits double-counting before the next iteration with different constraints on each step.

On the first step, the seed is built out of three points generally located in the pixel detector, which has a high granularity and is able to provide three-dimensional measurements. The reconstruction is then performed moving inside-out. 90% of the particles leads to three hits in the pixel system, however, combinations of pixel hits, beam-spot or vertex constraints and strip hits are used to increase the efficiency of track reconstruction and to take into account particles not coming directly from the colliding point.

The second step, the pattern recognition, is based on the Kalman filter method. The Kalman filter starts with the seed and try to find the next hits in the following layer. A search window is defined on the next layer given the trajectory state parameters and their uncertainties, fitted with an helix trajectory. The uncertainties matrix assumes Gaussian errors, which is a reliable approximation for the majority of the tracks, but not for electrons which suffer from bremsstrahlung radiation leading to highly non Gaussian energy loss. For each compatible sensor available in the acceptance window, the expected hit is evaluated given the trajectory direction with respect to the surface. The distance between the reconstructed and the expected hit is called the "residual", and is used in the computation of the chi-square of the track candidate. For n compatible measurement, $n + 1$ new track candidates are created (the n compatible hits plus one with the hypothesis of an undetected passage through one sensor, called "invalid hit") and the finding procedure is iterated with these new track candidates. To avoid time-consuming computations, the number of parallel trajectory states is limited to a given number, the less compatible ones being dropped (based on normalized chi-square when fitting the measurement hits on the extrapolated trajectory and on the number of valid and invalid hits). The iteration is stopped when the last layer is reached or when a stopping condition (i.e., when the reconstruction is sufficiently reliable for the given purpose) is satisfied. Finally, an ambiguity resolution procedure is then applied to conserve just one track candidate among the multiple track candidates sharing partially the same hits.

In the next step, a more correct trajectory estimation is obtained by applying a final fit, re-evaluating the uncertainty based on the track parameters of the reconstructed track. All the expected hits are first re-evaluated (this is done without taking into account the vertex or the beam-spot), then a second filter initialized from the first one is applied

backwards to smooth again the result. Some reconstructed hits in the track candidate can be affected by other effects (presence of another close track, electronic noise, ...), and this can be visible via the chi-square after the last smoothing. These spurious hits with residual above the threshold are then rejected and replaced by an invalid hit, the track candidate is refitted, and the procedure is repeated until all the residuals are below the threshold.

Finally, the last step consists of applying a track selection. The track candidate rejected by this selection are dropped and all their hits are available in the next reconstruction iteration. The selection cuts are different at each iteration, depending on the transverse momentum, the pseudorapidity and the number of layers crossed by the track-candidate (knowing that non physical tracks decrease exponentially with the number of crossed layers, the tracks crossing more than 10 layers do not really need additional selection, the cuts thresholds are then very loose) and are based on several variables: the normalized chi-square, the transverse distance between the beam-spot and the point of closest approach, the longitudinal distance between the HLT primary vertex and the point of closest approach, and the relative uncertainties of these two last variables.

The primary vertex reconstruction starts with the vertex finding. A selected set of reliable tracks are regrouped according to the z coordinate of their point of closest approach with respect to the beam line. Clusters of tracks separated from other clusters by a given distance z_{sep} on the z axis constitutes vertex candidates. For the data used in this study, this z_{sep} is set at 1 cm, but has afterwards been reduced for better performance in presence of the pile-up (in presence of two vertices, the merging probability has been found to be 23% for $z_{sep} = 1$ cm [102]). The vertex fitting is then applied, using the Adaptive Vertex Fitter [103], on the vertex candidates, to weight each contributing tracks according their compatibility with the vertex. The vertex fitting allows the determination of the best estimate for the vertex position and its covariance matrix and the evaluation of the number of degree of freedom, based on the weights of the constituent tracks. From these variables, a subset of reliable vertices are defined. An usual definition of good quality vertices in early data, as it is used in this thesis, is based on a number of degrees of freedom higher than four, a distance in the transverse plane smaller than 2 cm and a longitudinal position in the z axis between -24 cm and 24 cm.

2.2.3 Performance of track reconstruction

The performance of the track reconstruction algorithm for a track produced inside the CMS detector depends first on the acceptance of the detector, i.e., the probabil-

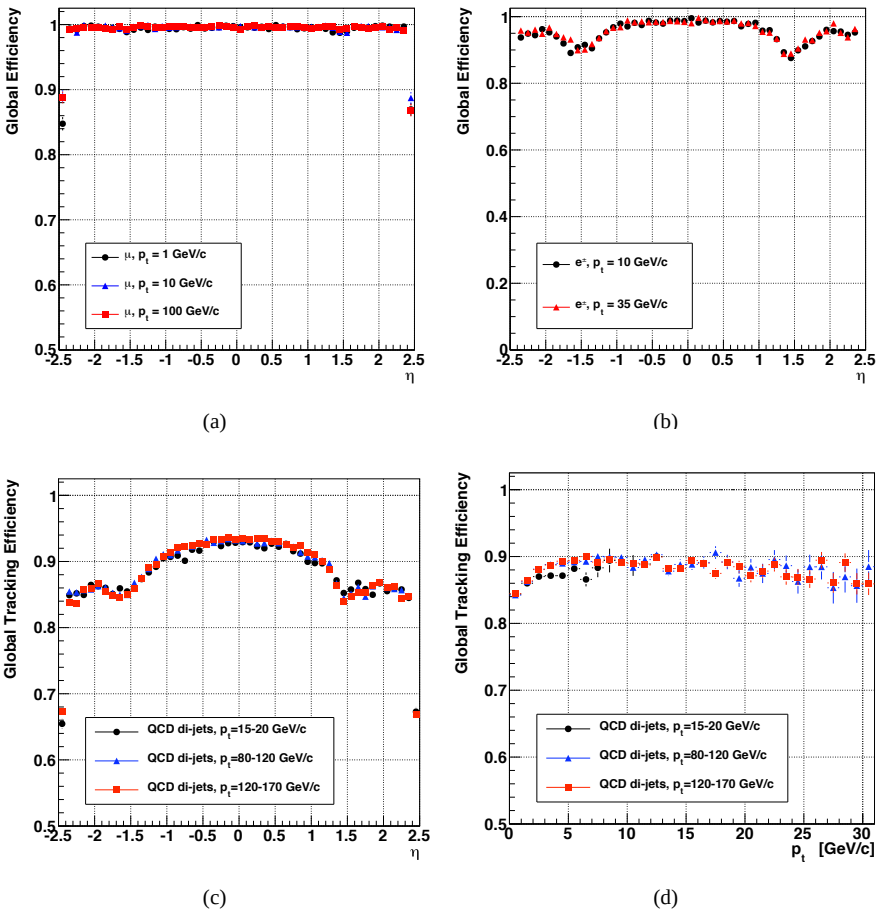


Figure 2.3: Global reconstruction efficiencies in simulation, for isolated muon tracks (a), isolated electron tracks (b) and all tracks in dijet events (c) with respect to η, and for all tracks in dijet events with respect to p_T (d). From [101].

ity that a charged particle produces a sufficient number of hits in the tracker to be reconstructed by the track finding algorithm. If a track satisfies this requirement, it will be reconstructed if its hits are used to reconstruct a track with parameters representative of those of the physical particle. This latter probability is called “algorithm efficiency”. The total probability for a track being reconstructed, combining the acceptance and the algorithm efficiency, is called “global efficiency”. These efficiencies can be evaluated in simulation, for isolated particle tracks or tracks in dijet events, by comparing simulated tracks and reconstructed tracks (two tracks are matched if they

share 75% of the hits). As visible in Fig. 2.3, the global efficiency is high for isolated muons and electrons, and reaches 90% for $p_T > 5\,\text{GeV}/c$ tracks in dijet events. The global efficiency can also be evaluated from data for specific scenario. Two methods have been developed [104]: the track-embedding method which consists of incorporating simulated hits in data and to check if the reconstruction leads to the initial track, and the tag and probe method exploiting the dimuon resonances and using a reconstructed track as a tag and a muon reconstructed only in muon chambers as a probe. The results are compatible with the simulation within one percent, and an uncertainty on the tracking efficiency for isolated muons is evaluated to 1-2%.

The parameters of the reconstructed tracks play an important role in physics analysis because they are used as discriminant to discern the physical origin of the track. The validity of these parameters can be tested by checking the discrepancies between simulations and data [98]. Other reconstructions, such the invariant mass of mesons (K^0_s, $D^\star$, J/Ψ mesons), can be used to evaluate the momentum scale to apply in data. As visible on Fig. 2.4 for the p_T, η and the significance of the point of closest approach ($\sigma(d_0)$), the comparison between simulation and data shows good agreement when the parameters of the Monte-Carlo generator are correctly tuned. For the J/Ψ invariant mass reconstruction also visible in Fig. 2.4, the deviation between the expected momentum of the muon tracks and the results from the fit has been found to be of the order of the per mille, and the resolution on muon transverse momentum is been found to be within 5% of the predictions of the simulation.

The primary vertex resolution and efficiency can be evaluated from data using the "split method", which consists of splitting the set of tracks used in the vertex reconstruction and to compare the reconstructed vertex obtained independently from each set. For the resolution, the two sets contain half of the initial set with similar kinematic distribution. The resolution can be extracted from the distribution of the differences on the fitted vertex positions between the vertices reconstructed from the two sets, for a given number of tracks. The resolution for early data for vertices reconstructed from more than 30 tracks is 20 μm in the longitudinal axis and 25 μm in each transverse axes. This resolution is enhanced if the mean transverse momentum of the tracks used in the vertex reconstruction increase. For the efficiency, one of the set (probe set) contains 1/3 of the initial set while the other (tag set) contains 2/3 (the tag set is therefore more reliable than the probe set). The reconstructed vertices from the two sets are matched together according to the resolution of the vertices and the efficiency is extracted from the number of tag reconstructed vertices for which no matching probe reconstructed vertex can be found. For primary vertices with at least two tracks with $p_T > 0.5\,\text{GeV}/c$, the efficiency is estimated to be close to 100%.

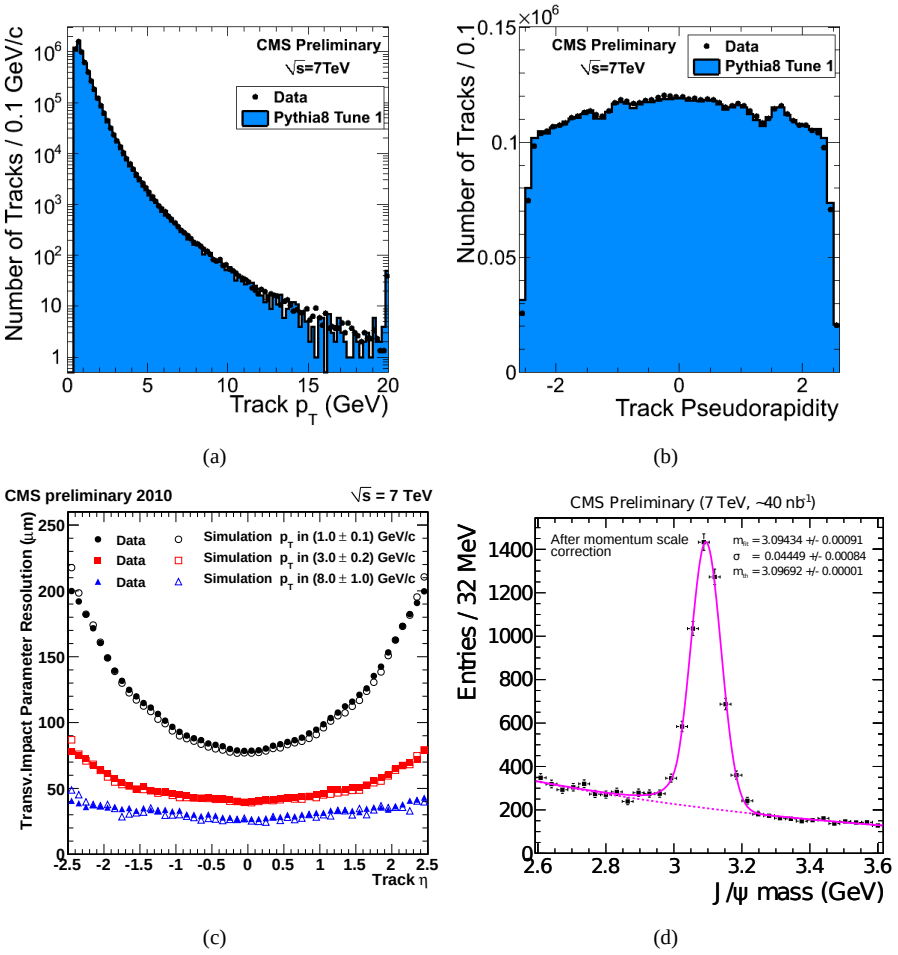


Figure 2.4: Comparison of data with simulation or standard candle. Track parameters such as p_T (a), η (b) and σd_0 (c) (in minimum bias events after $\sim 11\,\mathrm{nb}^{-1}$ of integrated luminosity) are compared with simulation. The J/Ψ meson mass reconstructed from the corrected tracks of the two produced muons can be fitted to evaluate the momentum scale (d) (in a dimuon sample triggered by a low threshold muon trigger after $\sim$40 nb^{-1} of integrated luminosity). From [102] (a, b and c) and [105] (d).

2.3 Reconstruction and identification of muons

Muons are produced in several processes of interest. The subdetectors dedicated to their detection also allow a fast reconstruction and identification for the trigger system.

The muons are detected in dedicated gas-ionization detectors located outside the solenoid. The reconstruction is complemented by information from the tracker or calorimeter subdetectors. This section describes the muon detection systems (muon chambers), the muon reconstruction and finally the observed performance of the muon detection in CMS for early data.

2.3.1 Muon chambers subdetector

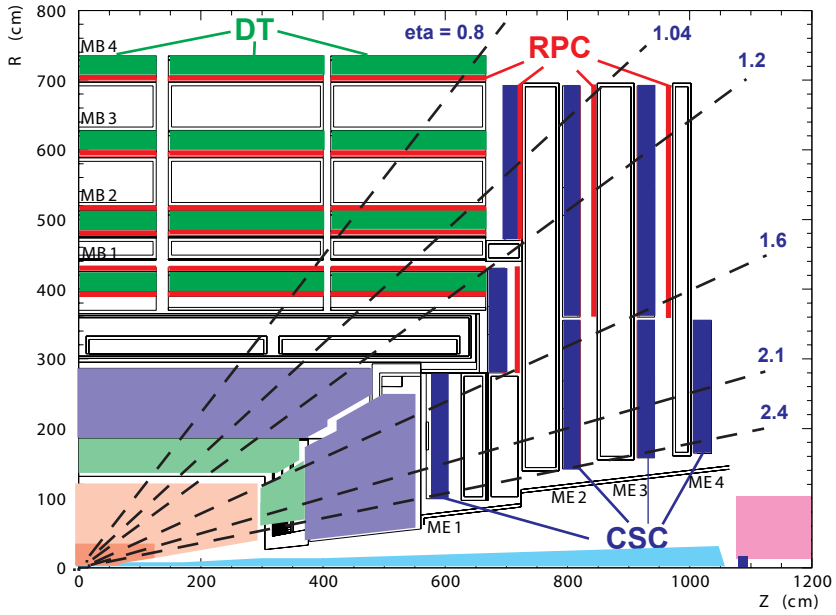


Figure 2.5: Schematic view of the muon system [106], in a quarter of the z-y plane.

The muon chambers [106], illustrated in Fig. 2.5, are composed of four layers of Drift Tube (DT) chambers installed in the barrel and Cathode Strip Chambers (CSC) installed in the endcaps, both complemented with Resistive Plate Chambers (RPC). These layers are separated with steel absorbers which are also used as return yoke for the magnetic field.

The drift tube chambers are distributed on 5 wheels of 12 sectors in 4 successive layers up to $|\eta| < 1.2$ for the first layer. Each chambers contains 60 drift tubes and they provide a spatial resolution of 100 μm for the position measurement and a resolution of 1 mrad on the direction. In the endcaps, where the magnetic field is less uniform and

the muon rate is higher, cathode strip chambers are used. 468 multiwire proportional chambers are arranged in four layers transverse to the beam axis, shifted in ϕ to avoid gap in acceptance. The chambers are trapezoidal section of the discs containing 7 layers of cathode strip panels oriented in radial direction and 6 layers of anode wire panels perpendicular to the strips. The endcaps cover an η region between 0.9 and 2.4. They provide a spatial resolution of 200 μm for the position measurement and a resolution of the order of 10 mrad for the direction. Additionally, resistive plate chambers have been disposed in front of all the drift tube chambers and in the back of the drift tube chambers for the two first layers, and in the back of the three first cathode strip chambers layer until $|\eta| < 1.6$. These fast and robust sensors are principally used for the triggering or to provide complementary information for the other muon detectors.

2.3.2 Muon reconstruction

The muon reconstruction [107] can be divided in three algorithms: the one using only the muon chambers information ("Standalone Muon"), the one combining muon chambers information and tracker information ("Global Muon") and the algorithm using mainly the tracker information complemented by some calorimeter and muon chambers information ("Tracker Muon").

The standalone reconstruction algorithm starts with the reconstruction of the track segments inside each chambers. Because the magnetic field is concentrated in the yoke, the hits in the chamber layers are fitted with a linear algorithm on the aligned hits taking into account their resolution. The best-fitted segment, based on the number of hits and the chi-square, are kept, and the possible hits double-counting are solved. This reconstruction takes also into account the fact that the muon is supposed to be originally from the centre of the detector. From the estimated parameters of the initial track segment, Kalman filter extrapolation to the segment (in DT) or hits (in CSC) of the outer chambers can be performed. For this extrapolation, the muon scattering and energy loss in the absorbers and the effect of the magnetic field is taken into account. As the track reconstruction in the tracker, the possibility of missing layer is taken into account, and when the outer layer is reached, the filter is reapplied backwards to refine the track reconstruction. Only tracks fitted with a sufficiently small chi-square and compatible with the nominal interaction point are kept.

The global reconstruction algorithm is based on the standalone muon track reconstructed as described previously. From this track, a rectangular η-ϕ region of interest in the tracker is defined, based on the reconstructed primary vertex and the standalone muon track variables. The primary vertex variables are its position, its longitudinal resolution (if no primary vertex is reconstructed, the beam-spot and a fixed longitudinal

resolution value are used) and a fixed radial spread value is used for the radial resolution. The standalone muon track variables are its direction and its error estimate (limited to fixed maximum values to avoid a too large region surface) and a lower threshold fixed at 60% of its transverse momentum. All the tracker tracks reconstructed from CTF algorithm (cf. Section 2.2.2) in this kinematical region are taken into account in a matching procedure to find the tracker track which should be associated to the standalone muon. The matching procedure is based on variables of the two tracks on a common surface. This surface is chosen to optimize the covariant error matrix of the propagated track parameters and is usually the tracker system outer boundary, the muon system inner boundary, the detector surface of the outermost tracker track hit (for low momentum muons) or the detector surface of the innermost muon track hit (for high momentum muons). Two tracks are matched based on variables from their position (for high momentum muons) and their momenta (for low momentum muons), taking into account the misalignment uncertainty. The reconstruction fails if no compatible tracker track is found, which corresponds to 1% of the muons from collisions. On the other hand, cosmic-ray muons have higher probability to fail the global reconstruction procedure due to the fact that they can miss the tracker acceptance or be incompatible with the beam-spot, leading to a higher cosmic origin contamination for the standalone reconstruction only muons (4 to 5 order of magnitude less favourable than the two other reconstruction procedures). For each association of a tracker track and a standalone muon track, a global refit is performed using selected hits from the tracker track and standalone muon track. Special treatment is dedicated to high momentum muons where electromagnetic showers are produced in the material leading to extra hits in muon chambers. This fit provides one global track per matched pair, and only the best chi-square global track is kept per standalone muon track.

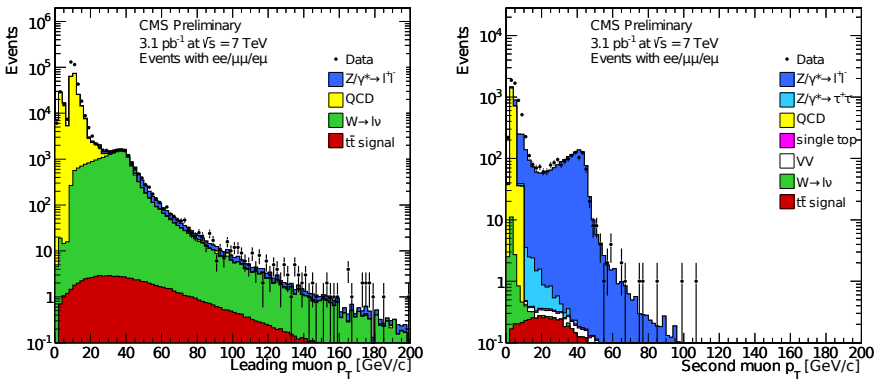
Finally, the tracker muon algorithm allows the muon reconstruction for muons not energetic enough to reach the outer part of the detector. These reconstructed muons can also be used to enhance the muon reconstruction performance or to improve the robustness. This algorithm is based on reconstructed tracker tracks above a given momentum threshold ($p_T > 0.5\,\text{GeV}/c$, $p > 2.5\,\text{GeV}/c$) compatible with at least one segment in muon chambers. No combined track reconstruction is however performed. The tracker track is propagated through the calorimeters until the muon chambers, and is considered as a tracker muon if at least one segment is found in a DT or CSC chamber. The possible double-counting of segments or tracks in the total set of reconstructed tracker muons can be resolved by different algorithms.

A final set of reconstructed muon candidates is defined by merging the candidates sharing same segments or same tracker track among the three reconstruction procedure results. Additionally, depending on the analysis, several variables can be exploited to reduce reconstructed muon candidates that do not correspond to muons (hadron

punch-through, duplication created during the reconstruction due to instrumental effects or imperfections in the pattern recognition) or that are considered as background for the given analysis (e.g., muons from heavy or light flavour quarks for the analysis described in this thesis), with different efficiency and purity compromises. These variables regroup information collected in the tracker, such as the transverse impact parameter and tracker track parameters (momentum, chi-square, number of valid hits, ...), information collected in the calorimeter, such as the energy deposit, information collected in the muon chambers, such as number of missing layers and standalone muon track parameters, or even global information, such as the global track parameters and matching variables.

2.3.3 Muon identification for the dileptonic $t\bar{t}$

The analysis that I have performed in this thesis requires the selection of events containing prompt muons from W bosons. Therefore, I have optimized the muon selection to select these muons and to reject other sources of muons, in particular the ones present in multijet QCD events. The reconstructed muons used for this analysis are obtained from global reconstruction and tracker only reconstruction.



(a) (b)

Figure 2.6: Distribution of the transverse momentum for the leading muon candidate (a) and second (when it exists) muon candidate (b) passing the muon selection described in the text (excepted the p_T cut), after 3.1 pb^{-1} of integrated luminosity. The plain histograms are from simulation only, rescaled to correspond to the same luminosity. The number of entries in the left (right) plot corresponds to the number of events with at least one (two) selected muon candidate. Only visible processes are listed in the legend.

The leptons produced in the $t\bar{t}$ process have a momentum peaking around $M_W/2$. Conversely, a large part of muons from jets are results of a long decay chain, leading to a small momentum. A threshold on the transverse momentum is thus applied: $p_T > 20\,\text{GeV}/c$. This selection is not optimized for the muons from tau lepton decay and then penalizes slightly this contribution. The angular distribution is limited to $|\eta| < 2.5$. This cut is effective for muons even if the acceptance of muon chambers is limited to $|\eta| < 2.4$ because a muon can be produced with $|z| > 0$. The distribution of the transverse momentum for the leading muon (muon with the highest transverse momentum in the event) and the second leading muon (muon with the second highest transverse momentum in the event, when it exists) are shown in Fig. 2.6. These distributions are shown after the complete muon selection described in this section in order to focus on muons coming from W boson, and it appears clearly that the $p_T < 20\,\text{GeV}/c$ region is dominated by background. Discrepancies are also visible between data and simulation for the $p_T < 20\,\text{GeV}/c$ region, which corresponds to a bad description of the low p_T muons, but this is beyond the scope of this analysis which focuses on energetic leptons.

The leptonic decay of the W bosons does not produce any other visible particles colinear to the muon direction. This isolation with respect to other particles can be quantified by evaluating the deposits in the tracker and the calorimeter in a geometrical cone around the reconstructed muon. A relative isolation with respect to the lepton transverse momentum which combines the calorimeter and tracker deposits gives the best compromise between background rejection and signal selection. The isolation is defined according to:

$$I_{\mu}^{comb} = \frac{\sum_i^{tracks} p_T^i + \sum_j^{ECAL} E_T^j + \sum_k^{HCAL} E_T^k}{p_T^{\mu}}, \tag{2.1}$$

where the index i runs over the tracks (excluding the track corresponding to the muon itself) and j and k over the calorimeter deposits in ECal and HCal respectively excluding the energy deposits corresponding to the muon, in a $\Delta R = 0.3$ cone around the reconstructed muon direction. Muons with $I \geq 0.15$ are considered as not isolated and are rejected. The distribution of the isolation is visible in Fig. 2.7. This distribution is shown after the complete muon selection described in this section in order to focus on muons coming from W boson. The fact that muons from W or Z bosons are isolated in contrast to muons from multijet QCD processes is clearly visible. The data agree well with simulations, except that data are slightly less isolated due to the presence of pile-up. This however does not impact the selection for the given isolation threshold.

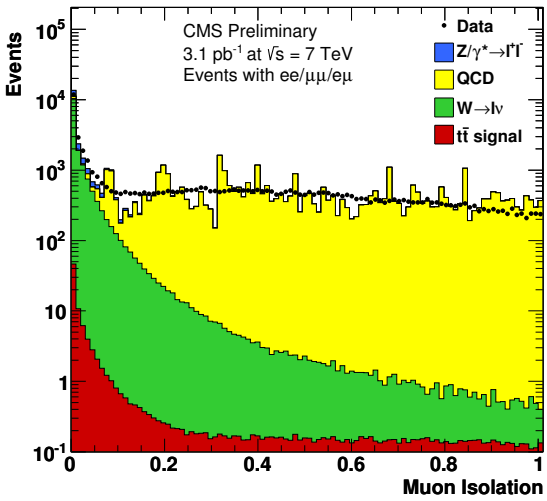


Figure 2.7: Distribution of the isolation variable for the muon candidates passing the muon selection described in the text (excepted the isolation cut), after 3.1 pb^{-1} of integrated luminosity. The plain histograms are from simulation only, rescaled to correspond to the same luminosity. Only visible processes are listed in the legend.

Additional requirements can be used to reject muons not coming from W bosons. Because the top quarks and the W bosons possess a too small time of flight to be observed in the detector, the muons produced in the $t\bar{t}$ process are emitted at the primary vertex position. This characteristic can be quantified by the transverse impact parameter with respect to the interaction vertex, d_0 (in practice, the transverse impact parameter is calculated with respect to the reconstructed beam-spot obtained from observed beam parameters). This transverse impact parameter is computed from the track reconstructed in the tracker subdetector only, and muons with d_0 higher than 0.02 cm are rejected.

The number of valid hits in the tracker detector have to be higher than 10 and the reconstructed muon has to satisfy the prompt tight identification procedure defined in Tab. 2.1.

The summary of the muon selection for the $t\bar{t}$ analysis is given is Tab. 2.2.

Table 2.1: Thresholds used for the prompt tight muon identification.

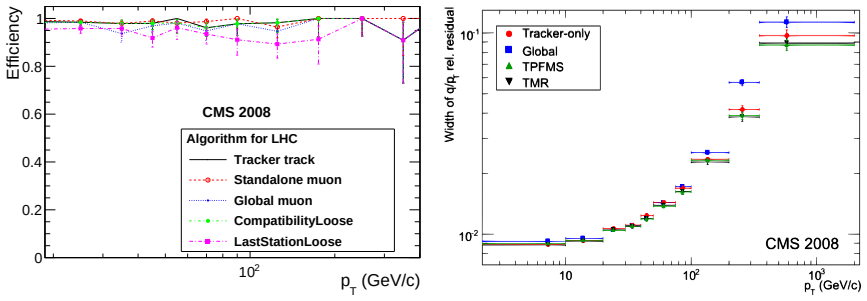
Reconstruction algorithm	GlobalMuon
χ^2/ndof(global track)	< 10
Number of valid muon chambers hits (in global fit)	> 0

Table 2.2: Summary of the muon selection for the $t\bar{t}$ analysis. The Lepton ID is described in Tab. 2.1.

Algorithm	GlobalMuon and TrackerMuon
p_T	$> 20\,\text{GeV}/c$
$\lvert\eta\rvert$	< 2.5
Isolation	< 0.15
d_0	$< 0.02\,\text{cm}$
Lepton ID	GlobalMuonPromptTight

2.3.4 Performance of muon reconstruction and identification

The efficiencies of the different reconstruction algorithms have been evaluated using cosmic-ray muons [107]. Cosmic events have been chosen to contain muons similar to those expected from collisions at LHC (i.e., with trajectories pointing to the nominal interaction point). The fact that the algorithms designed for collision events reconstruct two muons in opposite direction in cosmic events can be exploited to evaluate the reconstruction efficiency, using one of the two reconstructed muons as reference. As visible on Fig. 2.8 (a), the efficiency is higher than 95%, without strong dependence in p_T. The momentum resolution can be evaluated through the width of the relative q/p_T residuals, $R(q/p_T) = ((q/p_T)^{\text{upper}} - (q/p_T)^{\text{lower}})/(\sqrt{2}(q/p_T)^{\text{lower}})$ with "upper" label (resp. "lower" label) referring to the muon reconstructed in the upper (resp. lower) detector half. The Fig. 2.8 (b) shows the width of the relative residuals for the tracker muons and global muons. The TPFMS (Tracker Plus the First Muon Station) and TMR (Truncated Muon Reconstructor) are two additional algorithms reducing the resolution degradation in the global muon reconstruction by using less muon stations. The standalone muon reconstruction is not shown in this figure, but gives worse resolutions.



(a) (b)

Figure 2.8: (a) Efficiency in the central region ($|\eta| < 0.8$) from cosmic events, for the tracker track reconstruction (plain black line), the standalone muon reconstruction (red open circle), the global muon reconstruction (blue dot) and for two muon identification algorithms (green plain circle and pink square). (b) Widths of the Gaussian fits to the relative residuals distributions in the central region ($|\eta| < 0.8$), for tracker muon reconstruction (red circle), global muon reconstruction (blue square) and two additional algorithms using less muon stations than the global muon reconstruction (green triangle and black triangle). From [107].

Different techniques have also been developed to study muon performance in $\sqrt{s} = 7$ TeV data. Firstly, a trivial check consists into comparing the kinematic and tracking variables distributions between data and simulation. For minimum bias collision events, the muons are mainly originating from decays-in-flight of pions and kaons, but also contains muons from heavy flavour decays and a small contribution from punch-through hadrons. Results for $0.47\,\text{nb}^{-1}$ of integrated luminosity in a minimum bias collision events show a good agreement, which confirms that the simulation describes well these contributions. The only discrepancy visible is in the number of hits in the global track which can be related to calibration and alignment conditions at start-up. Tag and probe methods based on the J/Ψ meson resonance can also be used to check the reconstruction efficiency for low momentum muons, leading to good agreement (the most significant deviation corresponding to an observed efficiency about 5-10% higher than expected in the barrel) between data and simulation for $84\,\text{nb}^{-1}$ of integrated luminosity. Muon misidentification and non isolated muons can also be studied via resonances decaying only in kaons, pions and protons and in tag and probe methods in multijet QCD respectively. Again, the simulation seems to describe correctly what is observed in data.

In the framework of this $t\bar{t}$ analysis, the efficiency of muon selection as described in Section 2.3.3 has been evaluated with real data by a tag and probe method. The selection has been factorized between identification and isolation. The global efficiency uncertainty is obtained by assuming that identification and isolation are uncorrelated (this hypothesis has been found reasonable in a check in simulation). The identification and isolation efficiencies are evaluated with probe passing the isolation and identification criteria respectively. The tag corresponds to the muons passing the full lepton selection. No visible dependence on p_T and η has been found for the identification efficiency, measured at 0.992 in simulation and 0.992$\pm$0.002 (statistical uncertainty) in data, and no visible dependence on η has been found for the isolation efficiency, measured at 0.980 in simulation and 0.981$\pm$0.003 for $p_T < 40$ GeV/c and 0.971 in simulation and 0.970$\pm$0.005 in data for $p_T > 40$ GeV/c. The differences between simulations and data for identification, isolation and the global selection are respectively 0%, 0.5% and 0.5% for muons and are taken as an estimation of the systematic uncertainty of these efficiency. These results are obtained for Drell-Yan process. But dileptonic $t\bar{t}$ provides more hadronic activity than Drell-Yan events, and these efficiencies have to be corrected according to this different environment. A first approach can be to select events from data with at least two jets. Unfortunately, the low integrated luminosity used in this study yields large statistical uncertainty. It has then been decided to use Drell-Yan efficiencies corrected for dileptonic $t\bar{t}$ by a factor obtained in simulation (for signal simulation passing the full selection). When doing this, it appears, as expected, that the identification efficiency remains unaffected while the isolation efficiency decreases in dileptonic $t\bar{t}$. In average, the isolation efficiency for the Drell-Yan events has to be corrected by 4% to correspond to dileptonic $t\bar{t}$ events.

2.4 Reconstruction and identification of electrons

The electrons created inside CMS can be deduced from a presence of a track in the tracker and an electromagnetic shower in the electromagnetic calorimeter. Some specific care is taken to avoid mismeasurement due to frequent bremsstrahlung and multiple scattering inside the tracker, or due to the possibility for a pion to mimic a typical electromagnetic shower in the electromagnetic calorimeter. This section describes the electromagnetic calorimeter, the principal subdetector for electron detection, then the algorithm applied for the electron reconstruction and, finally, details the performance of the electron reconstruction in CMS.

2.4.1 Electromagnetic calorimeter subdetector

The electromagnetic calorimeter (ECal) [108], illustrated in Fig. 2.9, is the subdetector surrounding the tracker and dedicated to the detection of electromagnetic particles (electrons and photons).

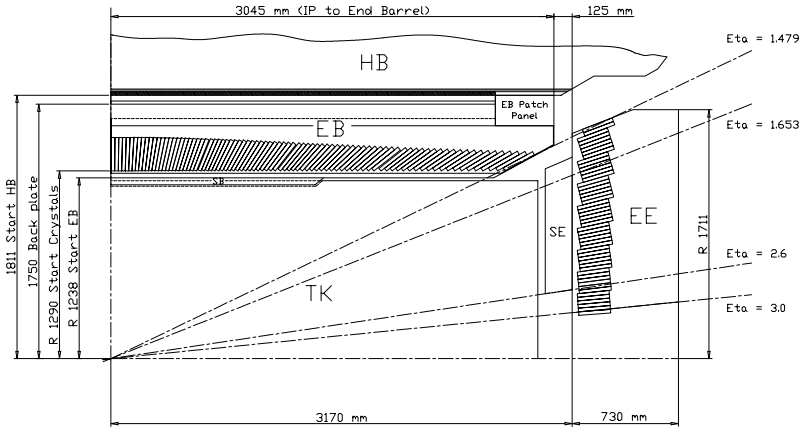


Figure 2.9: Schematic view of the CMS electromagnetic calorimeter [108], in the $z - y$ plane.

It is composed of one layer of 75 848 lead tungstate ($PbWO_4$) crystal scintillators, 61200 in the barrel (EB) and 7 324 in each endcaps (EE). The $PbWO_4$ crystals allow a good granularity in a compact volume with a scintillation decay time of the same order of magnitude of the LHC bunch crossing time.

The electromagnetic calorimeter barrel is 609 cm long and the inner radius is 129 cm far from the beam axis, allowing a coverage until $|\eta| = 1.479$. The pseudorapidity range is complemented by the electromagnetic calorimeter endcaps, Fig. 2.10, located at 315.4 cm from the interaction point, that spread until $|\eta| \leq 3.0$. In the barrel, the crystals are pointing to the interaction point, but with a small deviation of 3° in both η and ϕ. In the endcaps, they are pointing to a point located 1300 mm before the nominal point of interaction on the beam axis. They have 230 mm long (25.8 X_0) in the barrel and 220 mm long (24.7 X_0) in the endcaps. They are connected to avalanche photodiodes in the barrel and to vacuum phototriodes in the endcaps. The crystals in the barrel are grouped by submodule of 400 or 500 crystals, while the crystals in the endcaps are grouped in supercrystals composed of 5×5 crystals.

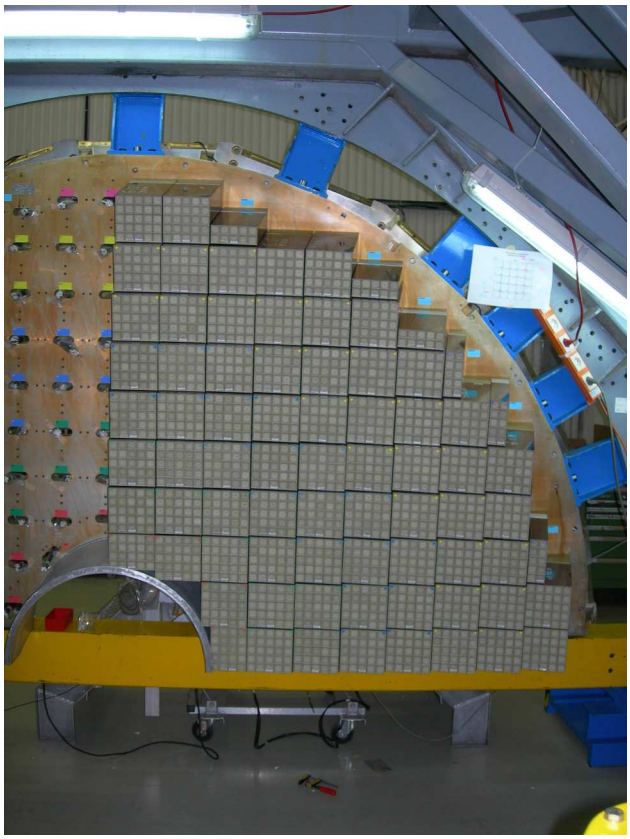


Figure 2.10: The first complete quadrant of the CMS electromagnetic calorimeter Endcap in 2007.

An additional calorimeter, called “preshower detector”, is also added between the EE and the TEC, covering the region between 1.653 and 2.6 in pseudorapidity. This 20 cm thick detector, composed of a lead radiator followed by silicon strip sensors, provides additional hits enhancing the shower shape description. This is useful for neutral pions identification, but also for the electron identification and position determination.

2.4.2 Electron reconstruction

The electron reconstruction [109] starts with the clustering of the energy deposits in the ECal. Some of these clusters are taken as seeds to propagate the electron candidate inside the tracker where an electron track is reconstructed. Finally, further

electron identification procedures are applied to discriminate between real electrons and reconstructed candidates which do not correspond to a real electron.

The clustering in the ECal aims at recovering all the energy spread by the bremsstrahlung radiation of the electron. While the electron trajectory is bent in the ϕ direction, the clustering algorithm covers more surface in ϕ (0.3 radiant in both direction) than in η. In the barrel, the "hybrid algorithm" is used to reconstruct superclusters: a set of clusters is created, seed clusters are found among these clusters, and superclusters are built associating some surrounding clusters to the seed cluster. In the Endcaps, the "multi 5×5 algorithm" is applied: cluster are built with 5×5 crystals and superclusters are formed by grouping clusters in ϕ direction. A supercluster is then selected if its total transverse energy is above a given threshold (4 GeV). A hadronic veto is also applied: the sum of the HCal towers in a 0.15 cone above the supercluster should be small with respect to the supercluster energy ($H/E < 0.15$ where H is the HCal energy and E the ECal energy).

The supercluster seed gives information on the momentum and on the position of the electron at the ECal surface for a non-radiative trajectory (the bremsstrahlung energies and positions being taken into account). Windows in the inner tracker region (ϕ-z window in the pixel system, ϕ-r_T window in the TEC region) are then defined by extrapolation of two trajectories (one for the electron hypothesis, one for the positron hypothesis). In these two windows, tracker seeds are defined, using pixel triplets or pixel pairs plus strip hits from TEC to increase the efficiency for large η. The dedicated track reconstruction taking into account the radiative interactions of the electron is, at this stage, too slow and too complicated. The standard Kalman Filter, which leads to low-number-of-hits tracks for electrons suffering bremsstrahlung, is used. The electrons not affected by radiative interactions lead to correctly reconstructed tracks matching with an ECal supercluster. For the others, the badly reconstructed tracks are refitted by the electron dedicated track reconstruction algorithm and a multivariate analysis is performed on several variables to increase the pre-identification efficiency and the hadron rejection. This method is less efficient in case of electrons in jets (because the supercluster E_T can be biased by neutral particles) or low p_T electrons. In these cases, an alternative method based on Particle Flow algorithm [110] can be used.

From these seeds, an electron dedicated track reconstruction algorithm, called "Gaussian Sum Filter" (GSF) [111], is applied. The algorithm reproduces mainly the Kalman Filter algorithm, but estimates correctly the uncertainties on the next hits using a Bethe Heitler modelling of the radiative energy losses. The chi-square compatibility is set at a very loose value, but the penalty on an invalid hit is increased in order to avoid a merging between an electron and its photon conversion leg. During the final fit procedure, the energy loss at each layer is approximated by a weighted sum of Gaussian distribution from which the momentum can be estimated.

For a track reconstructed from ECal driven seeds, the electron candidate is the association of the track and the initial supercluster. An alternative tracker seed finding method not using ECal information can complement the process. For these tracker driven seed cases, an association between the track and the ECal deposit is performed using Particle Flow techniques: at each energy losses of the electron track, the corresponding photon is extrapolated until the ECal surface, and the set of all clusters, including the one corresponding to the remaining electron, is associated to the track to form an electron candidate. Further preselection cuts are applied to increase the rejection of jets faking electrons. For the ECal driven seeds case, in addition to the $E_T > 4\,\text{GeV}$ supercluster threshold and the $E/H < 0.15$ hadronic veto, cuts on η and ϕ differences between the supercluster position and the extrapolation from the track are applied ($|\Delta\eta| < 0.02$, $|\Delta\phi| < 0.15$). For the tracker driven seeds case, a cut is performed on a multivariate analysis output based on several quantities (p_T, η and several track-cluster matching variables). A last selection is applied to deal with the possibility to associate a primary electron track with an electron leg of a photon conversion from one of its bremsstrahlung. as consequence, several tracks or several superclusters are created, and the ambiguity is mainly solved by rejecting the worse E/P candidate.

The electron momentum determination relies on the track momentum and the ECal energy, but the corrections are dependant of the type of electron candidate. Four classes, based on the number of clusters in the supercluster, the E/P ratio and the bremsstrahlung fraction f_{brem} defined as the relative difference between the inner track momentum and the outer track momentum, are defined:

- "golden" electrons: with a supercluster composed of a single cluster, a high E/P ratio (> 0.9) and a low bremsstrahlung fraction ($f_{brem} < 0.5$);
- "big brem" electrons: with a supercluster composed of a single cluster, a high E/P ratio (> 0.9) but a high bremsstrahlung fraction ($f_{brem} > 0.5$);
- "showering" electrons: affected by bremsstrahlung losses and not in the previous categories;
- "crack" electrons: whose the supercluster starting crystal is close to an η boundary (between ECal barrel modules or between ECal barrel and endcaps).

The corrections also depend on η and p_T. Furthermore, for the low energy electrons, the track momentum measurement is favoured with respect to the ECal energy measurement. The charge identification of the electron candidate is based on the track curvature. Due to the distribution of material, the $1.1 < |\eta| < 2.5$ region suffers from more interactions which can lead to misidentification of the charge. The charge

misidentification increases also as a function of the p_T. To resolve this issue, three charge estimators are defined (the GSF track charge, the associated CTF track charge: a CTF track is associated if enough hits are shared with the GSF track, and the supercluster charge: evaluating the curvature between the first tracker hit and the supercluster position with respect to the beam-spot) and the decision is taken with respect to the majority agreement.

Finally, additional requirements are to be used if one wants to reduce the contamination of misreconstructed electrons. Because the level of purity and the efficiency are analysis dependant, the electron collection is not filtered, but additional variables are associated to each electron candidate and can be used for further selection. Several electron identification are available, divided into two selection methods:

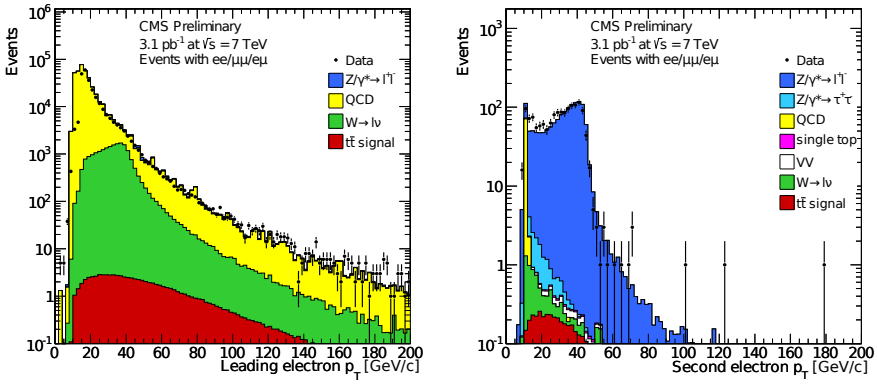
- the fixed-threshold identification: designed for early analysis, it is aimed to be simple and robust. It consists of rectangular cuts based on several variables (H/E, the geometrical matching between the extrapolated position at the ECal surface from the track and the supercluster position: $\Delta\phi$ and $\Delta\eta$, and the calorimeter shower shape in η: $\sigma_{\eta\eta}$), with a set of fixed threshold for the barrel region and the endcap regions;
- the category-based identification: three classes are defined using the same principle of the previously defined electron classes ("low-brem electrons", "bremming electrons" and "bad track"), divided into the barrel region and the endcap regions. The selection are based on matching variables: E/H, $\Delta\phi$, $\Delta\eta$, E/p_{in}, E_{seed}/p_{in}, E_{seed}/p_{out}, on shower shape variables: $\sigma_{\eta\eta}$, Σ_9/Σ_{25} (the ratio of the sum of the energy of the 3×3 domino of crystals centred on the most energetic one and the 5×5 domino), on photon conversion variables: d_0 (the impact parameter with respect to the reconstructed vertex), number of missing hits, and on isolation variables in the tracker, the ECal and the HCal. Cut-based selection or multivariate analysis can be applied to discriminate between signal and background.

Several level of tightness are also proposed for each selection method.

2.4.3 Electron identification for the dileptonic $t\bar{t}$

As stated in Section 2.3.3, the lepton selection for the dileptonic $t\bar{t}$ analysis is optimized for leptons from W bosons. I use only the reconstructed electrons obtained from a calorimeter seed with electromagnetic calorimeter transverse energy higher than 10 GeV. In order to reject electron reconstructed from an inner bremsstrahlung induced by a muon, I also apply a veto on electron which are close to any reconstructed

muons from global reconstruction or tracker only reconstruction ($\Delta R > 0.1$ between the electron and all available global or tracker muon).



(a) (b)

Figure 2.11: Distribution of the transverse momentum for the leading electron candidate (a) and second (when it exists) electron candidate (b) passing the electron selection described in the text (excepted the p_T cut), after 3.1 pb^{-1} of integrated luminosity. The plain histograms are from simulation only, rescaled to correspond to the same luminosity. The number of entries in the left (right) plot corresponds to the number of events with at least one (two) selected electron candidate. Only visible processes are listed in the legend.

As for the muons, a transverse momentum and an angular cuts are applied. Based on the same consideration than for the muon, the electrons with $p_T < 20\,\text{GeV}/c$ are rejected. The angular distribution is limited to the detector acceptance for the emitted electron, $|\eta| < 2.5$. The distribution of the transverse momentum for the leading electron (electron with the highest transverse momentum in the event) and the second leading electron (electron with the second highest transverse momentum in the event) are shown in Fig. 2.11. These distributions are shown after the complete electron selection described in this section in order to focus on electrons coming from W boson. The low p_T region is limited by the other selection requirements (mainly $E_T^{\text{calo}} > 10\,\text{GeV}$). Regardless, it is however clear that the $p_T < 20\,\text{GeV}/c$ region is dominated by background. As in the muon case, discrepancies between data and simulation are visible for the $p_T < 20\,\text{GeV}/c$ region, but again, this does not affect the analysis which focuses on energetic electrons. The fact that electrons are more contaminated from fake leptons from multijet QCD process than muons is also visible. Even if the agreement between data and simulation is fairly good for the $p_T > 20\,\text{GeV}/c$ region, the

fake lepton background is estimated from data in the analysis (Section 3.2.2).

As for the muons, the isolation of electrons can be used to discriminate with respect to electrons not coming from W (or Z) boson. The isolation is defined in a similar way than for the muons, in Eq. 2.2 and Eq. 2.3, except that the barrel region and endcap regions are treated separately and that I_e^{barrel} takes into account the subtraction of a 1 GeV pedestal.

$$I_e^{barrel} = \frac{\sum_i^{tracks} p_T^i + \max\left(0., \sum_j^{ECAL} E_T^j - 1.\right) + \sum_k^{HCAL} E_T^k}{E_T^e} \quad (2.2)$$

$$I_e^{endcap} = \frac{\sum_i^{tracks} p_T^i + \sum_j^{ECAL} E_T^j + \sum_k^{HCAL} E_T^k}{E_T^e} \quad (2.3)$$

In these equations, the index i runs over the tracks (excluding the track corresponding to the electron itself) and j and k over the calorimeter deposits in ECal and HCal respectively (excluding the electron footprint in ECal), in a cone of size $\Delta R = 0.3$ around the reconstructed lepton direction. As for muons, electrons with $I \geq 0.15$ are considered as not isolated and are rejected. The distribution of the isolation for electrons is visible in Fig. 2.12. This distribution is shown after the complete electron selection described in this section in order to focus on electrons coming from W boson. Again, the fact that electrons from W or Z bosons are isolated in contrast to the electrons in multijet QCD process is visible. Again, the data agree well with simulation, excepted for a slight shift in the low value region with no impact on the selection.

The electron selection can also include additional requirements to reject electrons not coming from W boson. As in the muon case, a transverse impact parameter d_0 cut is applied. It is computed from the track reconstructed with the GSF algorithm, and electrons with d_0 higher than 0.04 cm are rejected.

A special treatment is used to reduce the contamination from photon conversions. In addition to the d_0 requirement, only one missing expected hit in the inner tracker is allowed and no electron partner behaving like a result of a photon conversion has to be found. This last condition is based on the presence of another track with opposite sign and a similar η, and on a discriminating variable based on the distance d in the transverse plane for the two points where the tracks are parallel. If a second track with opposite charge and $|\Delta(\cot(\theta))| < 0.02$ and $|d| < 0.02$ cm is found, the electron is rejected. In order to reject electron reconstructed from anomalous noise in the electromagnetic calorimeter, it is also required to have a significant amount of energy stored in the neighbours of the highest crystal corresponding to the ECal cluster of the

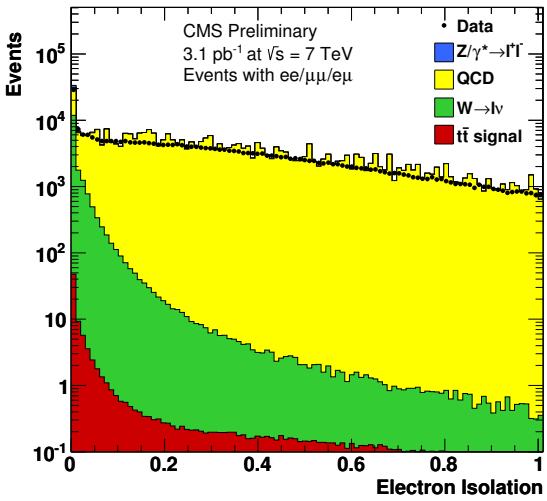


Figure 2.12: Distribution of the isolation variable for the electron candidates passing the electron selection described in the text (excepted the isolation cut), after 3.1 pb^{-1} of integrated luminosity. The plain histograms are from simulation only, rescaled to correspond to the same luminosity. Only visible processes are listed in the legend.

electron. The $S = 1 - E_4/E_1$ variable is computed, with E_1 being the energy in the highest crystal and E_4 the sum of the energy of the four surrounding crystals (two in the z axis and two in the ϕ axis). The electron is rejected if $S < 0.95$. Finally, the electron identification procedure that shows the best compromise between robustness, signal efficiency and background rejection appears to be the so-called "90% efficiency working point" fixed-threshold cuts developed in the study of vector boson processes, detailed in Tab. 2.3. For real data, these variables are corrected for misalignments between the tracker and the endcap electromagnetic calorimeter.

The summary of the electron selection for the $t\bar{t}$ analysis is given is Tab. 2.4.

2.4.4 Performance of electron reconstruction and identification

The electron momentum determination obtained by combining tracker and calorimeter information as described previously leads to a resolution around 2% (resp. 5%)

Table 2.3: Thresholds used for the 90% efficiency working point fixed-threshold electron identification.

Variable	Barrel	Endcap
$H/E <$	0.12	0.05
$\lvert\Delta\phi\rvert <$	0.8	0.7
$\lvert\Delta\eta\rvert <$	0.007	-
$\sigma_{\eta\eta} <$	0.01	0.03

Table 2.4: Summary of the electron selection for the $t\bar{t}$ analysis. The Lepton ID is described in Tab. 2.3.

Algorithm	GSF ECalDriven + $E_T^{\text{calo}} > 10\,\text{GeV}$
Muon veto	$\Delta R(\mu) > 0.1$
p_T	$> 20\,\text{GeV}/c$
$\lvert\eta\rvert$	< 2.5
Isolation	< 0.15
d_0	$< 0.04\,\text{cm}$
γ-conv. cut	Inner tracker missing hit ≤ 1
	If an opposite charge tracks is found:
	$\lvert\Delta(\cot(\theta))\rvert \geq 0.02$ or $\lvert d\rvert \geq 0.02\,\text{cm}$
Lepton ID	Fixed threshold WP90
ECal cleaning	$1 - E_4/E_1 < 0.95$

for isolated golden electrons (resp. isolated showering electrons) with $p_T > 20\,\text{GeV}/c$. The transverse momentum residual distribution, as well as the the residual distributions for momentum direction in η and ϕ are presented in Fig. 2.13 for electrons from $Z \rightarrow ee$ decay simulation. The charge misidentification rate for back-to-back electrons uniformly distributed in p_T and η is described in Fig. 2.14. It has also been evaluated with real data in the dileptonic $t\bar{t}$ analysis framework, based on the percentage of same sign electrons events near the Z mass. 2.1$\pm$0.4(stat.)% and 1.5% is found for data and simulation respectively, which leads to a conservative charge misidentification rate difference between data and simulation of 0.4%, taken as systematic uncertainty.

The efficiency of reconstruction from simulation can be obtained by comparing the number of generated electrons in the fiducial region and in a given p_T range and the

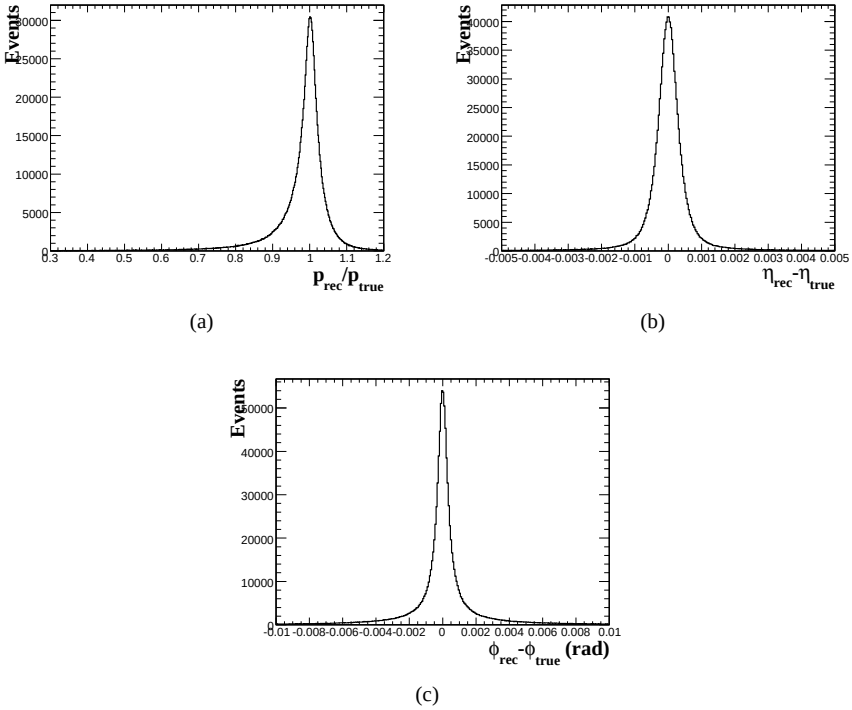


Figure 2.13: Resolution distributions for electrons from $Z \rightarrow ee$ decay simulation of the momentum magnitude (a), the momentum direction in η (b) and the momentum direction in ϕ (c). From [112].

presence of reconstructed electrons matching with these generated electrons. In case of electrons with $p_T > 20\,\text{GeV}/c$ from Drell-Yan process, the electron reconstruction efficiency is evaluated at 98.5% in barrel region and 96.1% in endcap regions. Two methods are used to evaluate the reconstruction efficiency from data [113] for electron with $p_T > 20\,\text{GeV}/c$. The first one is a tag and probe method based on Drell-Yan process: the tag is an ECal driven seeds reconstructed electrons with supercluster transverse energy higher than 20 GeV and standard identification and isolation requirement, and the probe is an ECal supercluster with transverse energy higher than 20 GeV and leading to a dileptonic invariant mass close to the Z boson mass. The second method is an unbinned maximum likelihood fit on the reconstructed W boson mass from selected events with a supercluster (with transverse energy higher than 20 GeV, seed satisfying some shower shape requirements and isolated in the tracker and the calorimeters) absence of jets (no jets with corrected transverse energy higher than 25 GeV in $|\eta| < 3$ region) and presence of missing transverse energy (more than 30% of the transverse

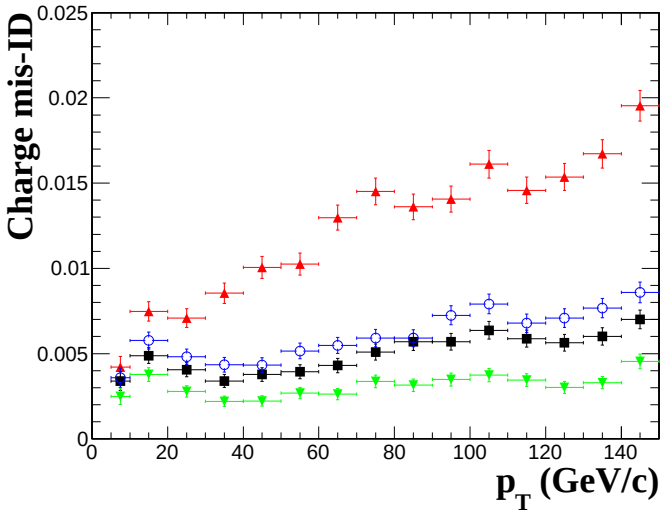


Figure 2.14: Electron charge identification performance from simulated back-to-back electrons uniformly distributed in p_T and η as a function of p_T. The different methods are represented: the GSF track charge (upward red triangles), the associated CTF track charge (blue circles), the supercluster charge (black square) and the combined charge (downward green triangles). From [112].

energy of the supercluster). The maximum likelihood fit is applied with free background shape but with a constraint on the signal shape from simulation, and allows the extraction of the reconstruction efficiency by comparing the selected supercluster with respect to the reconstructed electron. For an integrated luminosity of $\sim 200\,\mathrm{nb}^{-1}$, the reconstruction efficiency obtained by combination of these two methods is found in good agreement with simulation in both barrel and endcap regions with a ratio compatible with the unity, as shown by Fig. 2.15. For lower p_T, the consistency between simulation and real data can be checked by comparing the shape of the distribution of standard variables, e.g., p_T, η, E/P (ratio between the energy of the supercluster and the momentum of the GSF track) and $\Delta\eta$ (difference in η of the supercluster position and the track extrapolation from the innermost measurement). The simulated distributions are rescaled in order to correspond to the amount of real data, and if the expected efficiency is not correct, shape discrepancies should be observed. Such validation has been performed with minimum bias data after $\sim 3\,\mathrm{nb}^{-1}$ of integrated luminosity, directly at the reconstruction level without any other selection on the electron. At this level, the list of reconstructed electron candidates is mainly constituted by misreconstructed electron candidates from hadrons (61.5%), electrons from photon conversion

(33.9%) and real electrons from jets ($\sim$4.6%, composed at 82% of B and D mesons decays, at 14% of Dalitz decays, and then of J/ψ decays). Apart from effects of understood imperfect calibration of the EE and misalignment between ECal and Tracker, the distributions show a good agreement between simulation and data.

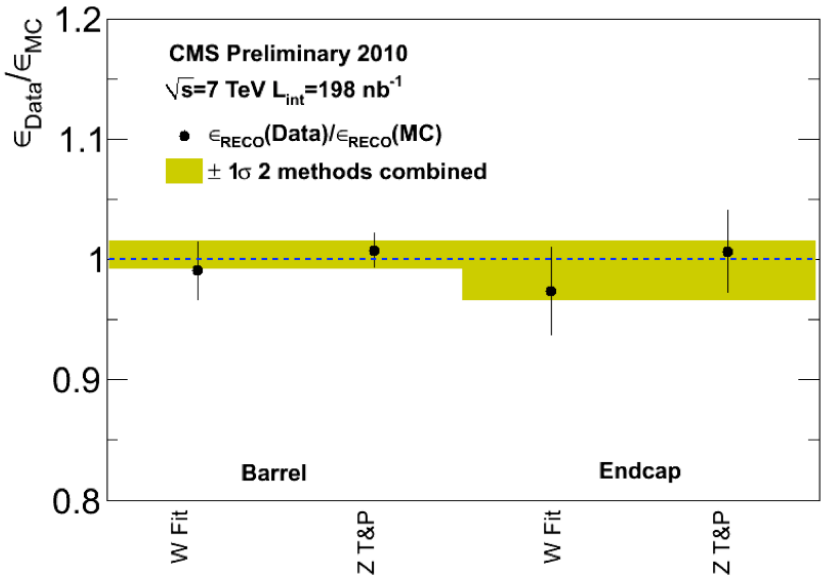


Figure 2.15: Ratio between the measured electron reconstruction efficiency (after $\sim 200\,\mathrm{nb}^{-1}$ of integrated luminosity) and the expected efficiency from simulation. The results are divided in barrel and endcap regions and for the two methods describe in the text. The yellow band represents the two methods combined results with one sigma errors. From [113].

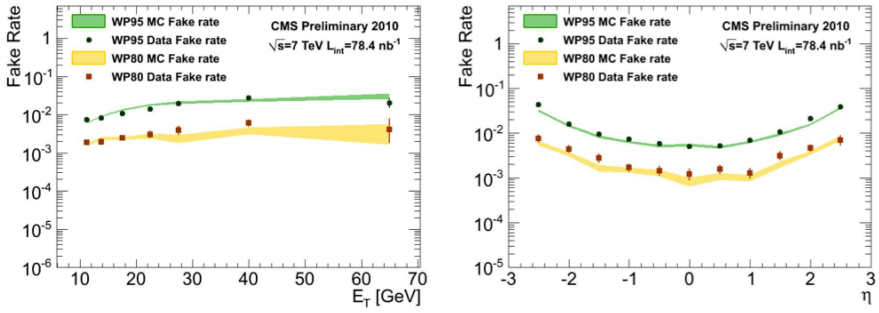
Identification, isolation and conversion rejection selection have been defined, in simulation, in several working point corresponding to a given final efficiency (for a prompt electron with $E_T > 20\,\mathrm{GeV}$, in the case of fixed-threshold identification) or to a given signal over background ratio (for single electron, in the case of category-based identification). These selections rely on different variables, for which the expected distribution from simulation can be compared in barrel and endcap regions with real data in different process selection (e.g., $W \to e\nu$) to validate the expected efficiency. The comparisons show a good agreement between the simulation and the real data. As in the muon case, the selection efficiency for the dileptonic $t\bar{t}$ analysis has been directly extracted from data. Because no selection at all leads to too many backgrounds for the electrons, the selection has been factorized between identification and isolation.

The identification and isolation efficiencies can be evaluated with probe passing the isolation and identification criteria respectively. The tag corresponds to the electrons passing the full lepton selection. These efficiencies have been evaluated for different p_T and η ranges and compared with pure Drell-Yan simulation or with Drell-Yan + W+Jets simulation, and are listed in Tab. 2.5. The estimations of the difference between simulations and data for identification, isolation and the global selection are respectively 2%, 1% and 2.5% for electrons and are taken as an estimation of the systematic uncertainty of these efficiency. As in the muon case, these efficiencies are corrected to corresponds to the $t\bar{t}$ environment.

Table 2.5: Identification and isolation efficiencies for electrons, compared between data, and Drell-Yan or Drell-Yan and W+Jets simulations. 100% uncertainty on the simulated background has been taken into account in the final simulated uncertainty, and the uncertainty for data is statistical.

Identification	$p_T < 40\,\text{GeV}/c$	$p_T > 40\,\text{GeV}/c$	$\lvert\eta\rvert < 1.5$	$\lvert\eta\rvert > 1.5$
DY MC	0.924	0.937	0.947	0.871
DY & W+jet MC	$0.906_{\pm 0.007}$	$0.933_{\pm 0.004}$	$0.942_{\pm 0.005}$	$0.863_{\pm 0.008}$
Data	$0.918_{\pm 0.008}$	$0.935_{\pm 0.008}$	$0.934_{\pm 0.006}$	$0.886_{\pm 0.013}$
Isolation	$p_T < 40\,\text{GeV}/c$	$p_T > 40\,\text{GeV}/c$	$\lvert\eta\rvert < 1.5$	$\lvert\eta\rvert > 1.5$
DY MC	0.961	0.984	0.970	0.972
DY & W+jet MC	$0.948_{\pm 0.013}$	$0.980_{\pm 0.004}$	$0.961_{\pm 0.009}$	$0.962_{\pm 0.010}$
Data	$0.966_{\pm 0.005}$	$0.988_{\pm 0.004}$	$0.974_{\pm 0.004}$	$0.972_{\pm 0.007}$

A good understanding of reconstructed electron candidates not corresponding to a real electron is also essential to the determination of the expected amount of background. In this case, an unbiased sample has to be selected, using jet triggers and jet selection. Usual electron variable distributions can be obtained from events with one jet and one electron candidate not close to the selected jets (W boson processes can also be rejected with a threshold on the missing transverse energy). The comparison between simulation and real data shows good agreement, except for some shift to lower $\sigma_{\eta\eta}$ in endcaps (due to a miscalibration) and some excess in the tail of the bremsstrahlung fraction in endcaps. The fake rate (i.e., the rate of the reconstructed electron candidates not corresponding to a real electron) can also be evaluated in real data and be compared with simulation, as visible in Fig. 2.16 in the case of fixed-threshold identification at 80% and 95% working point.



(a) (b)

Figure 2.16: Electron fake rate per reconstructed electron candidate as a function of E_T (a) and η (b), for fixed-threshold identification at 80% and 95% working point in simulation and real data (after $\sim 80\,\text{nb}^{-1}$ of integrated luminosity). From [113].

2.5 Reconstruction and identification of jets

Due to the confined nature of the coloured particles, the phenomenon of hadronization happens when a quark or a gluon is produced by a particle interaction, leading to a shower of hadrons. These particles are boosted in the original quark direction, and are then generally mainly contained in a geometrical cone around this direction. This particle shower is detected by the calorimeters in form of energy deposits. The jet reconstruction consists in clustering the calorimeter deposits (or more complex reconstructed objects) in order to reconstruct the direction and the energy of the initial parton, useful for the understanding of the event.

This section describes the hadronic calorimeter, used for the jet reconstruction. The different kind of jet reconstruction algorithms that have been used in the dileptonic $t\bar{t}$ analysis in the next chapter are detailed. And finally, the missing transverse energy and the b-tagging algorithm, directly related to jets, are presented.

2.5.1 Hadronic calorimeter subdetector

The Hadronic Calorimeter (HCal) is a sampling calorimeter [114], composed of alternate layers of plastic scintillators and metallic absorbers, dedicated to the detection of hadronic particles. A picture of the barrel during its construction in 2002 is shown on Fig. 2.17, a longitudinal view is shown on Fig. 2.18. The main requirement of this subdetector, knowing that the resolution will be anyway degraded by pile-up, magnetic field effect and parton fragmentation, is a good granularity. This calorimeter is

Figure 2.17: The CMS Hadronic Calorimeter Barrel in 2002.

organized in towers pointing to the interaction point. The granularity in the eta-phi space is 0.087 $\times$ 0.087 below $|\eta| = 1.6$ and 0.17 $\times$ 0.17 between $|\eta| = 1.6$ and $|\eta| = 3$, which is the maximum coverage of the endcaps. The layers of absorbers and scintillators have been specially designed to avoid the presence of dead material in eta-phi plane, enabling a perfect hermeticity. The absorbers are mainly composed of flat brass plates, which is a non-magnetic material mandatory in a strong magnetic environment. The external absorber layers are also partially composed of stainless steel to support the structure. To facilitate the manipulation and to avoid dead zones, the scintillator are grouped in trays, divided in eta section, containing also the optical decoding unit, linked to the scintillators and to external read-out boxes by wavelength fibres, and some laser alignment and radioactive calibration facilities. Several layers of absorbers are used to obtain a thickness corresponding to a sufficiently high number of interac-

tion lengths. Some layers are however reduced in the last eta towers of the barrel of the first eta towers of the endcaps to allow space to the read-out boxes.

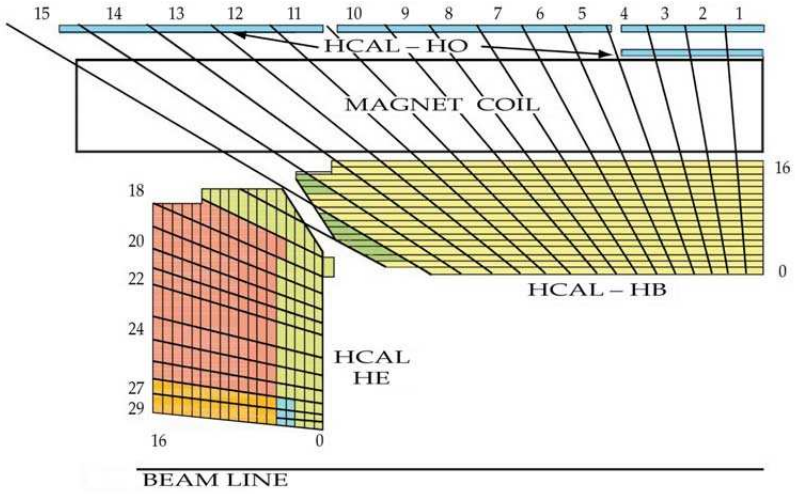


Figure 2.18: Schematic view of the CMS hadronic calorimeter [114], in a quarter of the z-y plane.

The hadronic calorimeter barrel (HB) is 860 cm long and covers a pseudo-rapidity reaching $\eta = 1.3$. Its inner radius is 177 cm and its outer radius is 295 cm. It is divided in two symmetrical and identical half-cylinders, one covering the positive part of the z axis (HB+) and the other one the negative part (HB-). They are formed by 18 portions in the transverse plane. Each portion layer contains 4 scintillator trays, parallel to the beam axis and dividing again the transverse plane in 4 (which lead to 18 $\times$ 4 segments, corresponding to $\Delta\phi = 0.087$). The scintillator trays are divided in 16 sections of $\Delta\eta = 0.087$ each. Radially, there are 16 absorber layers and 17 layers of scintillators (excepted for high η towers). The absorbers are about 50 mm of thickness (6 $\times$ 56.5 mm and 8 $\times$ 50.5 mm of brass absorbers and 75 mm for the first layer and 40 mm for the last layer in stainless steel), and corresponding to a minimum (i.e., at $|\eta| = 0$) of 5.82 interaction lengths and a maximum (i.e., at $|\eta| = 1.3$) of 10.6 radiation lengths. Two different plastic scintillators are used: one (9 mm-thick) appears once and is located in front of the first absorber layer and recovers the showers starting in the inert material between the HCal and the ECal (the material of ECal correspond to about 1.1 interaction length), and one used in all the barrel (3.7 mm-thick or 9 mm for the last layer). The scintillators located just after the last absorber are thicker and recover the late showers leaking out the back of the HCal.

The hadronic calorimeter endcaps (HE) are located from $|z| = 388$ cm to $|z| = 570$ cm, with thickness corresponding to 10 interaction lengths, and cover a pseudo-rapidity from $|\eta| = 1.3$ to $|\eta| = 3$. There are 17 layers, excepted for the low η towers where the latest layers is reduced to allow space for the read-out boxes. To compensate, an extra-layer (layer -1) is added to the tower 18. The scintillator trays are arranged in megatile dividing the transverse plane in 36 portions.

2.5.2 Jet reconstruction

In a particle detector, a reconstructed jet is a cluster of deposits which are supposed to be resulting from the hadronization of a parton. If all theses deposits correspond exactly to the secondary particles coming from the initial parton, then the reconstructed jet corresponds to the physical jet. In this case, the kinematics of the secondary particles can be deduced and the initial parton kinematic can be reconstructed, with a precision depending on the energy deposits resolution. Of course, in an environment rich in jets, which is the case for proton-proton collisions, the probability of missing a deposit for a given jet or of adding a deposit from another jet is not null and degrades the resolution of the reconstructed jet.

To cluster these deposits, several algorithms, called "Jet Algorithms", implemented in the FASTJET package [115] can be used.

Ideally, these algorithms should be infrared safe (i.e., the addition of a small deposit cannot turn the jets result in a drastically different set) and colinear safe (i.e., changing a deposit into two deposits constituting one half of the initial deposit each cannot turn the jets result in a drastically different set). In some case, time consumption of the algorithm has also to be taken into account.

The main algorithms used in this study are:

- The iterative cone algorithm: this algorithm is not infrared safe and colinear safe, but it is quite simple and fast. Therefore, the jet reconstruction in the trigger uses this algorithm. It is based on the creation of a protojet around a seed (which is the deposit with highest energy available) regrouping all the deposits contained inside a cone of a given radius. When the protojet is created, the jet axis is recomputed and a new protojet is evaluated around this new axis. The iteration stops when the protojet appears to be stable. In this case, the deposits inside the cone are associated to this jet and are not available for another jet reconstruction. In this study, the iterative cone is only used at trigger level.
- The anti-k_T algorithm [116]: this is a sequential recombination algorithm, clustering elements two-by-two at each iteration. From all the elements available

(which are the energy deposits for the first iteration, or the deposits and the protojets formed from the combination for the further iterations), distances are defined. For an element i and an element j, the d_{ij} parameter is defined by $d_{ij} = min(k_{ti}^{-2}, k_{tj}^{-2})\Delta R^2/R^2$, where k_{ti} is the transverse momentum of the element i, ΔR is the distance in the η-ϕ plane, and R is a free parameter. For this element i, $d_{iB} = k_{ti}^{-2}$ is also defined and correspond to the distance to the beam. The iterative clustering consists into finding, for all the elements, the smallest d. If it corresponds to a d_{ij}, a protojet is formed by combining the elements i and j. If it corresponds to a d_{iB}, then the element i is called a jet and is removed from the list of available element. The multiple advantages of this algorithm are that it is fast, infrared safe, colinear safe and leads to more intuitive results (the jets are more conical than with other sequential algorithms). This algorithm is applied to reconstruct all the jets used in the analysis of this study, with the parameter $R = 0.5$.

The jets can be obtained from simulation, using the appropriate stable particles from the generator information as input for the clustering algorithm. These jets are called "generator jets" or "GenJets". From the generator information, it is also possible to deduce the initial parton which is responsible for the jet creation. When a jet is provided directly from the jet algorithm, without any correction, (called "raw jet"), the jet energy obtained can differ from the corresponding generator jet. These discrepancies between raw jet energy and generator jet energy are due to the non-uniform and non-linear response of the CMS calorimeters, but also to the electronic noise and the additional energy coming from underlying events and pile-up interactions. The correction procedure is factorized in several steps, each step using the corrected value from the previous step. The first correction level is the offset correction, subtracting the mean energy contribution of electronics noise and the pile-up events. The second correction level ("L2") is the relative correction, where the energy is rectified by a factor depending in η and p_T determined with respect to a control region (the central region, $|\eta| < 1.3$) in order to flatten the η distribution of the energy values. The third correction ("L3") is the absolute correction, depending on the p_T and applied in order to adjust the mean reconstructed energy value at the generator jet energy value. Other correction levels, based on the electromagnetic content or the flavour type of the jets can also be applied, but are not used in the following study. Finally, these corrections relying on simulation have to be controlled with real data [117]. Due to the very successful CMS jet energy response simulation, it has been decided that the simulation based corrections will be used and an optional residual calibration can be applied to data jets in order to correct the small discrepancies observed between simulations and data. In the following study with early data, this residual calibration have not been applied.

Four kinds of jets are used in CMS:

- The calorimeter jets ("CaloJets") [118]. These jets are based only on electromagnetic and hadronic calorimeters information. The deposits used are obtained from the calorimeter towers. These towers regroup one HCal cell and the corresponding ECal cells. In order to reduce noise, only the energy of cells above a given energy threshold (which can depend on the subdetector or the region) is summed and constitutes the tower energy. In order to reduce the pile-up effect, only the towers with energy above a given threshold are used in the jet reconstruction algorithm. They are treated as massless particles with a direction determined by the centre of the tower and the centre of the detector.

- The tracker jets ("TrackJets") [119]. These jets are based only on tracker information. In this sense, they are totally independent with respect to the CaloJets. Tracks with some good quality requirements originating from a common vertex are used as inputs for the jet algorithm. If needed, several primary vertices can be used to reconstruct TrackJets associated with different vertices but consistent with respect to the origin of their constituents. They gain from the good tracker resolution at low transverse momentum, but do not account for neutral particles. They also provide better angular resolution, due to the fact that the direction is evaluated at vertex, while charged particles used in CaloJets has been deflected before reaching the calorimeters, and are more robust with respect to the pile-up contribution, due to the easy rejection of contributions not belonging from the same vertex. Their energy resolution is similar to the CaloJets: the gain in resolution due to the tracker use is however compensated by the absence of the neutral particles contribution. But they do not suffer from large jet energy scale uncertainty. A more complete description of the reconstruction is available in Appendix A.

- The jet-plus-tracks jets ("JPTJets") [120]. These jets are energy corrected CaloJets, the correction being obtained from the tracker information. Single charged particle response in the calorimeter is evaluated from real data using isolated tracks. For each track coming from the production vertex, if its extrapolation to the calorimeter surface corresponds to an used calorimeter tower, the average expected energy of this track in the calorimeter is subtracted from the jet energy. If the track direction at vertex belongs inside the jet cone, the momentum measured in the tracker is added to the jet energy. Moreover, the contribution from tracks which are inside the jet cone at vertex but do not reached the calorimeter surface inside the jet cone is also added to the jet energy. Finally, some corrections from track finding inefficiency and lepton track consideration are also performed (the calorimeter response for muons is evaluated at 2 GeV, which is the mean value of muon energy deposition in the calorimeter, when the electron

response is considered as already calibrated). The jet direction is also corrected using the tracker information.

- The particle-flow jets ("PFJets") [121]. These jets are reconstructed from so-called "PFCandidates", which are reconstructed from all subdetectors and correspond to muons, electrons, photons, charged hadrons and neutral hadrons. Each kind of PFCandidates is reconstructed and commissioned specifically, and the information associated to a PFCandidate cannot be used in another PFCandidate, ensuring consistency of the set. For example, the charged hadrons are reconstructed from tracks. The neutral hadron are then reconstructed from the remaining deposits in the hadronic calorimeter after subtraction of the expected deposit from charged hadron. The good electromagnetic calorimeter granularity can also be exploited during the electrons and photons reconstruction. This optimal use of the detector information enhances the spatial resolution and the energy resolution.

Even if care is dedicated to reducing noise and fake signal at deposits level, it is still possible to reconstruct a jet from a non physical origin (mainly from calorimeter and/or readout electronics noise). Fortunately, specific jet variables can provide information for the rejection of non-physical jets (called "fake jets" or "fakes" in jet context). Different jet identification criteria can be defined, the choice of the procedure depending of the nature of the jets and the efficiency and purity needed for the study. These identification conditions are defined in pure noise data samples recorded in non-collision state (during cosmic test or LHC operation with empty bunches).

The typical variables for the CaloJets and JPTJets identification are the electromagnetic energy fraction (EMF), the minimum number of calorimeter hits clustered into a jet which contain 90% of the jet energy (N_{hits}^{90}) and the fraction of energy contributed by the highest energy HPD readout (f_{HPD}). The typical variables for the PFJets are the fraction of charged and neutral contribution in the HCal (respectively CHF and NHF) and in the ECal (respectively CEF and NEF), the number of constituents (NC) and the number of charged constituents (ChMult). There is no specific jet identification procedure for TrackJets, knowing that the fakes in the Tracker are particularly infrequent, amplified by the good quality requirement applied on the tracks before the jet algorithm.

The jet identification threshold used in the dileptonic $t\bar{t}$ analysis is summarized in Tab. 2.6.

Table 2.6: Thresholds used for CaloJets, JPTJets (a) and PFJets (b) identification. The description of the variables is available in the text.

Variable	Threshold
EMF	> 0.01
N_{hits}^{90}	> 1
for $\|\eta\| < 2.5$ only:	
f_{HPD}	< 0.98

a)

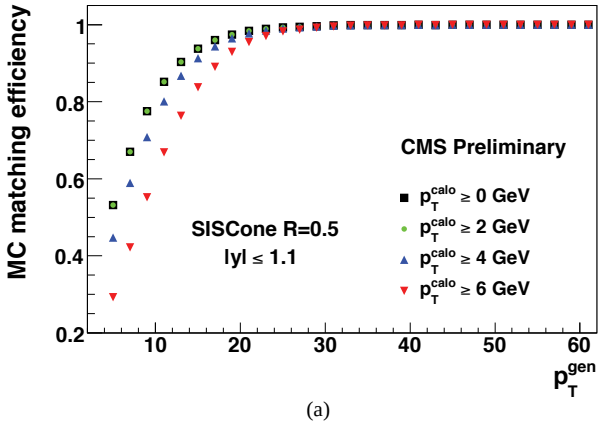
Variable	Threshold
NHF	< 0.99
NEF	< 0.99
NC	> 1
for $\|\eta\| < 2.4$ only:	
CHF	> 0
CEF	< 0.99
ChMult	> 0

b)

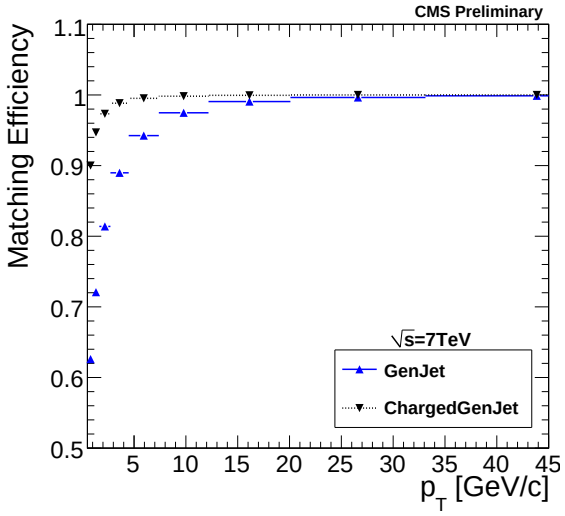
2.5.3 Performance of jet reconstruction and identification

The jet reconstruction efficiency can be estimated in simulation from the jet matching efficiency. For this purpose, the standard reconstructed CaloJets, JPTJets, PFJets and TrackJets are identified under the generic term "RecoJets", and generator jets ("GenJets") are obtained by applying jet algorithm on the set of stable particles at the generator level of the simulation (with special treatment of the neutrinos, the muons and the neutral particles, in order to simulate what should be expected by the RecoJets). For the TrackJet case, I have also produced ChargedGenJets, using charged generator-level particles only. The matching efficiency is obtained by evaluating the fraction of GenJets which do not match with any RecoJets. This method depends on the matching criterion, usually chosen as a geometrical matching of the jet axis position in the η-ϕ space: $\Delta R(\text{GenJet}, \text{RecoJet}) < R_{max}$. In Fig. 2.19, the matching efficiency for $R_{max} = 0.5$ is given for CaloJets and TrackJets with respect to the p_T of the GenJet (or ChargedGenJet). A $\sim$100% reconstruction efficiency is reached for $p_T > 30\,\text{GeV}/c$ (resp. $p_T > 20\,\text{GeV}/c$) for central CaloJets (resp. TrackJets). The reconstruction efficiency can also be evaluated from data, using tag and probe method in back-to-back Z+Jet data (where the tag is the TrackJet in the direction opposed to the Z boson and the probe is the CaloJet matching the TrackJet) [122].

The p_T resolution can also be evaluated from simulation for corrected CaloJets, JPTJets and PFJets, in a sample of QCD dijet events. The RecoJets and the GenJets are then associated by geometrical matching (the maximum distance being $\Delta R <$ 0.2, with smaller value for good position resolutions condition in order to conserve



(a)



(b)

Figure 2.19: Matching efficiency between the jets reconstructed from generator-level particles and CaloJets (a) and TrackJets (b) with respect to the generator jet p_T. The matching criterion is $\Delta R(\text{GenJet}, \text{RecoJet}) < 0.5$. The CaloJets are reconstructed with the SISCone 0.5 algorithm in a Z+Jets sample and is restricted to central jets for different p_T threshold. The TrackJets are reconstructed with the anti-k_T 0.5 algorithm in a minimum bias sample, is restricted to $|\eta| < 2$ and is compared with usual generator jets (GenJets) and charged-particle-only generator jets (ChargedGenJets). From [122] (a) and [119] (b).

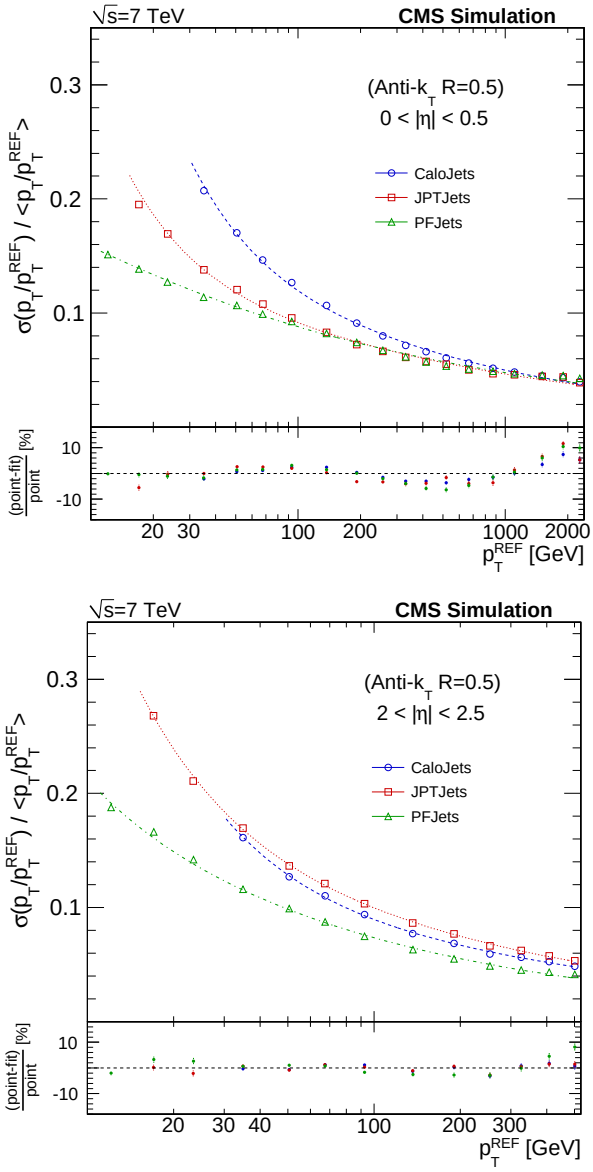


Figure 2.20: Jet p_T resolution from simulation. The σ of the Gaussian fit is shown for CaloJets (blue circles), JPTJets (red squares) and PFJets (green triangles) as a function of the p_T(GenJet), for the two extreme η regions ($0 < |\eta| < 0.5$ and $2 < |\eta| < 2.5$). From [123].

a similar matching efficiency between the three kind of jets). Only the two highest p_T matched GenJets are considered. The jet response for each pair is defined as $p_T(\text{RecoJet})/p_T(\text{GenJet})$. The jet resolution is then given by the standard deviation of the Gaussian fit in the core region of the jet response distribution, and can be evaluated by $|\eta|$ and p_T ranges (Fig. 2.20). In special case where the tails of the distribution have to be taken into account, which is not the case of the following study, a more complex fit can be used.

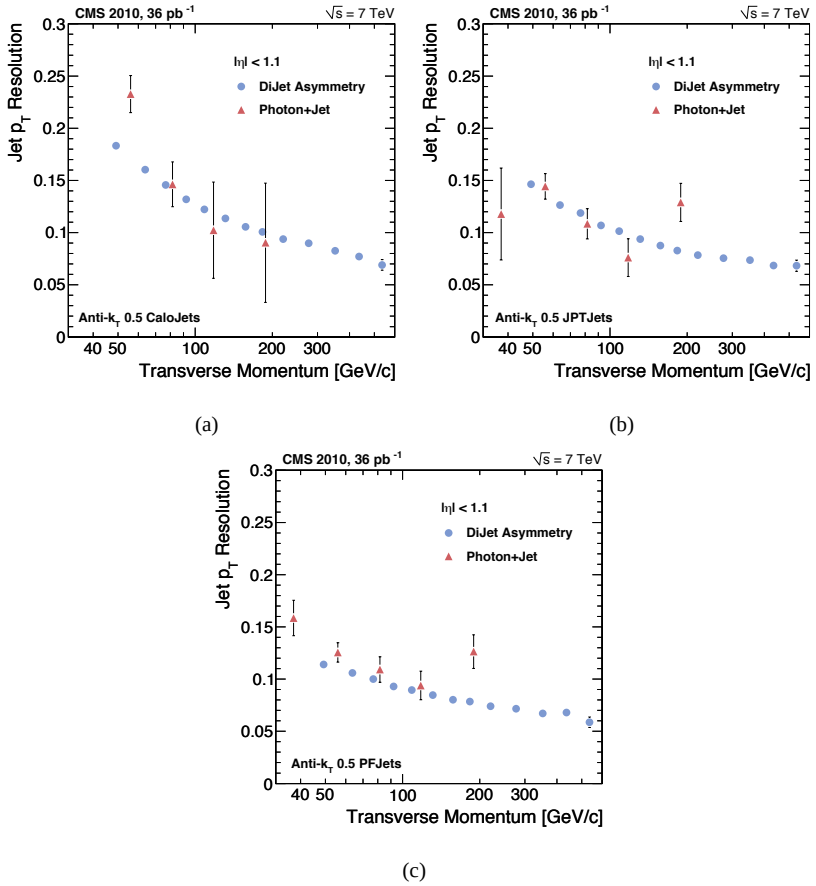


Figure 2.21: Jet p_T resolution from data for CaloJets (a), JPTJets (b) and PFJets (c), in the $0 \leq |\eta| \leq 1.1$ region (for $36\,\text{pb}^{-1}$ of integrated luminosity for proton-proton collision at $\sqrt{s} = 7\,\text{TeV}$). The dijet asymmetry method (blue dots) and the γ+Jet balance method (red triangles) are compared. From [117].

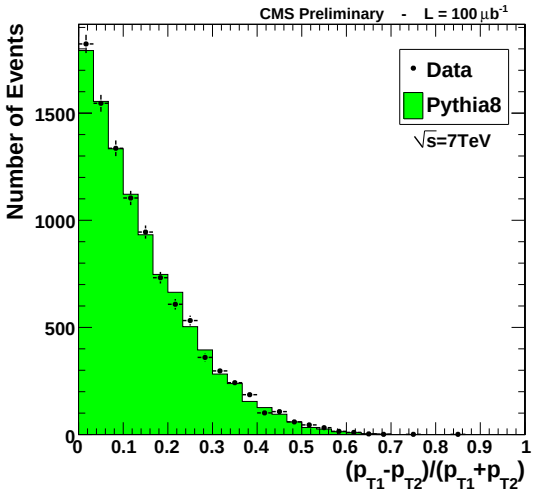


Figure 2.22: Dijet momentum asymmetry $(p_{\mathrm{T}1} - p_{\mathrm{T}2})/(p_{\mathrm{T}1} + p_{\mathrm{T}2})$ for di-TrackJet events with $p_{\mathrm{T}} > 10$ GeV/c (in dijet minimum bias events after 100 μb^{-1} of integrated luminosity). The distributions are normalized to the number of events in each plot. From [119]

The RecoJets p_T resolution can also be evaluated directly from data, using the dijet asymmetry method [117]. This method exploits the momentum balance in the transverse plane in the dijet production events. In a sample of collision data triggered by dijet p_T average triggers (15 and 30 GeV/c thresholds), the events with two jets azimuthally separated by $\Delta\phi > 2.7$ and a third jets below a given threshold are selected. The asymmetry variable is defined by $A = (p_T^{jet1} - p_T^{jet2})/(p_T^{jet1} + p_T^{jet2})$ ($jet1$ and $jet2$ are the two corrected jets satisfying the previous selection, not sorted in p_T). Neglecting the effect of the presence of soft radiation (limited by the third jet veto), the variance of the asymmetry distribution is directly related to the relative p_T resolution for a jet of $p_T = (p_T^{jet1} + p_T^{jet2})/2$. The third jet p_T can be used to correct the soft radiation effect. Additional physics effects, as proton remnants contribution or particles from the parton emitted outside the jet reconstruction cone, can be corrected by evaluating their impacts on the dijet asymmetry method in simulation. After correction, a reasonable agreement between simulation and data is observed. These results are also available for higher integrated luminosity and the γ+Jet balance method can also cross check the results [117]. Both results are shown in Fig. 2.21. The asymmetry variable has also been observed for TrackJets, in minimum bias data after $\sim 100\,\mu$b^{-1} of integrated luminosity, Fig. 2.22, and appears to agree with simulation.

It finally appears that p_T resolution for CaloJets, JPTJets and PFJets is the order of 10% for central jets with $p_T > 100\,\text{GeV}$, while for lower p_T jets, the resolution is clearly improved for JPTJets and PFJets. The final jet energy scale uncertainty used in the dileptonic $t\bar{t}$ study detailed in this thesis is taken at 5%.

The jet position resolution can also be evaluated, from simulation, using the same method as describe in the p_T resolution determination from simulation, by calculating the standard deviation of the Gaussian fit on the distribution of $\Delta\eta = (\eta(\text{RecoJet}) - \eta(\text{GenJet})) \times sign(\eta(\text{GenJet}))$ and $\Delta\phi = (\phi(\text{RecoJet}) - \phi(\text{GenJet}))$. These results can be cross-checked in data by using TrackJets as reference jets, knowing that they are based on tracks for which the position resolution is smaller and well modelled.

A detailed description of the commissioning of TrackJets with first data is also available in Appendix A.

2.5.4 Transverse Missing Energy and b-tagging

Additional reconstruction related to jets can be performed. The missing transverse energy and the b-tagging is briefly described in this section.

Missing transverse energy. The missing transverse energy ($\not\!\!E_T$) is the measure of the energy imbalance in the transverse plane. Although a small part of the energy imbalance can be the result of known detector artifacts, it indicates the presence of neutral weakly interacting particles which have escaped the detector without producing any direct response in any subdetector. Such particles are the neutrinos, but also the long-lived neutral weakly interacting particles predicted in beyond standard model scenarios. The missing transverse energy vector is the opposite of the vector sum of the transverse momenta of all final-state particles reconstructed in the detector. The missing transverse energy is then the magnitude of this vector. Several reconstruction techniques are available, mainly based on the jet reconstruction algorithm used to determine the momentum vector of the jets. The calorimeter-based $\not\!\!E_T$ ($\not\!\!E_T^{\text{Calo}}$) [125] is calculated from calorimeter tower energies with noise suppression threshold and correction for deposits associated to the muons. The track-corrected $\not\!\!E_T$ (tcMET) [126] is obtained from the calorimeter-based $\not\!\!E_T$ but also takes into account the information from tracker in a similar way than in the JPTJets reconstruction. The particle-flow $\not\!\!E_T$ ($\not\!\!E_T^{\text{PF}}$) [121] is calculated from the reconstructed PFCandidates in the particle-flow algorithm. The $\not\!\!E_T$ can benefit from the energy correction applied to jets (type1) and additional corrections from energy deposits not clustered into jets (type2). The $\not\!\!E_T$ resolution simulated and measured in multijet QCD events is visible in Fig. 2.23.

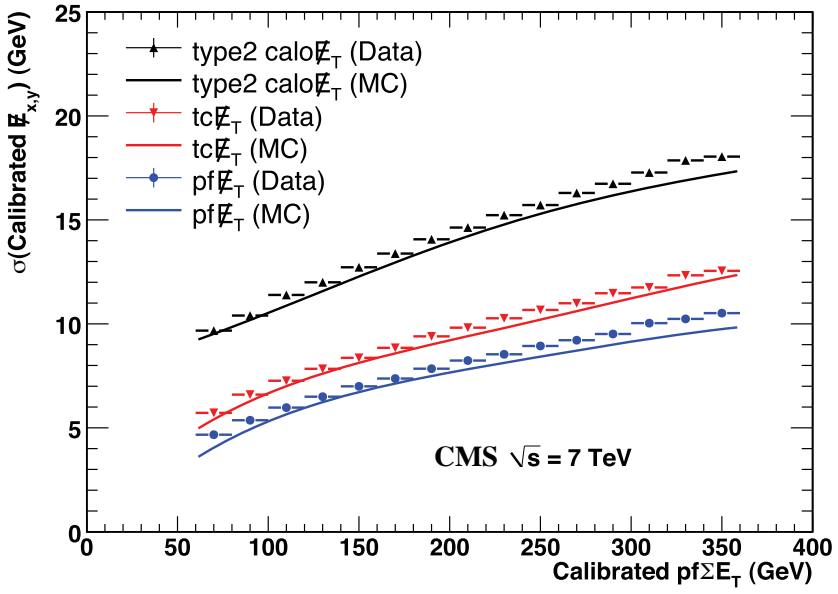


Figure 2.23: Calibrated $\not\!\!E_T$ resolution with respect to the scalar sum of the transverse energies of the PFCandidates for corrected $\not\!\!E_T^{\text{Calo}}$, tcMET and $\not\!\!E_T^{\text{PF}}$ in data and simulation (in minimum bias events containing at least two jets after $\sim 12\,\text{nb}^{-1}$ of integrated luminosity). From [124].

B-tagging algorithm. Several techniques [127] can be used to the identification of jets originating from the decay of a bottom quark, leading to different efficiency and purity compromises and systematic uncertainties. These b-tagging algorithms are based on the long lifetime of the heavy flavour particle or in the presence of soft lepton from a semileptonic decay of the heavy flavour particle.

The different b-tagging algorithms are the "Track Counting" method, based on the number of tracks with a significance of the impact parameter exceeding a given threshold, the "Simple Secondary Vertex" method, based on the reconstruction of at least one secondary vertex, the "jet probability algorithm", based on a discriminant built from the probability of each track to be originating from the primary vertex, and the "soft lepton" method, based on the reconstruction of muons or electrons inside the jet. For each discriminant, three standard operating point have been chosen, "loose", "medium" and "tight", which correspond to 10%, 1% and 0.1% of acceptance of jets from light partons respectively. Additionally, the two first methods are defined in a

high purity or a high efficiency scenarios, which can be used according to the need of a given analysis.

The distributions of input variables [98] used in the different discriminant have been tested in $\sqrt{s} = 7$ TeV data and the simulation have been found generally quite close the observation, which indicates that the ingredients of b-tagging are reasonably well understood.

Chapter 3

Dileptonic Analysis

This chapter describes the measurement of the top pair production ("$t\bar{t}$") cross section in the dileptonic channel after 3.1 pb^{-1} of integrated luminosity at CMS, for a centre-of-mass energy of $\sqrt{s} = 7$ TeV, as published in 2010 [1]. The reference theoretical cross section used in this analysis is $\sigma_{t\bar{t}} = 158^{+23}_{-24}$ pb, computed next-to-leading order for a top mass of 172.5 GeV/c^2, cf. Tab. 1.1. As stated before (cf. Section 1.1.1), the dileptonic $t\bar{t}$ final state is divided in three channels: the ee channel, composed of two opposite sign electrons, the $\mu\mu$ channel, composed of two opposite sign muons and the $e\mu$ channel, composed of two opposite sign leptons, one being an electron, the other being a muon. These electrons and muons are obtained from the direct decays of W bosons or through intermediate tau leptons which have decayed leptonically. These three channels correspond to $\sim$6.5% of the total $t\bar{t}$ cross section, taking into account the contribution of the leptonic decay of intermediate tau leptons. The dileptonic final state, accompanied by two jets and large missing transverse energy, is a clear signature, which suffers from really low background and can allow a first cross-section measurement after only few inverse picobarns of integrated luminosity.

The backgrounds that I have studied in this thesis have been described in Section 1.2.3, and the data and the simulation used in this analysis have been presented in Section 1.2.4. The reconstruction of the events and its different physical objects has been described in Section 2.

This chapter starts with the description of the selection of signal events, for a scenario that I have developed ("TrackJets scenario") and a scenario on which I have made major contributions ("baseline scenario"). Two additional confirmation techniques are also presented. Methods for estimation of background from data are then described.

The next section focuses on the systematic uncertainties study. Based on all these results, I extract the cross section. And finally, a brief survey of the latest results on cross-section determination is made in the last section.

3.1 Selection of the dileptonic $t\bar{t}$ topology

The dileptonic $t\bar{t}$ topology is based on the presence of two leptons coming from W bosons. The general reconstruction of the leptons (Section 2.3.2 and Section 2.4.2) provides a global set of reconstructed candidates usable in all studies. The selection of the reconstructed leptons that procures a sufficiently pure set of leptons coming from W bosons is performed in Section 2.3.3 and 2.4.3. From these selected leptons, event selections can be applied to discriminate the signal of the backgrounds. Two scenarios have been considered, and their selection methods are validated by comparing the observed data and the simulations. Two additional techniques are also presented to confirm that the selected events behave mainly as expected by $t\bar{t}$ events.

Table 3.1: Summary of the lepton selection for the $t\bar{t}$ analysis. For a better readability, the variables refer to CMS jargon, while a detailed description of each cut is available in Section 2.3.3 for muons and 2.4.3 for electrons.

	Muon	Electron
Algorithm	GlobalMuon and TrackerMuon	GSF ECalDriven + $E_T^{\text{calo}} > 10\,\text{GeV}$
Veto	/	$\Delta R(\mu) > 0.1$
p_T	$> 20\,\text{GeV}/c$	$> 20\,\text{GeV}/c$
$\lvert\eta\rvert$	< 2.5	< 2.5
Isolation	< 0.15	< 0.15
d_0	$< 0.02\,\text{cm}$	$< 0.04\,\text{cm}$
γ-conv. cut	/	flag $\neq 0$
Lepton ID	GlobalMuonPromptTight	Fixed threshold WP90
ECal cleaning	/	$S < 0.95$

3.1.1 Event selection with a minimal use of the calorimeters

Originally, our idea of providing an analysis that makes a minimal use of the calorimeters information was motivated by several reasons. A first reason was to provide a simple and consistent analysis without complex methods to combine information from different subdetectors. In this context, the tracker performance was well known thanks to its intensive tests before LHC start-up to justify a study relying only on this subdetector. A second reason was the possibility of recovering runs excluded from the global analysis, due to a malfunctioning of one specific subdetector or a part of it. The use of the tracker offers also the possibility to profit from the vertex position to increase the robustness against pile-up. Finally, this study allows complementary results with different systematic uncertainties with respect to a full detector analysis. The two first motivations have been weakened thanks to the outstanding CMS performance. However, the two last motivations are still valid.

This event selection, called "TrackJets scenario" in this thesis, is divided into two steps. The first one is based on the selected leptons defined in Section 2.3.3 and 2.4.3, summarized in Tab. 3.1. The second one relies on the reconstruction of jets from tracks.

Dileptonic selection. The dileptonic $t\bar{t}$ events are characterized by two opposite sign leptons from the same vertex. A first requirement is then obviously the presence of two selected isolated leptons with opposite charge. It is worth noting that, in this context, calorimeter information is used in lepton identification and isolation. Even if I have studied the possibility of a tracker only isolation, this information is localized around the lepton position, and is then less affected by presence of hot cells or increases of calorimeter activity due to the pile-up than global variable (e.g., $\not\!\!E_T$ or number of CaloJets). The two selected leptons should also be consistent with respect to their associated primary vertex. This selection discards the case of two single leptons processes from two different proton-proton interactions due to the presence of pile-up events at LHC regime. The lepton is associated to the closest good primary vertex (has defined in the vertex reconstruction description Section 2.2.2) in the longitudinal axis within a 1 cm window. If no such vertex is found, the event is rejected. I also reject the event if the two selected leptons have not the same associated primary vertex.

An event with exactly two selected leptons is trivially classified as ee, $\mu\mu$ or $e\mu$ event. In case of more than two selected leptons, the chosen pair of opposite sign leptons is the one that maximizes the scalar sum of transverse momenta. In the following, the

Table 3.2: Expected signal and background yields after the selection of two opposite sign isolated leptons, using only simulation to estimate the different processes, after 3.1 pb^{-1} of integrated luminosity, for the ee channel, the $\mu\mu$ channel, the $e\mu$ channel and for all channels combined. The uncertainties are from the number of simulated events and are small enough to be neglected. The "Data" line gives the number of observed events in data. "S/B" and "$S/\sqrt{B}$" lines give the signal-over-background ratio and the approximated statistical significance from simulation.

Process	ee	$\mu\mu$	$e\mu$	all
Dilepton $t\bar{t}$	3.1	3.5	6.5	13.2
Single top - tW	0.2	0.3	0.5	1.0
VV	0.8	0.9	1.3	3.0
Drell-Yan $ee, \mu\mu$	1088	1256	0.3	2345
Drell-Yan $\tau\tau$	2.8	2.2	5.3	10.3
Non-dilepton $t\bar{t}$	0.1	0.02	0.05	0.2
W+jets	1.1	0.06	1.3	2.5
Total simulated	1096	1263	15	2375
Data	1065	1224	17	2306
S/B	2.8×10^{-3}	2.8×10^{-3}	0.8	5.6×10^{-3}
$S/\sqrt{B}$	0.09	0.10	2.2	0.27

term "selected leptons" refers to these two leptons. Using only simulation to estimate the different processes, the results of this selection are displayed in Tab. 3.2. The dominant background at this step is clearly the Drell-Yan in the ee and $\mu\mu$ channels while the other backgrounds are already of the order of the signal. The reduction of the Drell-Yan in $\tau\tau$ with respect to the Drell-Yan in ee and $\mu\mu$ is due to the leptonic decay branching ratio of the tau lepton, but also of the effect of the transverse momentum cut in the lepton selection. In the $e\mu$ channel, even if its contamination is reduced by two order of magnitude, the Drell-Yan process is still the dominant background. The observed data are statistically compatible with the expectations from simulation.

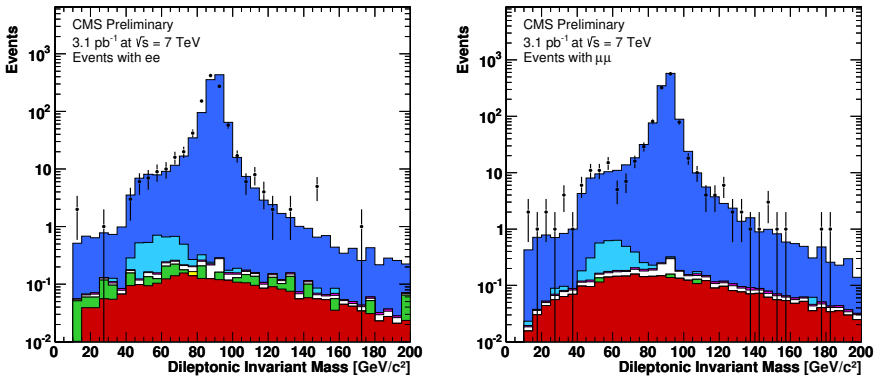
From the two selected leptons, the dileptonic invariant mass can be reconstructed, leading to distributions presented in Fig. 3.1. The two first figures, for the ee and the $\mu\mu$ channels, are similar and show the large peak corresponding to the decay of an on-shell Z^0 boson in the Drell-Yan process. The slight discrepancy in the ee channel between data and simulation concerning the position of the peak is explained by a

Table 3.3: Expected signal and background yields after the dileptonic selection, using only simulation to estimate the different processes, after 3.1 pb^{-1} of integrated luminosity, for the ee channel, the $\mu\mu$ channel, the $e\mu$ channel and for all channels combined. The statistical uncertainties from the number of simulated events are small enough to be neglected. The "Data" line gives the number of observed events in data. "S/B" and "$S/\sqrt{B}$" lines give the signal-over-background ratio and the approximated statistical significance from simulation.

Process	ee	$\mu\mu$	$e\mu$	all
Dilepton $t\bar{t}$	2.4	2.7	6.5	11.6
Single top - tW	0.2	0.2	0.5	0.9
VV	0.5	0.5	1.3	2.3
Drell-Yan $ee, \mu\mu$	94	112	0.3	206
Drell-Yan $\tau\tau$	2.6	2.1	5.3	10.0
Non-dilepton $t\bar{t}$	0.1	0.01	0.05	0.2
W+jets	1.1	0.04	1.3	2.3
Total simulated	101	118	15	234
Data	107	127	17	251
S/B	2.4×10^{-2}	2.3×10^{-2}	0.8	5.2×10^{-2}
$S/\sqrt{B}$	0.24	0.25	2.2	0.78

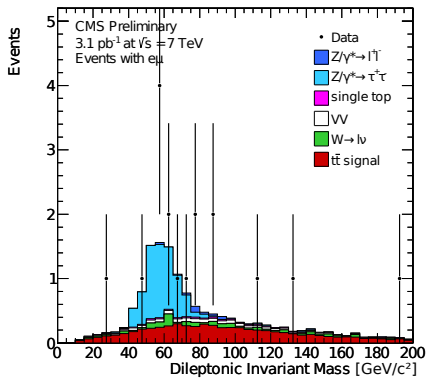
non-optimal calibration of the electromagnetic calorimeter. For the third figure, in the $e\mu$ channel, the Drell-Yan background contributes only via the decay of intermediate tau leptons, which leads to a purer selection with respect to the ee and $\mu\mu$ channels. The events with dileptonic masses below 10 GeV/c^2 correspond to a region containing a negligible part of the signal but possibly affected by fake leptons and processes not taken into account in the simulation (cf. Tab. 1.3), and are then rejected. In the ee and $\mu\mu$ channels, I reject the large Drell-Yan contribution by vetoing the events for which the dileptonic invariant mass is contained between 76 and 106 GeV/c^2 ("Z-veto"). As visible in Tab. 3.3 when compared with Tab. 3.2, this method reduces by $\sim$90% the amount of the Drell-Yan background while the signal is reduced by $\sim$10%, but the contribution of off-peak Drell-Yan events still constitutes the dominant background. The observed data are statistically compatible with the simulated expectations.

Jet selection. In addition to the two isolated leptons, the dileptonic $t\bar{t}$ signature is characterized by the presence of two jets arising from the hadronization of the two



(a)

(b)

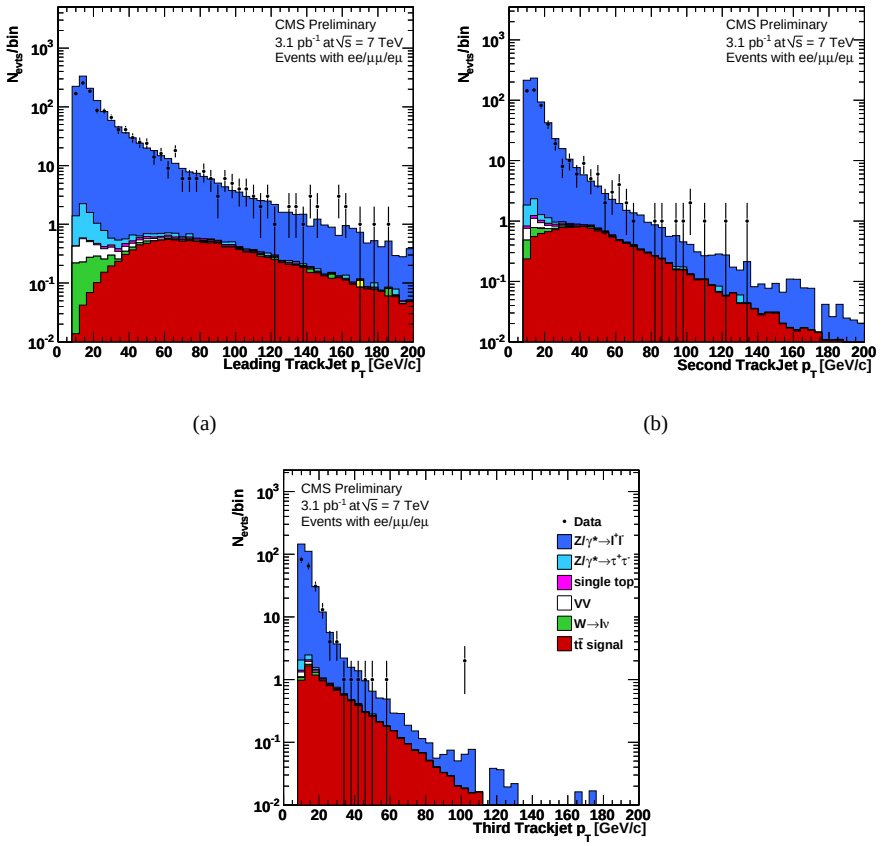


(c)

Figure 3.1: Distributions of the dileptonic invariant mass reconstructed from the two selected leptons, after the selection of two opposite sign isolated leptons, for the ee channel (a), the $\mu\mu$ channel (b) and the $e\mu$ channel (c), after 3.1 pb^{-1} of integrated luminosity. The plain histograms are from simulation only, and the uncertainty from the number of simulated events are small enough to be neglected. The black dots correspond to data, with statistical uncertainty.

bottom quarks. The dominant backgrounds at this stage are not expected to provide large hadronic activity, and the number of jets is then a powerful discriminant.

The jets used in this study are the TrackJets, described in Section 2.5.2 and Appendix A, which do not use any information from calorimeters. They are reconstructed from anti-k_T algorithm with clustering parameter equals to 0.5. The reconstruction algo-



(a) (b)

(c)

Figure 3.2: Distributions of the transverse momentum of the first (a), second (b) and third (c) corrected TrackJets, after the dileptonic selection without the Z-veto, for all channels combined, after 3.1 pb^{-1} of integrated luminosity. The jet selection described in the text is applied, excepted the p_T cut which has been relaxed up to > 10 GeV/c. The plain histograms are from simulation only, and the uncertainty from the number of simulated events are small enough to be neglected. The black dots correspond to data, with statistical uncertainty.

rithm of a TrackJet takes into account only the tracks which are originating from the same primary vertex. Consistently, I only taken into account the TrackJets corresponding to the primary vertex associated to the two selected leptons, avoiding contamination from jets from pile-up. The jet momenta are corrected with L2 and L3 corrections, which compensates the absence of neutral particles. The selected jets are the ones with

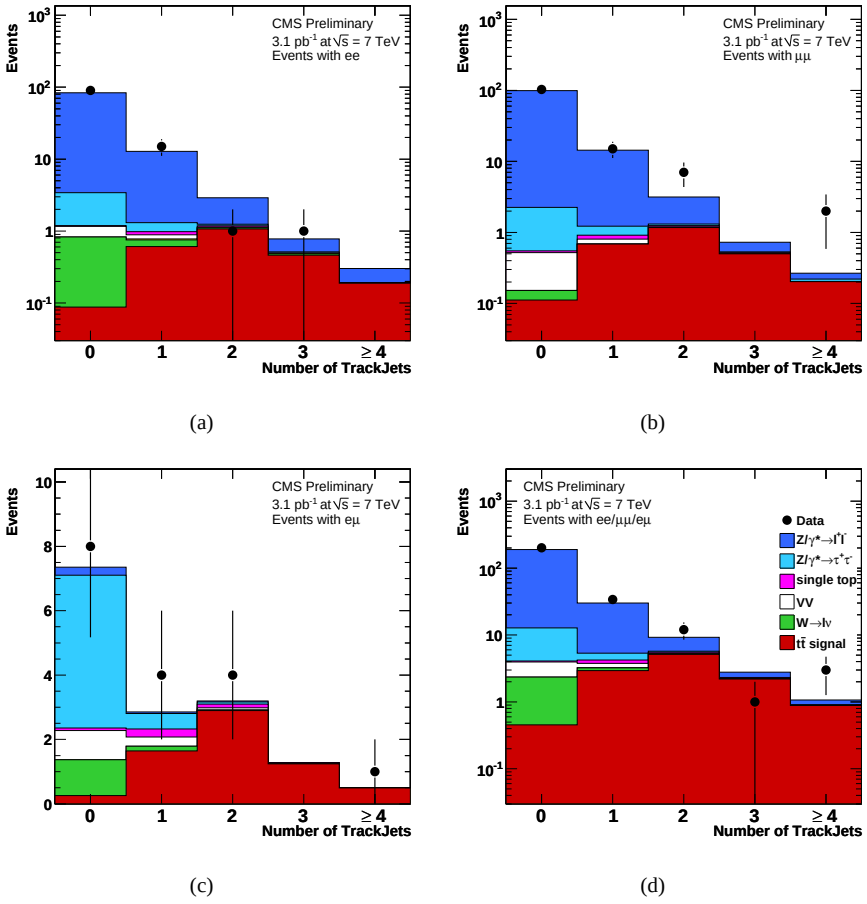
$p_T > 30$ GeV/c and $|\eta| < 2.5$. I apply a cleaning with respect to the two selected leptons: the reconstructed jets close to one selected leptons ($\Delta R < 0.4$) are vetoed.

Table 3.4: Expected signal and background yields after the TrackJets scenario selection, using only simulation to estimate the different processes, after 3.1 pb^{-1} of integrated luminosity, for the ee channel, the $\mu\mu$ channel, the $e\mu$ channel and for all channels combined. The statistical uncertainties from the number of simulated events are small enough to be neglected. The "Data" line gives the number of observed events in data. "S/B" and "$S/\sqrt{B}$" lines give the signal-over-background ratio and the approximated statistical significance from simulation.

Process	ee	$\mu\mu$	$e\mu$	all
Dilepton $t\bar{t}$	1.72	1.87	4.63	8.23
Single top - tW	0.06	0.06	0.15	0.27
VV	0.03	0.03	0.08	0.14
Drell-Yan $ee, \mu\mu$	2.00	2.10	0.03	4.10
Drell-Yan $\tau\tau$	0.08	0.10	0.07	0.24
W+jets	0.06	<0.02	0.02	0.08
Total simulated	4.0	4.2	4.9	13.1
Data	2	9	5	16
S/B	0.75	0.80	17	1.7
$S/\sqrt{B}$	1.1	1.2	8.9	3.7

Fig. 3.2 shows the distributions for the transverse momentum of the first, the second and the third jets by decreasing transverse momentum order for all channels combined, after the dileptonic selection without the Z-veto. The signal distributions peak at ~ 50 GeV/c for the first jet and at ~ 40 GeV/c for the second jet, which corresponds to the presence of the two expected jets from bottom quark hadronization. The low jet production of the background is also visible, justifying a cut at $p_T > 30$ GeV/c. Observed data appear to be in good agreement with expectations from simulation for jets beyond the 30 GeV/c threshold. The events with less than two selected jets are rejected, leading to the yields listed in Tab. 3.4. 98% (resp. 30%) of the expected backgrounds (resp. signal) have been rejected. The expected signal proportion in the total expected events, which was ~ 5% before the jet selection, has increased up to ~ 60% for a total of 13.1 expected events, while 16 selected events are observed in data. The presence of $t\bar{t}$ process is also directly visible in TrackJets multiplicity distributions after the dileptonic selection shown in Fig. 3.3. These distributions illustrate the low jet multi-

plicity of backgrounds while the signal is peaked at two jets, as expected from the two bottom-quark decay.



(a) (b)

(c) (d)

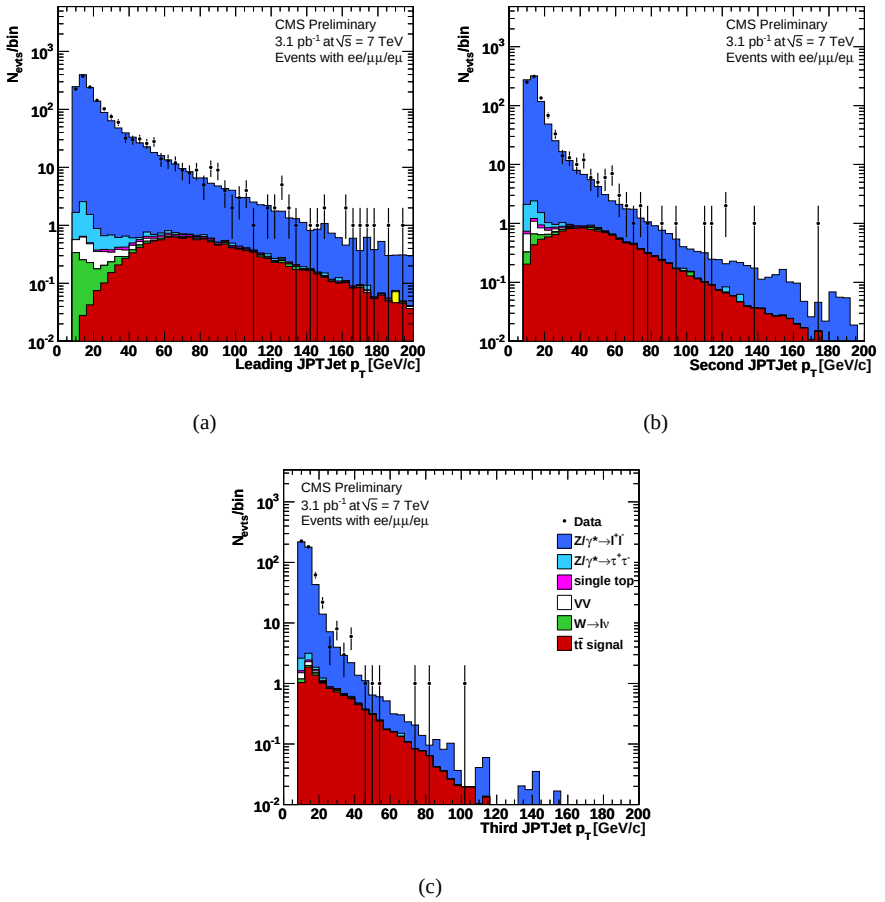
Figure 3.3: Distributions of corrected TrackJet multiplicity for jets with $p_T > 30\,\text{GeV}/c$ after the TrackJets scenario selection without the two jets cut, for the ee channel (a), the $\mu\mu$ channel (b), the $e\mu$ channel (c) and for all channels combined (d), after $3.1\,\text{pb}^{-1}$ of integrated luminosity. The plain histograms are from simulation only, and the uncertainty from the number of simulated events are small enough to be neglected. The black dots correspond to data, with statistical uncertainty.

3.1.2 Event selection with the full CMS detector

As developed in each section of Chapter 2, it has rapidly appeared that the CMS detector behaved outstandingly. This has allowed the use of more complex reconstruction techniques with the whole detector, which can be used in the selection called "baseline scenario" in this thesis. However, several characteristics of an early analysis are maintained. Special care is still applied on the robustness of the selection and the cross-section extraction method is kept as simple as in the TrackJets scenario. With respect to this previous scenario, the baseline study applies exactly the same dileptonic selection. The status of the selection after the dileptonic selection is thus identical to the one developed in the previous section, and allows a direct and easy comparison for the next selection steps. The calorimeters information is then exploited in the jet reconstruction and in the missing transverse energy. B-tagging methods are not used because no significant gain has been observed in simulation.

Jets selection. The jets are reconstructed with the methods described in Section 2.5.2, and results can be obtained using CaloJets, JPTJets and PFJets (the TrackJets being used in the previous section). There is no big difference in the selection efficiency using CaloJets, JPTJets and PFJets in simulations, the choice of objects has then been determined by the resolution and the systematic uncertainties associated to it. CaloJets relies only on calorimeters, which have lower resolution for jets close to the applied p_T threshold. On the other hand, it appears that the understanding of the more complex methods for the JPTJets and PFJets reconstruction is better than expected in the conservative approach adopted before data taking. This leads to good reliability and lower systematic uncertainties. Because there is no big advantage at this stage to favour one kind on the other, the study can be realized for both. But to avoid to present similar results for both choices, I present here the final results done for JPTJets only. These jets are reconstructed with the same parameter than in the TrackJets scenario (the possible effect of the clustering parameter which does not correspond to the same amount of particles at the tracker level and at the calorimeter surface appears to be negligible after correction). They are reconstructed with anti-k_T algorithm with 0.5 clustering parameter, corrected at L2 and L3 levels and the same threshold of $p_T > 30$ GeV/c is applied. A lepton cleaning (the jets with $\Delta R < 0.4$ with respect with one of the two selected leptons are vetoed) and $|\eta| < 2.5$ cuts have also been applied. Unlike the TrackJets where it appears to be not necessary, an additional rejection of non physical or misreconstructed jets can be provided by jet identification criteria. The jet identification criteria required for CaloJets and JPTJets and for PFJets are detailed in Tab. 2.6. Additional corrections are also available for these jets, in particular the small residual correction, adjusting the simulation-based correction to the

real data observations. This correction being small, it has not been applied. The more advanced reconstruction techniques dealing with the pile-up and being consistent with a given primary vertex have not been used in this study. Therefore, no condition on the primary vertex origin of the jets has been applied on the contrary of the TrackJets scenario case.



(a) (b)

(c)

Figure 3.4: Distributions of the transverse momentum of the first (a), second (b) and third (c) corrected JPTJets, after the dileptonic selection without the Z-veto, for all channels combined, after 3.1 pb^{-1} of integrated luminosity. The jet selection described in the text is applied, excepted the p_T cut which has been relaxed up to $>$ 10 GeV/c. The plain histograms are from simulation only, and the uncertainty from the number of simulated events are small enough to be neglected. The black dots correspond to data, with statistical uncertainty.

As done in the TrackJets study, the distribution of the three leading jets can be studied. These distributions are really similar among the different kind of jets. This is illustrated by Fig. 3.4 displaying these distributions for the JPTJets, which can be directly compared with Fig. 3.2. The events with less than two selected jets are rejected.

Missing transverse energy. Due to the leptonic decays of the two W bosons, the dileptonic $t\bar{t}$ contains also neutrinos which are not detected and lead to some missing transverse energy. Again, this characteristic is not expected in dominant backgrounds, and the value of the $\not\!\!E_T$ can be used as a discriminant. The missing transverse energy reconstruction is addressed in Section 2.5.4. Consistently with the choice of the JPTJets and PFJets, the tcMET and the $\not\!\!E_T^{\text{PF}}$ have been used. The $\not\!\!E_T^{\text{Calo}}$ corrected for jet and muon response appears to be less efficient for the Drell-Yan rejection. The background rejection efficiency for tcMET and $\not\!\!E_T^{\text{PF}}$ are similar, while $\not\!\!E_T^{\text{PF}}$ has a slightly better resolution (but that appears to be negligible in the current study).

Table 3.5: Expected signal and background yields after the baseline scenario selection, using only simulation to estimate the different processes, after 3.1 pb^{-1} of integrated luminosity, for the ee channel, the $\mu\mu$ channel, the $e\mu$ channel and for all channels combined. The statistical uncertainties from the number of simulated events are small enough to be neglected. The "Data" line gives the number of observed events in data. "S/B" and "$S/\sqrt{B}$" lines give the signal-over-background ratio and the approximated statistical significance from simulation.

Process	ee	$\mu\mu$	$e\mu$	all
Dilepton $t\bar{t}$	1.50	1.68	4.48	7.65
Single top - tW	0.05	0.05	0.15	0.25
VV	0.03	0.03	0.08	0.13
Drell-Yan $ee, \mu\mu$	0.14	0.28	0.01	0.43
Drell-Yan $\tau\tau$	0.04	0.07	0.07	0.18
Non-dilepton $t\bar{t}$	0.05	0.01	0.09	0.15
W+jets	0.03	$<$0.01	0.06	0.09
Total simulated	1.8	2.1	4.9	8.9
Data	3	3	5	11
S/B	5.0	4.0	10.7	6.1
$S/\sqrt{B}$	2.7	2.6	6.9	6.8

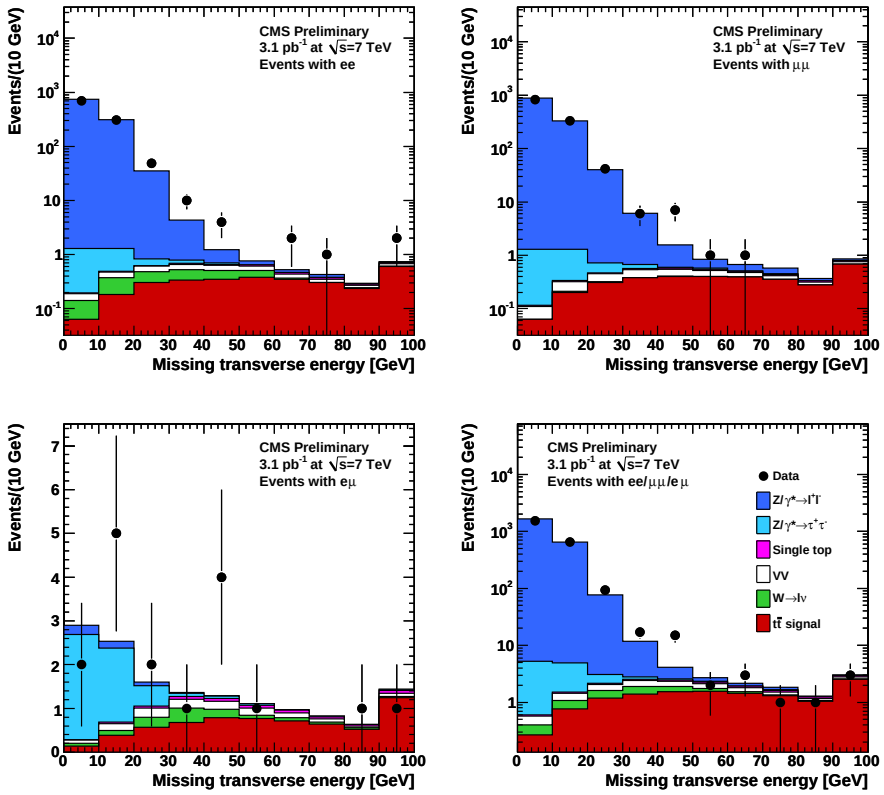
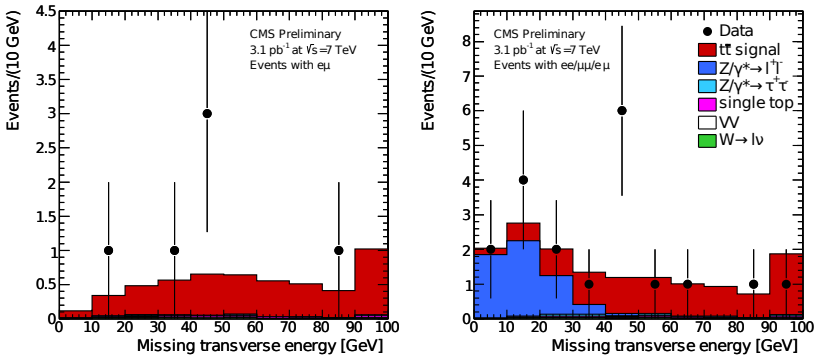


Figure 3.5: Distributions of the tcMET after the dileptonic selection without the Z-veto, for the ee channel (a), the $\mu\mu$ channel (b), the $e\mu$ channel (c) and for all channels combined (d), after 3.1 pb^{-1} of integrated luminosity. The plain histograms are from simulation only, and the uncertainty from the number of simulated events are small enough to be neglected. The black dots correspond to data, with statistical uncertainty. The last bin contains overflow events.

Fig. 3.5 shows the distributions for tcMET in the different channels after the dileptonic selection without Z-veto. Some discrepancies can be observed for values around 40 GeV, in ee and $\mu\mu$ channels. It seems to be an excess of Drell-Yan events with respect to expectation, because no pathologies have been identified in missing energy or jet energy estimation and because the excess events are consistent with Drell-Yan



(a) (b)

Figure 3.6: Distributions of the tcMET after the baseline scenario selection without the $\not\!\!E_T$ cut (with corrected JPTJets), for the $e\mu$ channel (a) and for all channels combined (b), after 3.1 pb^{-1} of integrated luminosity. The plain histograms are from simulation only, and the uncertainty from the number of simulated events are small enough to be neglected. The black dots correspond to data, with statistical uncertainty. The last bin contains overflow events.

(with respect to the light flavour characteristics of the jets or the dileptonic invariant mass shape). A method to estimate the amount of Drell-Yan background from data has been developed, cf. Section 3.2.1, which avoids thus any effect of this discrepancy on the cross section determination. Fig. 3.6 shows the distributions for tcMET in the $e\mu$ channel and for all channels combined after the dileptonic selection and the jet selection. Because the Drell-Yan background is more important in the ee and $\mu\mu$ channels, the cut on the $\not\!\!E_T$ is harder: events with $\not\!\!E_T$ below 30 GeV are rejected. For the $e\mu$ channel, even if it appears to be unnecessary from simulation study, a missing transverse energy cut is also applied, in order to reject possible unexpected data events which do not have the kinematical characteristics of $t\bar{t}$ events: events with $\not\!\!E_T$ value below 20 GeV are rejected. The events yields after this selection step is given in Tab. 3.5. The expected signal proportion in the total expected events is $\sim$ 86% for a total of 8.9 expected events, while 11 selected events are observed in data. 8 events are selected in both TrackJets and baseline scenarios. The non-common events are the consequence of the absence of $\not\!\!E_T$ cut in the TrackJets scenario (6 events are kept in the TrackJets scenario while they do not satisfy the tcMET cut, which is consistent with the simulation expectation) and of the different resolution of the jets in the two scenarios, where the second jet can be close to the threshold (2 (resp. 3) events are selected in the TrackJets scenario and not in the baseline scenario (resp. in the baseline

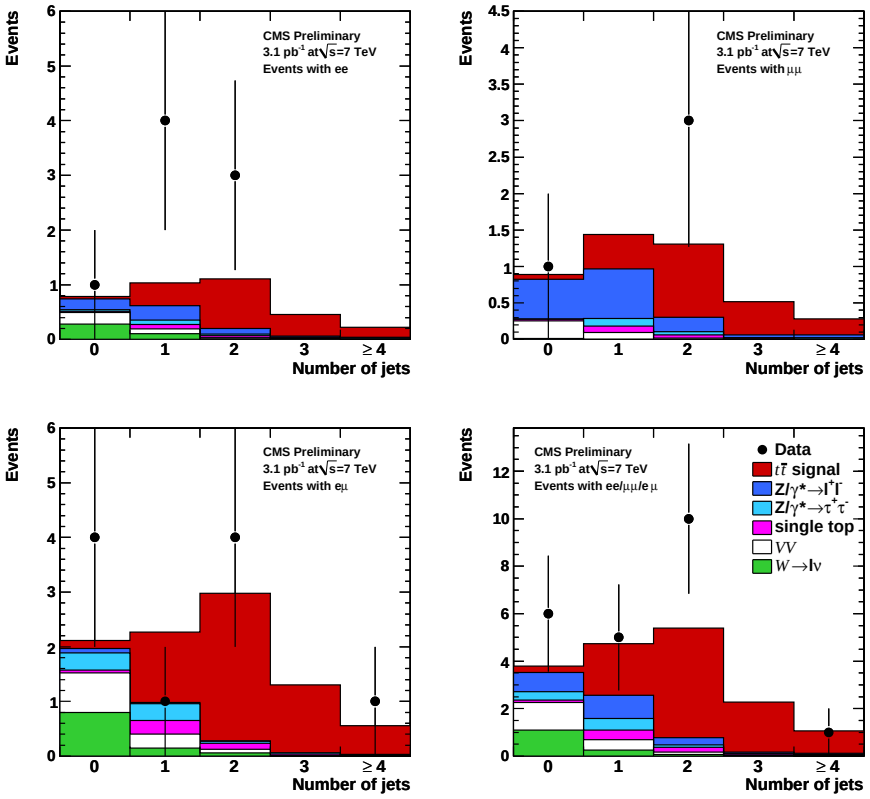


Figure 3.7: Distributions of corrected JPTJet multiplicity for jets with $p_T > 30$ GeV/c after the baseline selection without the two jets cut, for the ee channel (a), the $\mu\mu$ channel (b), the $e\mu$ channel (c) and for all channels combined (d), after 3.1 pb^{-1} of integrated luminosity. The plain histograms are from simulation only, and the uncertainty from the number of simulated events are small enough to be neglected. The black dots correspond to data, with statistical uncertainty.

scenario and not in the TrackJets scenario) due to this effect). The JPTJets multiplicity after the dileptonic selection and the missing transverse energy is shown in Fig. 3.7.

3.1.3 Additional $t\bar{t}$ characterizations

The previous selections allow the creation of a pure $t\bar{t}$ sample. But two other characteristics of the $t\bar{t}$ events can also be exploited to increase the confidence on the fact that the selected events are $t\bar{t}$ events. The first one is the fact that the two jets are obtained from bottom quarks, which can be discriminated with respect to other jets from lighter quarks or from gluons using b-tagging algorithms. The second characteristic is the fact that the top-quark mass measurement techniques for dileptonic $t\bar{t}$ return a more suitable observed mass values for $t\bar{t}$ events than for background events.

These two characteristics are not suited for an early analysis. The gain after b-tagging selection, studied in simulation, appeared to be small: the signal over background ratio is enhanced, but the statistical uncertainty, which is the dominant uncertainty, is also increased, and the systematic uncertainty from the b-tagging (large due to the lack of calibration from data at the time of this analysis) has also to be taken into account. This explains why no b-tag is applied in the baseline scenario despite its good performance in data. Concerning the top-quark mass extraction, these complex methods have to be carefully applied and should benefit of a larger data sample. Two preliminary methods have been exploited to extract a value related to the top-quark mass. It is worth noting that these methods were not fully optimized for a mass measurement at the time of this early study and that the presented results are not to be taken as a real mass measurement but more as an additional discriminant to confirm the signal presence. These methods have later been improved and applied to a larger amount of data [128].

B-tagging. The b-tag algorithm used is the track-counting tagger loose working point (cf. Section 2.5.4). The jet is identified as a b-jet if the impact parameters of at least two tracks satisfies the loose working point requirement. This loose working point has been chosen in order to reduce the rejection of signal events. This tagger has an 80% efficiency on b-jets and suffers of 10% of fake rate from jets not originating from bottom quark. The distribution of the number of b-tagged jets is shown in Fig. 3.8. The uncertainties considered here are only determined from the scale factor observed in data [127], and do not account for the correlation between bins and the jets momentum dependency. Even if the statistical uncertainty is large, the data show indeed a tendency to contain one or two b-tagged jets, as it is expected from $t\bar{t}$ events from simulation.

Mass measurement methods. Because the top-quark measurement in itself is beyond the scope of this thesis, the methods are only briefly described, no uncertainties

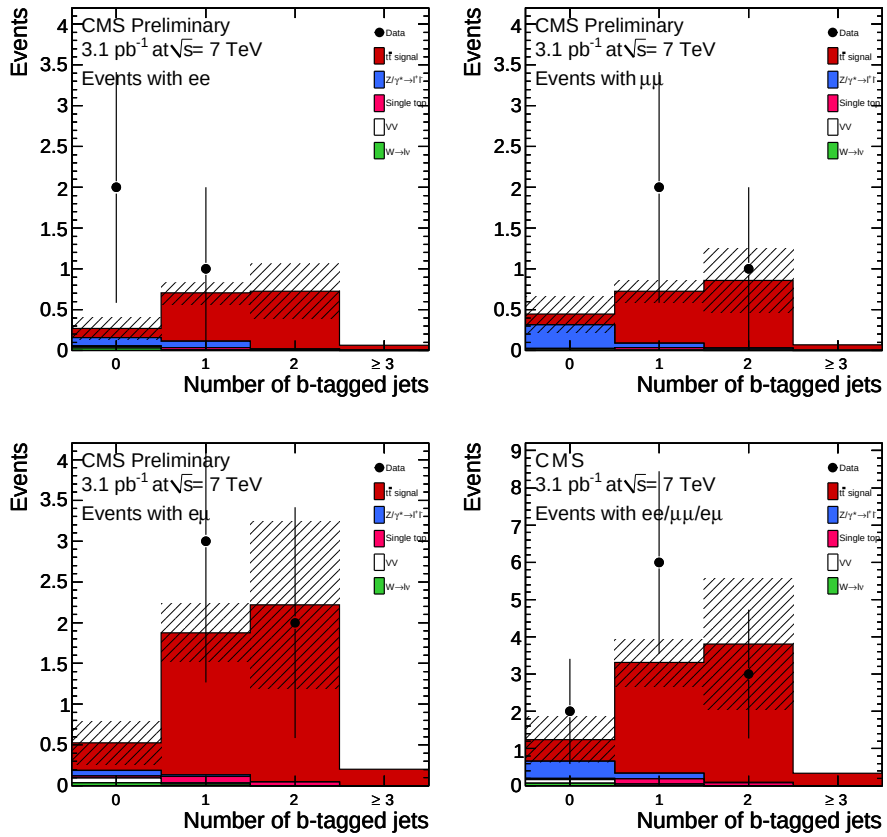


Figure 3.8: Distributions of b-tagged JPTJet multiplicity for jets with $p_T > 30$ GeV/c and track-counter loose working point b-tagger, after the baseline scenario selection, for the ee channel (a), the $\mu\mu$ channel (b), the $e\mu$ channel (c) and for all channels combined (d), after 3.1 pb^{-1} of integrated luminosity. $Z \to \ell\ell$ contains $Z \to \tau\tau$. The plain histograms are from simulation only, and the uncertainty from the number of simulated events are small enough to be neglected. The shaded area displays the uncertainty on the number of signal events corresponding to the uncertainty in the b-tag selection, based on the b-tagging knowledge for early data. The black dots correspond to data, with statistical uncertainty. From [1].

study is developed and no measured top-quark mass value are unveiled. Two methods have been tested, and the consistent results between the two strengthen the confidence in these methods.

The "Kinematic Method" has been developed and used at the CDF experiment [129]. The dileptonic $t\bar{t}$ topology has six particles in its final states: two charged leptons, two neutrinos and two bottom quarks. Each particle being described by four quadrimomentum components, there are twenty-four independent parameters (including known masses). The eight parameters corresponding to the charged leptons (P_{ℓ^-} and P_{ℓ^+}) are directly measured. The eight parameters corresponding to the bottom quarks (P_b and $P_{\bar{b}}$) are given by the two leading jets of the events, even if non negligible uncertainty on these values is introduced by the detector resolution. The six parameters corresponding to the neutrinos, supposed to be massless, are constrained by the measurement of the missing transverse energy, $\vec{\not{E}}_T = \vec{p}_{T,\nu_1} + \vec{p}_{T,\nu_2}$. Again, large uncertainty due to the resolution of the detector has to be taken into account. The world average value for the W boson mass is assigned to the two W bosons being produced on-shell in the top decay and the condition that both top and antitop quark have the same mass, $m_t = m_{\bar{t}}$, is also imposed. These measurements and constraints lead to a system with one degree of freedom. This extra degree of freedom is related to the longitudinal balance with which the $t\bar{t}$ system is produced described by the variable p_z:

$$p_z = p_{z,t} + p_{z,\bar{t}} = p_{z,b_1} + p_{z,\ell_1} + p_{z,\nu_1} + p_{z,b_2} + p_{z,\ell_2} + p_{z,\nu_2}, \tag{3.1}$$

which is not expected to have a strong dependence in the top-quark mass. Assuming the p_z value from the Gaussian distribution obtained from simulation, the system is then totally constrained for each of the two lepton-jet assignment. Varying the jets and missing transverse energy values according to their reconstruction resolutions and varying the p_z value according to the simulated Gaussian distribution, all solutions for which $|m_{t_1} - m_{t_2}| < 3\,\text{GeV}/c^2$ are accepted. The lepton-jet assignment chosen for the mass measurement is the one that provides most of the solutions. The equations been quadratic, more than one solution might be found for a given lepton-jet assignment and a given smearing configuration. It has been found that the solution corresponding to the softer value of $[P_{b_1} + P_{\ell_1} + P_{\nu_1}]^2 + [P_{b_2} + P_{\ell_2} + P_{\nu_2}]^2$ reflects better the true kinematics of the $t\bar{t}$ events. This solution is then chosen such that there is at most one solution for each smearing configuration. For a given event, the final top mass value is extracted from a Gaussian fit on the $(m_{t_1} + m_{t_2})/2$ distribution.

The "Matrix Weighting Algorithm" has been developed and used at the DØ experiment [130]. It uses, for each event, the solutions kinematically consistent with a dileptonic $t\bar{t}$ for a given top-quark mass hypothesis. Each of these solutions are then weighted by the matrix element from the two leptons momenta for a top-quark mass hypothesis. Repeating the process for each top-quark mass hypothesis and summing the weights of each solution leads to a distribution of weights with respect to the top-

quark mass. The higher weight mass value is assigned to the event, and finally, the masses distribution is fit with simulated templates to extract the top-quark mass.

The kinematic event reconstruction is based on the measured leptons and jets variables and on the total transverse momentum of all other jets and particles ($\vec{p}_{T,t\bar{t}}$), which corresponds to the transverse momentum of the top quarks pair.

Using the W boson and top-quark masses, the positively charged lepton momentum and the associated bottom quark momentum allows the reduction of the top quark transverse momentum ($\vec{p}_{T,t}$) to an ellipse in the $p_x - p_y$ plane (and inversely for the negatively charged lepton and the antibottom quark for the antitop quark). Knowing that $\vec{p}_{T,t} + \vec{p}_{T,\bar{t}} = \vec{p}_{T,t\bar{t}}$, the solutions for the two top quarks transverse momenta correspond to the intersection of the ellipses defined for $\vec{p}_{T,t}$ and $\vec{p}_{T,t\bar{t}} - \vec{p}_{T,\bar{t}}$. The ellipses intersection leads to a maximum number of solution of four, which is doubled due to the ambiguity of the lepton-jet association. Due to the detector resolution and the effect of hard gluon radiation, the jets momenta do not always exactly correspond to the bottom quarks momenta and lead to system with no solution (in $\sim 15\%$ of the cases). This effect is compensated by computing solutions many times per event (1 000 for data events, 100 for simulation events to save CPU time) with smeared jets momenta. The smearing is a normal distribution centred on the measured value with a width corresponding to the expected detector resolution. In summary, for each event, for each top quark mass hypothesis and for each jets smearing configuration, a maximum of four solutions are found when the positively charged lepton is associated to the leading jet, and a maximum of four solutions are found when the positively charged lepton is associated tot the second leading jet. For each top quark mass hypothesis, the total number of solution for one event is given by the sum of all the solutions for each jets smearing configuration.

In order to emphasize the most probable kinematical solutions for a m_t top quark mass hypothesis, a matrix element weight w using the two leptons momenta is computed:

$$w = \sum F(x_1)F(x_2)p(E^{\star}_{l^+}|m_t)p(E^{\star}_{l^-}|m_t) \tag{3.2}$$

where x_1 and x_2 are the Björkén x values for the initial state partons, $F(x)$ is the parton distribution function evaluated at $Q^2 = m_t^2$. The sum is done over the possible leading order initial state partons ($u\bar{u}$, $\bar{u}u$, $d\bar{d}$, $\bar{d}d$ and gg). $P(E^{\star}|m_t)$ is the probability density of observing a charged lepton of energy $E^{\star}$ in the rest frame of the top quark.

$$p(E^{\star}|m_t) = \frac{4m_t E^{\star}(m_t^2 - m_b^2 - 2m_t E^{\star})}{(m_t^2 - m_b^2)^2 + M_W^2(m_t^2 - m_b^2) - 2M_W^4} \tag{3.3}$$

It leads to a weight curve distribution with respect to the top mass hypothesis, which has been evaluated for 100 to 400 GeV/c^2 by 1 GeV/c^2 step. The value of the mass

for which the weight is the higher, called “peak mass”, is used as measured top quark mass before correction, which is biased by the simplified assumptions in the reconstruction but strongly correlated with the top mass.

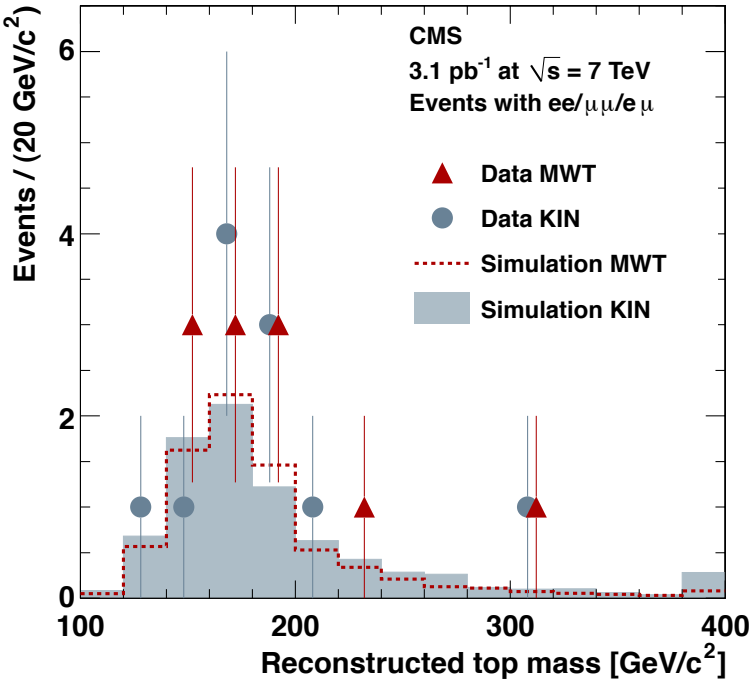


Figure 3.9: Distributions of the top mass values extracted with the Matrix Weighting Algorithm method (WMT) and the Kinematic method (KIN) detailed in the text, for the eleven events (dots) passing the baseline scenario selection and for simulation (dashed line and plain histogram), after 3.1 pb^{-1} of integrated luminosity. The last bin contains overflow. From [1].

The top quark masses distributions, obtained from the eleven data events passing the selection, and the expected distributions from simulations, are given in Fig. 3.9, for the two methods. The top quark mass can then be extracted by comparing these masses distributions to simulated distribution templates for a range of top quark masses hypothesis. The background contribution can also be taken into account using templates from simulation. These templates can be normalized from simulation, from data-driven background estimation or as a free parameter in the fit procedure.

3.2 Data-driven background estimate

Even if the simulation appears to describe well the data in most of the cases, expectations induced by real data are preferred, especially for the dominant backgrounds or for behaviour depending of the tails of distributions, such as the probability of a non-prompt lepton to pass the selection optimized to select the prompt leptons. In order to reduce this dependence of the simulation, I use two methods [131] to determine directly from data the amount of backgrounds:

- The Drell-Yan estimation from data, which uses the large amount of Drell-Yan events in the Z mass peak in the invariant mass distribution in order to estimate the amount of Drell-Yan events in the selected region;
- The estimation of fake leptons, which uses the probability of a fake lepton to pass the lepton selection, obtained from data, to estimate the contamination of these fake leptons into the selection.

The remaining backgrounds are the single-top tW, the diboson processes not contributing to the Z mass peak and the Drell-Yan into two tau leptons. These backgrounds are small with respect to the others and are therefore predicted with simulation with a conservative uncertainty on their normalization (50%).

I apply both data-driven background estimate techniques to the TrackJets scenario and the baseline scenario.

3.2.1 Estimation of Drell-Yan background

The characteristic peak corresponding to the Z^0 boson mass in the dileptonic invariant mass distribution is rejected in the analysis, but can be used as a control region to evaluate the number of Drell-Yan events selected. Assuming that the simulated shape of the Drell-Yan dileptonic invariant mass distribution is reliable, the scale factor between the amount of observed Drell-Yan events in the peak region and the predicted events in simulation can be use to rescale the amount of predicted events outside the peak region. In this context, the peaking background processes have to be distinguished from the non-peaking processes. The peaking background are the processes containing a Z^0 boson decaying into two electrons or two muons, which is the case of the Drell-Yan and diboson processes. This is shortened in DY in the notation, due to the dominance of the Drell-Yan contribution.

The scale factor $R_{out/in}$ is first determined in simulation after dileptonic selection without Z-veto by comparing the number of DY background events outside the Z

window ($N^{out}_{DY\ MC}$) and the number of DY background events inside the Z window ($N^{in}_{DY\ MC}$):

$$R_{out/in} = \frac{N^{out}_{MC\ DY}}{N^{in}_{MC\ DY}} \tag{3.4}$$

The estimated number of selected events for these DY processes can then be estimated from data, using the number of these DY processes in the Z window region:

$$N^{out}_{estimated\ DY} = N^{in}_{obs\ DY} \cdot R_{out/in} \tag{3.5}$$

Of course, the number $N^{in}_{obs\ DY}$ is not directly known, due to the presence of other processes. To evaluate these processes, the channel $e\mu$, free of DY background, is considered. With respect to the $e\mu$ channel, the non-DY processes yield in the $\mu\mu$ channel incurs a factor 1/2, but has also to be corrected for the difference between the electron and the muon selection efficiencies: ϵ_μ/ϵ_e. This efficiency correction factor can be extracted from the number of dileptonic events obtained from a loose selection (events passing the dileptonic selection without Z-veto in the TrackJets scenario, and the same selection plus the jet selection without $\not{E}_T$ for the baseline analysis), assuming a same geometrical acceptance ($A_{\ell\ell}$) for both channels:

$$\frac{N^{obs\ loose}_{\mu\mu}}{N^{obs\ loose}_{ee}} = \frac{N^{true}_{\mu\mu}}{N^{true}_{ee}} \cdot \frac{A_{\mu\mu}}{A_{ee}} \cdot \frac{\epsilon^2_\mu}{\epsilon^2_e}\ ; \tag{3.6}$$

The correction factor $k_{\mu\mu}$ is therefore:

$$k_{\mu\mu} = \frac{1}{2}\sqrt{\frac{N^{obs\ loose}_{\mu\mu}}{N^{obs\ loose}_{ee}}} \tag{3.7}$$

which can be evaluated in an identical way for the ee case.

This leads to the final equation of the estimation of DY events outside the peak window for a $\ell\ell$ channel:

$$N^{out}_{estimated\ DY\ \ell\ell} = (N^{in}_{obs\ \ell\ell} - k_{\ell\ell} \cdot N^{in}_{obs\ e\mu}) \cdot R^{\ell\ell}_{out/in} \tag{3.8}$$

The results in the TrackJets scenario are presented in Tab. 3.6. The estimate of the Drell-Yan contribution ($N^{out}_{estimated\ DY}$) can be compared to the total number of data events ($N^{out}_{total\ data}$) which also contain non Drell-Yan contributions. For $N_{jets} = 0$ and $N_{jets} = 1$ bins where the Drell-Yan process dominates, it shows a consistent results. The $N_{jets} \geq 2$ bin provides the final number of the estimation of the Drell-Yan used in the cross-section determination. The closure test performed with simulations is also shown. I have evaluated the systematic uncertainty on the background

Table 3.6: Results of the data-driven estimation of the Drell-Yan-like background to ee and $\mu\mu$ channels in the TrackJets scenario, scaled to 3.1 pb^{-1} of integrated luminosity, in function of the TrackJet multiplicity. $R_{out/in}$ is the ratio of Drell-Yan-like background events in and out the Z^0 peak region evaluated by simulation. $k_{\ell\ell}$ is the correction factor for the lepton efficiency evaluated in data. $N^{out}_{estimated\ DY}$ is the estimation of the Drell-Yan contamination in the out region. $N^{out}_{total\ data}$ is the total number of events observed in the out region, containing also the non Drell-Yan data. $N^{out}_{MC\ DY\ est.}$ is the estimation of the Drell-Yan contamination in the out region using only simulated data. $N^{out}_{MC\ DY}$ is the expected Drell-Yan-like contamination in the out region in simulation. Uncertainties are based on the used statistic.

ee channel	$N_{jets} = 0$	$N_{jets} = 1$	$N_{jets} \geq 2$
$R_{out/in}$	0.093 ± 0.002	0.104 ± 0.002	0.102 ± 0.002
k_{ee}	0.465 ± 0.029	0.470 ± 0.082	0.463 ± 0.179
$N^{out}_{estimated\ DY}$	77.912 ± 3.138	10.422 ± 1.062	2.041 ± 0.458
$N^{out}_{total\ data}$	90	15	2
$N^{out}_{MC\ DY\ est.}$	80.287 ± 3.198	11.673 ± 1.127	2.118 ± 0.467
$N^{out}_{MC\ DY}$	80.259 ± 1.056	11.645 ± 0.377	2.063 ± 0.166
$\mu\mu$ channel	$N_{jets} = 0$	$N_{jets} = 1$	$N_{jets} \geq 2$
$R_{out/in}$	0.098 ± 0.002	0.105 ± 0.002	0.088 ± 0.002
$k_{\mu\mu}$	0.537 ± 0.036	0.532 ± 0.098	0.540 ± 0.223
$N^{out}_{estimated\ DY}$	90.875 ± 3.540	14.709 ± 1.281	2.367 ± 0.458
$N^{out}_{total\ data}$	103	15	9
$N^{out}_{MC\ DY\ est.}$	97.143 ± 3.697	13.311 ± 1.215	2.161 ± 0.438
$N^{out}_{MC\ DY}$	97.130 ± 1.181	13.284 ± 0.435	2.108 ± 0.174

yield estimate from this method at 25%, based on discrepancies between inclusive and $N_{jets} \geq 2$ bins (15%) and in the $N_{jets} = 0$ bin (25%). I have also evaluated the effect of possible discrepancies between the real and the simulated Drell-Yan process shape by varying the window edges. A 2 GeV/c^2 shift of the two boundaries in the same direction, corresponding to a shift of the Z peak, leads to a $\sim$6% variation of the estimated background yield. A shift of the two boundaries in the opposite direction, corresponding to a wider or a narrower Z peak, leads to a higher effect of $\sim$12%. The accuracy of the Z peak simulation can be evaluated in Drell-Yan dominated region, and the possible effect is contained in the conservative uncertainty associated to this method.

Table 3.7: Results of the data-driven estimation of the Drell-Yan-like background to ee and $\mu\mu$ channels in the baseline scenario, scaled to 3.1 pb^{-1} of integrated luminosity, in function of the JPTJet multiplicity, with and without the missing transverse energy selection. $R_{out/in}$ is the ratio of Drell-Yan-like background events in and out the Z^0 peak region evaluated by simulation. $N^{out}_{estimated\ DY}$ is the estimation of the Drell-Yan contamination in the out region. $N^{out}_{total\ data}$ is the total number of events observed in the out region, containing also the non Drell-Yan data. $N^{out}_{MC\ DY}$ is the expected Drell-Yan-like contamination in the out region in simulation. Uncertainties are based on the used statistic.

without $\not\!\!E_T$ selection:

ee channel	$N_{jets} = 0$	$N_{jets} = 1$	$N_{jets} \geq 2$
$R_{out/in}$	0.092	0.11	0.11$\pm$ 0.01
$N^{out}_{estimated\ DY}$	76 ± 3	12 ± 1	2.9 ± 0.6
$N^{out}_{total\ data}$	87	15	6
$N^{out}_{MC\ DY}$	79	13	2.6 ± 0.1
$\mu\mu$ channel	$N_{jets} = 0$	$N_{jets} = 1$	$N_{jets} \geq 2$
$R_{out/in}$	0.096	0.12	0.11 ± 0.01
$N^{out}_{estimated\ DY}$	88 ± 3	17 ± 1	3.8 ± 0.7
$N^{out}_{total\ data}$	99	21	7
$N^{out}_{MC\ DY}$	96	15	2.9 ± 0.1

with $\not\!\!E_T$ selection:

ee channel	$N_{jets} = 0$	$N_{jets} = 1$	$N_{jets} \geq 2$
$R_{out/in}$	0.26 ± 0.05	0.17 ± 0.03	0.10 ± 0.02
$N^{out}_{estimated\ DY}$	$0.01\ ^{+0.3}_{-0.01}\ ^{+0.2}_{-0.01}$	$0.3 \pm 0.2 \pm 0.2$	$0.8 \pm 0.4 \pm 0.4$
$N^{out}_{MC\ DY}$	0.20 ± 0.04	0.27 ± 0.05	0.14 ± 0.03
$\mu\mu$ channel	$N_{jets} = 0$	$N_{jets} = 1$	$N_{jets} \geq 2$
$R_{out/in}$	0.44 ± 0.06	0.28 ± 0.03	0.15 ± 0.03
$N^{out}_{estimated\ DY}$	$0.0\ ^{+0.3}_{-0.0}\ ^{+0.2}_{-0.0}$	$1.7 \pm 0.7 \pm 0.8$	$0.6 \pm 0.4 \pm 0.3$
$N^{out}_{MC\ DY}$	0.54 ± 0.07	0.68 ± 0.07	0.28 ± 0.05

The results in the baseline scenario are presented in Tab. 3.7. The systematic uncertainty for this method is dominated by the dependence of $R_{out/in}$ on event selection, mainly on the $\not\!\!E_T$ cut and slightly on the overall scale of the event. The systematic

uncertainty is then evaluated at 50%, based on the differences between the results with jets with and without $\not\!\!E_T$ cut, which is conservative in the $N_{jets} \geq 2$ bin.

3.2.2 Estimation of fake leptons

The multijet QCD processes, the W+Jets process and the semileptonic $t\bar{t}$ process can contaminate the signal selection if some reconstructed and selected leptons are in fact non-prompt from W boson. These leptons are called "fake leptons" and regroup the reconstructed leptons that do not correspond to a real lepton but also the leptons from heavy flavour decays, decays in flight of mesons or photon conversions. These leptons pass the lepton selection if they are located in the tails of the discriminant distributions for which the simulation is less reliable.

In order to evaluate the rate of fake lepton passing the lepton selection, the fake rate (FR) or tight-to-loose ratio (R_{TL}, or R as used in the following formulae) is determined [132]. The previously defined lepton selection is called "tight selection" and a second selection, looser than the tight selection, the "loose selection" is defined. The fake rate is defined as the ratio R, obtained for different p_T and η regions, between the number of leptons passing the tight selection and the number of leptons passing the loose selection.

The loose selection for muons consists of relaxing the cut on the normalized χ^2 of the global fit of the track to < 50, the cut on the transverse impact parameter with respect to the beam-spot to < 2 mm and the cut on the combined isolation to < 0.4. The loose selection for electrons consists of removing the electron identification and the cut on the impact parameter with respect to the beam-spot.

The fake rate is evaluated from a multijet QCD sample. It is assumed that the number of leptons in the event does not affect the fake rate and that the evaluation of the fake rate in multijet QCD sample corresponds to the fake rate in the signal region. The initial idea was to obtain a QCD sample triggered by jets. Unfortunately, the low p_T jet triggers are heavily prescaled, which leads to insufficient statistics. Single lepton triggers have been used instead, to obtain a QCD sample populated with lepton, and a jet trigger has been simulated by an offline jet requirement (at least one jet with $p_T^{uncorr} > 15$ GeV/c). If there is no lepton identification or isolation requirements in the lepton trigger (which is the case here), this leads to a similar sample than the one that should be obtained by asking a jet trigger and an offline lepton requirement. In case of events with one lepton and one jet, these two objects should be well separated ($\Delta R > 1$) to avoid trigger bias. The lepton considered for the fake rate evaluation is the one corresponding to the lepton trigger ($\Delta R < 0.4$ between the offline and the online lepton). In this sample, there is a contamination of prompt leptons, mainly

from W decays. This contamination can be reduced by requiring low missing energy (tcMET < 20 GeV) and a low transverse mass ($M_T < 25$ GeV), which has no significant impact on results from multijet QCD process. The W contamination is however still significant for leptons with p_T higher than 35 GeV/c, especially for muons. To avoid this and because the leptons in multijet QCD background are expected to populate low p_T region, the fake rate is evaluated up to $p_T = 35$ GeV/c, and this last value is used as the fake rate for leptons with higher p_T. The obtained values of the fake rate for muons and electrons for different p_T and $|\eta|$ bins are given in Tab. 3.8 and Tab. 3.9.

Table 3.8: Fake rate (R) values for muons obtained for different η and p_T bins from the method discussed in the text.

muons	$0. \leq \lvert\eta\rvert < 1.$	$1. \leq \lvert\eta\rvert < 1.479$	$1.479 \leq \lvert\eta\rvert < 2.$	$2. \leq \lvert\eta\rvert \leq 2.5$
$10 < p_T \leq 15$	0.320 ± 0.004	0.357 ± 0.007	0.352 ± 0.007	0.381 ± 0.012
$15 < p_T \leq 20$	0.246 ± 0.009	0.257 ± 0.014	0.306 ± 0.014	0.345 ± 0.025
$20 < p_T \leq 25$	0.241 ± 0.017	0.289 ± 0.030	0.255 ± 0.026	0.253 ± 0.048
$25 < p_T \leq 30$	0.206 ± 0.026	0.218 ± 0.047	0.340 ± 0.049	0.147 ± 0.061
$30 < p_T \leq 35$	0.243 ± 0.051	0.143 ± 0.059	0.343 ± 0.080	0.188 ± 0.098

Table 3.9: Fake rate (R) values for electrons obtained for different η and p_T bins from the method discussed in the text.

electrons	$0. \leq \lvert\eta\rvert < 1.$	$1. \leq \lvert\eta\rvert < 1.479$	$1.479 \leq \lvert\eta\rvert < 2.$	$2. \leq \lvert\eta\rvert \leq 2.5$
$10 < p_T \leq 15$	0.191 ± 0.006	0.226 ± 0.008	0.205 ± 0.009	0.227 ± 0.007
$15 < p_T \leq 20$	0.180 ± 0.009	0.257 ± 0.016	0.291 ± 0.019	0.236 ± 0.012
$20 < p_T \leq 25$	0.160 ± 0.015	0.294 ± 0.028	0.336 ± 0.031	0.243 ± 0.018
$25 < p_T \leq 35$	0.193 ± 0.022	0.256 ± 0.035	0.396 ± 0.034	0.291 ± 0.021

Dileptonic events can then be divided in three categories: events containing two leptons passing the tight selection (nn events), events containing one lepton passing the tight selection and a second lepton passing only the loose selection and not the tight selection ($n\bar{n}$ events) and events containing two leptons passing only the loose selection and not the tight selection ($\bar{n}\bar{n}$ events).

The multijet QCD events passing the selection are dominated by two fake leptons. The fake rate (R) being the probability of having a fake lepton passing the tight selection, the number of multijet QCD events having two leptons passing the selection can be

estimated from the number of $\bar{n}\bar{n}\bar{n}$ events ($N_{\bar{n}\bar{n}\bar{n}}$):

$$N_{nn}^{QCD} = \sum_{i,j} \frac{R_i R_j}{(1-R_i)(1-R_j)} N_{\bar{n}\bar{n}}^{ij}, \tag{3.9}$$

where i and j stand for the p_T-η binning and the flavour of each lepton.

In the same way, the W+Jets background (and also the semileptonic $t\bar{t}$) being mainly composed of one prompt lepton and one fake lepton, the estimation of the number of events passing the selection is obtained from the number of $n\bar{n}$ events ($N_{n\bar{n}}$):

$$N_{nn}^{Wj,raw} = \sum_{i,j} \frac{R_j}{(1-R_j)} N_{n\bar{n}}^{ij}. \tag{3.10}$$

However, the $N_{n\bar{n}}$ number can be contaminated by multijet QCD events where only one fake lepton passes the tight selection. In the opposite way, this number can also contain signal events where one prompt lepton does not pass the tight selection. In order to correct for these contribution, the final estimate for W+Jets is given by:

$$N_{nn}^{Wj} = N_{nn}^{Wj,raw} - 2N_{nn}^{QCD} - \Delta_{signal}. \tag{3.11}$$

The QCD contamination corresponds indeed to 2 N_{nn}^{QCD}, and the signal spillage is $\Delta_{signal} = N_{nn} \cdot SR_{\ell\ell'}$, with the spillage rate $SR_{\ell\ell'}$. The signal spillage for the ee and $\mu\mu$ cases is evaluated from a sample dominated by pure prompt and isolated leptons dileptonic events. This is achieved by selecting Z events passing the loose selection. For example, in the ee case:

$$SR_{ee} = \frac{1}{N_{nn}^{Zee}} \sum_{i,j} \frac{R_j}{1-R_j} N_{n\bar{n}}^{ij,Zee} \tag{3.12}$$

And the signal spillage for the $e\mu$ case is given by $SR_{e\mu} = 0.5(SR_{ee} + SR_{\mu\mu})$. The fact that the signal region contains two more jets with respect to the used Z sample is supposed to have a negligible effect in the ee case thanks to the choice of the loose selection for electron not relaxing the isolation. But it can have an effect on the $\mu\mu$ case, the isolation being affected by the presence of jets, which is then underestimating the spillage. For analysis with higher integrated luminosity, where a Z sample with two jets is statistically sufficient, this bias can be avoided by adding the jet selection.

The first source on systematic uncertainties comes from differences between the QCD sample used to determine the fake rate and the sample on which the fake rate is applied. The main differences are the jet momentum spectrum, which affects the lepton isolation, and the flavour content. Other subdominant biases are induced by the triggers used or by the electroweak signal contribution or statistical limitation during the

Table 3.10: Estimated contamination in the TrackJets scenario of multijet QCD (N_{nn}^{QCD}), uncorrected for spillage W+Jets ($N_{nn}^{Wj,raw}$) and for the positive defined total of fake leptons contamination after 3.1 pb^{-1} of integrated luminosity for each channel. This can be compared with the expected contamination from fake leptons in simulation, the total number of events expected in simulation and the total observed data events. Δ_{signal} is the signal spillage for W+Jets. The uncertainties are statistical, systematic + statistical or not given if negligible or straightforward. The uncertainty for 0 is computed assuming 1 event at $R = 0.3$.

Channel	ee	$\mu\mu$	$e\mu$
N_{nn}^{QCD}	$0.3 \pm 0.3 \pm 0.3$	$0.2 \pm 0.1 \pm 0.2$	$0.0\ ^{+0.2}_{-0.0}\ ^{+0.1}_{-0.0}$
$N_{nn}^{Wj,raw}$	$0.2 \pm 0.2 \pm 0.1$	$0.2 \pm 0.2 \pm 0.2$	$0.0\ ^{+0.4}_{-0.0}\ ^{+0.2}_{-0.0}$
Simulation with fakes	0.06	0.02	0.02
Total from simulation	5.0	4.2	5.0
Events in data	2	9	5
Δ_{signal}	0.10 ± 0.07	0.11 ± 0.04	0.06 ± 0.03
Fakes in data	$-0.19 \pm 0.42 \pm 0.35$	$-0.05 \pm 0.26 \pm 0.29$	$-0.06\ ^{+0.47}_{-0.0}\ ^{+0.20}_{-0.0}$
Fakes in data pos. def.	$0.0\ ^{+0.42}_{-0.0}\ ^{+0.35}_{-0.0}$	$0.0\ ^{+0.26}_{-0.0}\ ^{+0.29}_{-0.0}$	$0.0\ ^{+0.47}_{-0.0}\ ^{+0.20}_{-0.0}$

fake rate determination. The systematic uncertainty on the fake rate has been conservatively estimated to $\pm$50% for electrons and $^{+50}_{-100}$% for muons (because the loose selection relaxes the isolation condition, which increases the bias of the jet momentum spectrum). This leads to an uncertainty of 50% (resp. 100%) on the raw W+Jets estimation (resp. on the QCD estimation) for the electrons and $^{+50}_{-100}$% (resp. $^{+100}_{-100}$%) for the muons.

The results obtained with the fake lepton estimation method are given for each channel in Tab. 3.10 and Tab. 3.11 for the TrackJets scenario and the baseline scenario respectively. Due to the subtraction of the QCD contamination and the signal spillage, the final number can be negative. This is due to the statistical fluctuation of the independent samples used to evaluate the different contributions. The statistical uncertainties being taken into account, negative results are compatible with positive solutions, and I then use the positive defined solution with the same uncertainty in the cross-section determination. The results can also be detailed for each jet multiplicity bins for all channels combined, cf. Tab. 3.12 for the baseline scenario.

Table 3.11: Estimated contamination in the baseline scenario of multijet QCD (N_{nn}^{QCD}), uncorrected for spillage W+Jets ($N_{nn}^{Wj,raw}$) and for the positive defined total of fake leptons contamination after 3.1 pb^{-1} of integrated luminosity for each channel. This can be compared with the expected contamination from fake leptons in simulation, the total number of events expected in simulation and the total observed data events. Δ_{signal} is the signal spillage for W+Jets. The uncertainties are statistical, systematic + statistical or not given if negligible or straightforward. The uncertainty for 0 is computed assuming 1 event at $R = 0.3$.

Channel	ee	$\mu\mu$	$e\mu$
N_{nn}^{QCD}	$0.0\ ^{+0.1}_{-0.0}\ ^{+0.1}_{-0.0}$	$0.08 \pm 0.08 \pm 0.08$	$0.09 \pm 0.09 \pm 0.09$
$N_{nn}^{Wj,raw}$	$0.4 \pm 0.4 \pm 0.2$	$0.2 \pm 0.2 \pm 0.2$	$0.0\ ^{+0.4}_{-0.0}\ ^{+0.2}_{-0.0}$
Simulation with fakes	0.07	0.01	0.15
Total from simulation	1.8	2.1	4.9
Events in data	3	3	5
Δ_{signal}	0.13 ± 0.08	0.07 ± 0.04	0.17 ± 0.08
Fakes in data	$0.3 \pm 0.4 \pm 0.2$	$0.1 \pm 0.2 \pm 0.2$	$-0.3\ ^{+0.4}_{-0.1}\ ^{+0.2}_{-0.1}$
Fakes in data pos. def.	$0.3 \pm 0.4 \pm 0.2$	$0.1 \pm 0.2 \pm 0.2$	$0.0\ ^{+0.4}_{-0.0}\ ^{+0.2}_{-0.0}$

Table 3.12: Estimated contamination in the baseline scenario of multijet QCD (N_{nn}^{QCD}), uncorrected for spillage W+Jets ($N_{nn}^{Wj,raw}$) and for the positive defined total of fake leptons contamination after 3.1 pb^{-1} of integrated luminosity for JPTJet multiplicity bin. This can be compared with the expected contamination from fake leptons in simulation, the total number of events expected in simulation and the total observed data events. Δ_{signal} is the signal spillage for W+Jets. The uncertainties are statistical, systematic + statistical or not given if negligible or straightforward.

Mode	$N_{jets} = 0$	$N_{jets} = 1$	$N_{jets} \geq 2$
N_{nn}^{QCD}	$0.2 \pm 0.1 \pm 0.2$	$0.1 \pm 0.1 \pm 0.1$	$0.2 \pm 0.1 \pm 0.1$
$N_{nn}^{Wj,raw}$	$0.3 \pm 0.3 \pm 0.2$	$1.1 \pm 0.7 \pm 0.6$	$0.6 \pm 0.5 \pm 0.3$
Simulation with fakes	1.1	0.3	0.2
Total from simulation	3.8	4.8	8.9
Events in data	5	5	11
Δ_{signal}	0.2 ± 0.1	0.2 ± 0.1	0.4 ± 0.1
Fakes in data	$0.0\ ^{+0.3}_{-0.3}\ ^{+0.3}_{-0.3}$	$0.9 \pm 0.7 \pm 0.6$	$0.1 \pm 0.5 \pm 0.3$
Fakes in data pos.def.	$0.0\ ^{+0.3}_{-0.0}\ ^{+0.3}_{-0.0}$	$0.9 \pm 0.7 \pm 0.6$	$0.1\ ^{+0.5}_{-0.1}\ ^{+0.3}_{-0.1}$

3.3 Systematic uncertainties

I have considered several sources of systematic uncertainties in this study: the limitations or the approximations in theory used to model the simulated events, the experimental effects and biases which affect the event selections, and the determination of the integrated luminosity. The current section describes the two first point, the experimental uncertainties containing the effect of the pile-up and the selection efficiency. The theoretical and experimental systematic uncertainties on the backgrounds are treated together in the last part of this section. At the time of this study, the integrated luminosity uncertainty was assumed to be 11%, as stated in [90]. This uncertainty has been reexamined later to 4% without modification of the integrated luminosity value. However, the results presented in this thesis use the 11% uncertainty value to correspond to the published results. This uncertainty is quoted separately.

The summary of the systematic uncertainties that are developed in the following section is available in Section 3.4.2, in Tab. 3.14 and Tab. 3.15 for the TrackJets scenario and the baseline scenario respectively.

3.3.1 Theoretical uncertainties

In order to evaluate the systematic uncertainties on the selected events due to an effect of a theoretical tuning of the Monte Carlo generator, signal samples have been produced in different configurations.

The signal has been simulated with increased or reduced initial and final state radiations rate. Compared to the nominal setup, there is a 0.98 ± 0.01 factor (the uncertainty is statistical only) on the number of selected events for the larger radiation rate and 0.99 ± 0.01 for the smaller radiation rate. Within the statistical uncertainties, these results are consistent with an absence of any effect with 1% systematic uncertainty. I have not evaluated the Q^2 scale parameter uncertainty for this study. However, more recent results have shown that a reasonable variation of the Q^2 scale in simulation leads to uncertainties of the order of 2%. This uncertainty is not included in the final results.

In order to evaluate an effect in the simulation of the tau and the heavy flavour hadronic decays, the signal has been simulated with tools dedicated to model more accurately these processes and not used to generate the signal sample used in the analysis. These tools are EVTGEN [133] for the heavy flavour hadronic decays and TAUOLA [134] for the tau decays. Using or not these tools on the signal generation, the amount of selected events varies by 2%, which is the value used as systematic uncertainty.

The signal sample has been generated with MADGRAPH, which uses the leading order value of 1/9 as a branching ratio for the leptonic decay of the W boson. The current world average is 0.1080 ± 0.0009 [28]. Because the final result is the inclusive cross section extracted from a dileptonic selection, the simulated signal acceptance is tuned by a scale factor of $(0.1080 \cdot 9)^2 = 0.9448$, with a corresponding systematic uncertainty of 1.7%.

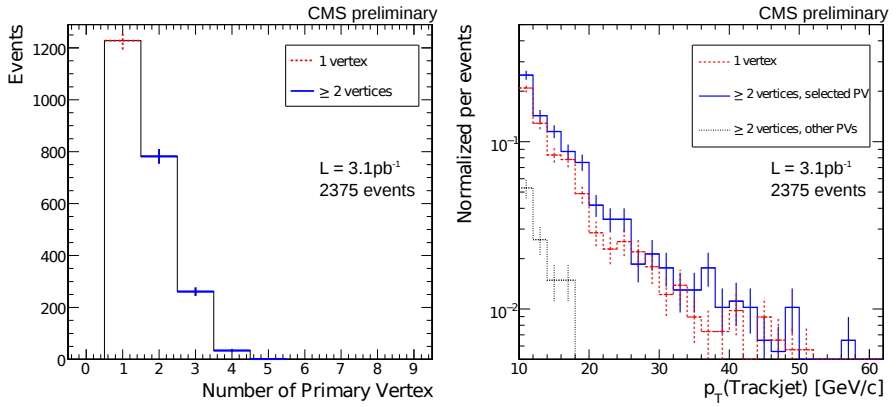
Some uncertainties also arise from parton distribution functions. Using the weight method [135] at generator level with the default CTEQ6.1 PDF in the MADGRAPH signal sample, the uncertainty on the acceptance can be obtained. The generator level selection is consistent with the analysis selection, asking two leptons with $|\eta| < 2.5$ and $p_{\mathrm{T}} > 20\,\mathrm{GeV}/c$ and two jets with $|\eta| < 2.5$, $p_{\mathrm{T}} > 30\,\mathrm{GeV}/c$ and not close to a selected lepton ($\Delta R > 0.4$). This leads to a relative uncertainty of 0.5%. Reweighting the central value of the acceptance to correspond to the MSTW 2008 NLO [14] at 68% C.L. and NNPDF2.0 PDFs [136] leads to a variation smaller than 0.1%, as well as a reweighting using the CTEQ6.6 PDFs [137] with maximal and minimal α_s. Because these variations are negligible, the PDF uncertainties are not taken into account in the total systematic uncertainty.

3.3.2 Pile-up uncertainties

During the harvesting of the $3.1\,\mathrm{pb}^{-1}$ of integrated luminosity, the proton beams conditions were such that during each bunch crossing, on average, two proton-proton interactions were expected. Even if an extra contribution due to the pile-up is softer than a dileptonic interaction (the probability of two hard interactions occurring simultaneously decreases exponentially), it can have an effect on the event selection, due to higher hadronic activity. Firstly, more jets and a higher missing energy are observed. This increases the probability of passing the jets and missing transverse energy selection for the background. For the signal, the effect is smaller, because rejected signal events for these cuts are already small. Secondly, the lepton isolation is deteriorated, reducing the probability to pass the selection. These two effects have thus opposite consequences on the selection, which contribute to the systematic uncertainty of the selection. The lepton isolation efficiency is however determined in data, and therefore takes into account pile-up effects.

In order to evaluate the total effect of pile-up on the selection, simulated signal samples with and without realistic pile-up conditions are compared. If the predicted effects on isolation and jets are indeed observed, the net effect appears to be negligible and no additional systematic uncertainties are assigned due to pile-up.

I have also tested pile-up effects on data, after the dileptonic selection without the Z-veto. If two proton-proton interactions happen in one event, there is a high probability that the corresponding two primary vertices will be reconstructed separately (cf. Section 2.2.2). Therefore, the subset of events with only one primary vertex is dominated by events where no additional proton-proton interaction has happened, while the complementary set of events with more than one primary vertex is dominated by events affected by pile-up. Comparing these two sets allows the determination of the effect of pile-up in data. The considered reconstructed vertices have to correspond to a vertex reconstructed from tracks, with a number of degrees of freedom higher than four for the vertex fit, and have to be located sufficiently close to the centre of the detector (the distance in the transverse plane (ρ) should be less than 2 cm and the longitudinal distance (z) should be less than 24 cm). The Fig. 3.10 (a) shows the vertex multiplicity in data, after the dileptonic selection without the Z-veto.



(a) (b)

Figure 3.10: Effect of the pile-up in data, after the dileptonic selection without the Z-veto for 3.1 pb^{-1} of integrated luminosity for all channels combined. The uncertainties are statistical. (a) Vertex multiplicity, for a vertex selection as described in the text. (b) Distribution of the transverse momentum of TrackJets for events with one primary vertex, and for events with more than one primary vertex separated for the TrackJets associated to the dileptonic interaction vertex and the others.

An effect on isolation is observed in the calorimeter isolation, but for a 0.3 cone isolation and a relative combined isolation cut at 0.15, this effect has no impact on the selection.

The effect of the pile-up on jets multiplicity can also be observed by comparing the one-vertex events and the more-than-one-vertex events. This effect is visible for low

p_{T} jets only, and appears to be negligible for a cut at 30 GeV/c on the jet transverse momentum as used in this study. The additional jets from the extra interactions have indeed low transverse momentum, as visible using TrackJets where jets from the dileptonic interaction primary vertex are separated from jets from another vertex, Fig. 3.10 (b).

Additional pile-up interactions can also affect the momentum of the jets belonging to the dileptonic interaction, if some particles from the additional interaction create deposits clustered inside the dileptonic interaction jet. The effect is already reduced by several techniques: the offset correction already rejects typical low energy pile-up deposits, and pile-up correction factor depending on the number of vertices can also be applied. Moreover, for track corrected jets, this effect can be avoided by asking that the considered tracks belong to the interesting primary vertex. This is however not the default configuration of the PFJets and JPTJets reconstructions and this feature has not been used in this study. In order to check this effect in data, TrackJets, which are not affected thanks to the default vertex association during the reconstruction (cf. Appendix A.4), can be used. I have performed a geometrical matching between the TrackJets and a CaloJet, and I have compared the p_{T} ratio between these two jets for one vertex events and for more than one vertex events. For the region of interest of this study, no statistically relevant shift in the CaloJet p_{T} has been observed.

3.3.3 Uncertainties from leptonic and jet selections

The leptonic selection efficiencies can be factorized in several parts: the trigger selection for dileptonic events, the reconstruction efficiency, the identification and isolation efficiency and the charge identification. These efficiency uncertainties have been discussed in Section 1.2.4 for trigger efficiency, Section 2.3.4 and 2.4.4 for the reconstruction efficiency, the identification and selection efficiency and the charge identification. These efficiencies are in general valid for Drell-Yan events, and have to be corrected for dileptonic $t\bar{t}$ events, as stated in Section 2.3.4.

Comparing the data and the simulation for these four efficiencies, it appears that the results are consistent and that no scale factor correction is needed for the lepton efficiency. The systematic uncertainties are evaluated as 100% of the differences between results from data and simulation. And the systematic uncertainty for the Drell-Yan to dileptonic $t\bar{t}$ correction has been taken as 50%. This is summarized in Tab. 3.13 for each channel and for the three channels combined.

The uncertainty due to the jet energy scale has also to be evaluated. In the TrackJets scenario, I have considered a p_T resolution of 5%, which is conservative with respect to the results for TrackJets (cf. Appendix A). Therefore, I have performed

Table 3.13: Summary of relative systematic uncertainties for dileptonic $t\bar{t}$ selection for each lepton selection contribution, for the each channel and the three channels combined.

Channel	ee	$\mu\mu$	$e\mu$	All
Trigger	0.1%	1%	0.3%	0.3%
Reconstruction	3%	3%	2.1%	2.1%
Identification and isolation	5%	1%	2.5%	2.5%
Opposite sign	0.4%	$<$0.1%	0.2%	0.2%
Simulation $t\bar{t}$ vs Z	4%	4%	2.8%	2.8%
Combined	7.1%	5.2%	4.4%	4.4%

the selection with a $\pm$ 5% variation on the momentum scale of the jets. The systematic uncertainty is evaluated as 100% of the variation of the selected events, which is 3.1%, 2.9% and 3.2% for the ee, $\mu\mu$ and $e\mu$ channels respectively. For the baseline analysis, the jet energy scale for jets in the interest region is evaluated at 5%, which is the recommended value for the jets corrected with respect to tracker information. The selection is performed with a simultaneous $\pm$ 5% variation on the hadronic energy of the jets and of the missing transverse energy. The systematic uncertainty is evaluated as 100% of the variation of the selected events, which is 3.8%, 4.0% and 3.4% for the ee, $\mu\mu$ and $e\mu$ channels respectively.

3.3.4 Background uncertainties

The backgrounds are affected by the same systematic uncertainties than the signal, but in addition, the uncertainty on their total cross sections, impacting directly their rate, has to be considered. The backgrounds uncertainty for the background predicted by data-driven methods is given by the results of these methods, as discussed in Section 3.2. For the Drell-Yan into two electrons or two muons, it is described in Section 3.2.1 and the systematic uncertainty is evaluated at 25% for the TrackJets scenario and 50% for the baseline scenario. For the multijet QCD, the W+Jets and the semileptonic $t\bar{t}$, it is described in Section 3.2.2 and the systematic uncertainties are visible in Tab. 3.10 and Tab. 3.11. The remaining backgrounds are small and totally rely on the simulation prediction. I have assigned a 50% fractional systematic uncertainty to these background, which is conservative (considering 15-20% for theoretical cross-section, 11% for luminosity, 5% for leptonic selection and 10% for energy scale). The background uncertainty is higher for the TrackJets scenario, where the signal-over-background ratio is higher.

3.4 First cross-section determination

This section describes the method used to extract the cross section, summarizes the expectations from simulations or data-driven methods and presents the final results.

3.4.1 Simple counting method

The cross-section $\sigma_{data}^{t\bar{t}}$ determination is based on a simple counting method for the sum of the events in the three channels:

$$\sigma_{data}^{t\bar{t}} = \sigma_{SM}^{t\bar{t}} \frac{S_{obs}}{SF \cdot S_{exp}} = \frac{S_{obs}}{SF \cdot \mathcal{L} \cdot A}, \tag{3.13}$$

where $\sigma_{SM}^{t\bar{t}}$ is the theoretical cross section, S_{obs} the number of observed signal events ($S_{obs} = N - B$ where N is the number of observed events and B the expected number of background events), S_{exp} the expected number of signal events directly from simulation and SF a scale factor taking into account the differences between simulation and real data. This equation can also be expressed in term of integrated luminosity $\mathcal{L}$ and the acceptance of the selection A obtained directly from simulation. The Standard Model NLO prediction for the theoretical cross section is $\sigma_{SM}^{t\bar{t}} = 157.5\,\text{pb}$ [24].

More complex methods to extract the cross section or to combine the different channel are possible, but at this early stage dominated by statistical uncertainty, no real gain is reached.

The scale factor SF is the product of all the factors to consider to adjust the simulation to reality and to take into account the systematic uncertainties. These corrections are:

- The correction due to the theoretical modelling of the fraction of events passing the selection;
- The differences in event selections, for leptons, for jets and missing transverse energy;
- The scaling factor between the observed integrated luminosity and the real integrated luminosity.

As shown in Section 3.3, the majority of these factors are centred on the unity. Only the branching ratio induces a factor 0.945, cf. Section 3.3.1, which leads to $SF = 0.945$.

The equations for statistical and systematic uncertainties can be derived from Eq. 3.13. The absolute statistical uncertainty on the observed cross section ($\delta_{stat}(\sigma^{t\bar{t}}_{data})$) is given by:

$$\delta_{stat}(\sigma^{t\bar{t}}_{data}) = \sigma^{t\bar{t}}_{SM}\frac{\sqrt{N}}{S_{exp}} = \sigma^{t\bar{t}}_{data}\frac{\sqrt{N}}{S_{obs}}, \tag{3.14}$$

where N is the total number of observed events. The absolute systematic uncertainty on the observed cross section ($\delta_{syst+lumi}(\sigma^{t\bar{t}}_{data})$) is given by:

$$\delta_{syst+lumi}(\sigma^{t\bar{t}}_{data}) = \sigma^{t\bar{t}}_{data}\sqrt{\left(\frac{\delta B}{S_{obs}}\right)^2 + (\delta SF)^2}, \tag{3.15}$$

where δB is the systematic uncertainty on the expected number of events for backgrounds, from simulations or data-driven methods, and δSF is the systematic uncertainty on the scale factor. Assuming all scale factor uncertainties to be uncorrelated and with Gaussian behaviour, $\delta SF = \sqrt{\sum_i (\delta SF_i)^2}$. It is informative to separate the integrated luminosity uncertainty from other systematic uncertainties, and for the following, $\delta_{syst+lumi}(\sigma^{t\bar{t}}_{data})$ is divided and renamed in $\delta_{syst}(\sigma^{t\bar{t}}_{data})$ and $\delta_{lumi}(\sigma^{t\bar{t}}_{data})$, where δ_{syst} does not contain the luminosity part.

3.4.2 Results with the first data

Using the data-driven methods described in Section 3.2, the summary of expected signal and backgrounds compared to observed data is presented in Tab. 3.16 and Tab. 3.17 for the TrackJet scenario and the baseline scenario respectively. For illustration, the JPTJets multiplicity after the full selection with data-driven expected background and combined background uncertainties is shown in Fig. 3.11.

Applying the number of events (summarized in Tab. 3.16 and Tab. 3.17) and the uncertainties values (summarized in Tab. 3.14 and Tab. 3.15) to the formulae described in Section 3.4.1, I have derived the cross section.

For the TrackJets scenario:

$$\sigma(\mathrm{pp} \to \mathrm{t\bar{t}}) = 221 \pm 81(\mathrm{stat.}) \pm 34(\mathrm{syst.}) \pm 24(\mathrm{lumi.})\mathrm{pb}$$

For the baseline scenario:

$$\sigma(\mathrm{pp} \to \mathrm{t\bar{t}}) = 194 \pm 72(\mathrm{stat.}) \pm 24(\mathrm{syst.}) \pm 21(\mathrm{lumi.})\mathrm{pb}$$

Table 3.14: Summary of relative systematic uncertainties on the observed cross section for each channel and the three channels combined, for the TrackJets scenario. The details of the extraction of these numbers are given in Section 3.3.

Channel	ee	$\mu\mu$	$e\mu$	All
Lepton selection	7.1%	5.2%	4.4%	4.4%
TrackJets momentum scale	3.1%	2.9%	3.2%	3.1%
ISR/FSR	1%	1%	1%	1%
Decay model	2%	2%	2%	2%
Branching ratio	1.7%	1.7%	1.7%	1.7%
Subtotal (no backgrounds, no luminosity)	8.2%	6.6%	6.1%	6.1%
Backgrounds ($\delta B/S_{exp}$)	51%	45%	11%	19%
Total without luminosity	51%	45%	13%	20%
Integrated luminosity	11%	11%	11%	11%

Table 3.15: Summary of relative systematic uncertainties on the observed cross section for each channel and the three channels combined, for the baseline scenario. The details of the extraction of these numbers are given in Section 3.3.

Channel	ee	$\mu\mu$	$e\mu$	All
Lepton selection	7.1%	5.2%	4.4%	4.4%
Energy scale	3.8%	4.0%	3.4%	3.7%
ISR/FSR	1%	1%	1%	1%
Decay model	2%	2%	2%	2%
Branching ratio	1.7%	1.7%	1.7%	1.7%
Subtotal (no backgrounds, no luminosity)	8.6%	7.1%	6.2%	6.4%
Backgrounds ($\delta B/S_{exp}$)	50%	40%	$^{+10}_{-5}$%	15%
Total without luminosity	50%	40%	$^{+12}_{-8}$%	16%
Integrated luminosity	11%	11%	11%	11%

This is consistent with less than one standard deviation with the NLO theoretical cross section of $157.5^{+23.2}_{-24.4}$ pb for a top quark mass of $\mathrm{m_t} = 172.5\,\mathrm{GeV}/c^2$. The dominant uncertainty is statistical. The two methods provide similar precision, with a total uncertainty of $\sim 40\%$, slightly better for the baseline scenario. This is due to the worse signal-over-background ratio of the TrackJets scenario which increases the systematic

Table 3.16: Expected signal and background yields after the TrackJets scenario selection, from simulation (MC) and data-driven estimation (data), after 3.1 pb^{-1} of integrated luminosity, for the ee channel, the $\mu\mu$ channel, the $e\mu$ channel and for all channels combined. The uncertainties are statistical and systematic. The last line gives the number of observed events in data.

Channel	ee	$\mu\mu$	$e\mu$	all
Drell-Yan (data)	2.0 ± 0.7	2.4 ± 0.7	N/A	4.4 ± 1.3
QCD (data)	0.3 ± 0.4	0.3 ± 0.2	0.0 ± 0.2	0.5 ± 0.5
W+Jets (data)	-0.5 ± 0.7	-0.2 ± 0.4	-0.06 ± 0.6	-0.8 ± 1.0
Others (MC)	0.16 ± 0.08	0.19 ± 0.09	0.29 ± 0.14	0.6 ± 0.3
Total background	2.2 ± 0.9	2.6 ± 0.8	0.3 ± 0.5	5.1 ± 1.6
Dileptonic $t\bar{t}$ (MC)	1.7 ± 0.1	1.9 ± 0.1	4.6 ± 0.3	8.2 ± 0.5
Total expected	3.9 ± 0.9	4.4 ± 0.9	4.9 ± 0.6	$13.3, \pm 1.6$
Observed data	2	9	5	16

Table 3.17: Expected signal and background yields after the baseline scenario selection (using JPTJets), from simulation (MC) and data-driven estimation (data), after 3.1 pb^{-1} of integrated luminosity, for the ee channel, the $\mu\mu$ channel, the $e\mu$ channel and for all channels combined. The uncertainties are statistical and systematic. The last line gives the number of observed events in data.

Channel	ee	$\mu\mu$	$e\mu$	all
Drell-Yan (data)	$0.8 \pm 0.4 \pm 0.4$	$0.6 \pm 0.4 \pm 0.3$	N/A	$1.4 \pm 0.5 \pm 0.5$
Evts with fakes (data)	$0.3 \pm 0.4 \pm 0.2$	$0.1 \pm 0.2 \pm 0.2$	$-0.3\,^{+0.4}_{-0.1}\,^{+0.2}_{-0.1}$	$0.1 \pm 0.5 \pm 0.3$
VV (MC)	0.03 ± 0.02	0.03 ± 0.02	0.08 ± 0.04	0.13 ± 0.07
Single top (MC)	0.05 ± 0.03	0.05 ± 0.03	0.15 ± 0.08	0.25 ± 0.13
Drell-Yan $\tau\tau$ (MC)	0.04 ± 0.02	0.07 ± 0.03	0.07 ± 0.04	0.18 ± 0.09
Total backgrounds	1.2 ± 0.7	0.8 ± 0.6	$0.3\,^{+0.4}_{-0.1}$	2.1 ± 1.0
Dileptonic $t\bar{t}$ (MC)	1.5 ± 0.1	1.7 ± 0.1	4.5 ± 0.3	7.7 ± 0.5
Total expected	2.7 ± 0.7	2.6 ± 0.6	4.8 ± 0.5	9.8 ± 1.1
Observed data	3	3	5	11

uncertainty due to the background and compensates the slightly better jet energy scale uncertainty and statistical uncertainty. For comparison, the last results from the CDF Run I [138] after 109 ± 7 pb^{-1} of integrated luminosity were accompanied by 48%

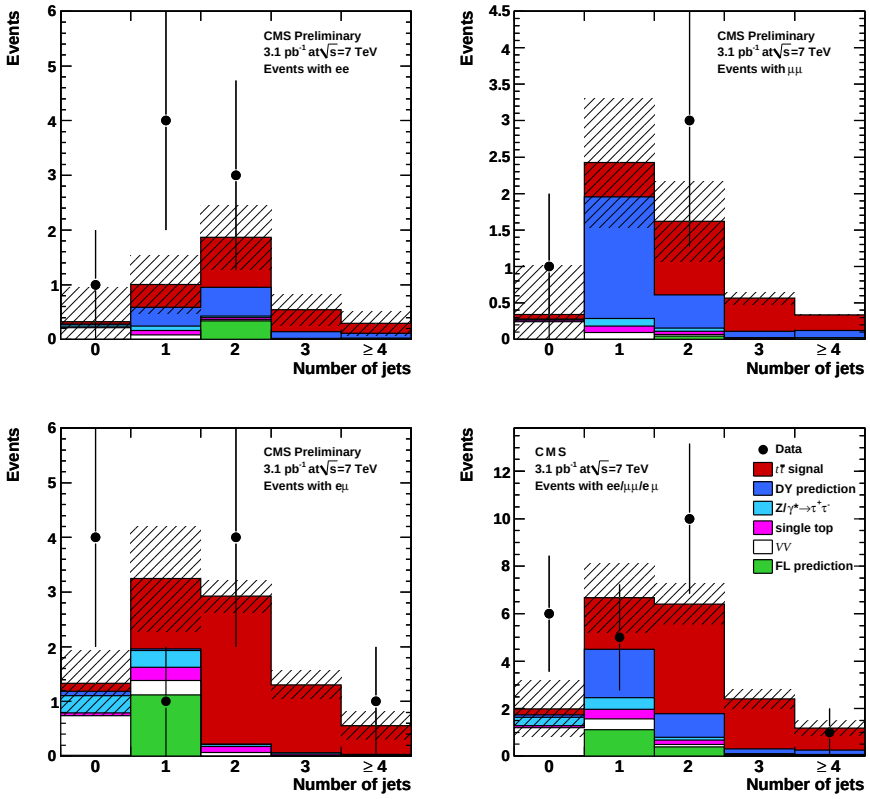


Figure 3.11: Distributions of corrected JPTJet multiplicity for jets with $p_T > 30\,\text{GeV}/c$ after the baseline scenario selection without the two jets cut, for the ee channel (a), the $\mu\mu$ channel (b), the $e\mu$ channel (c) and for all channels combined (d), after $3.1\,\text{pb}^{-1}$ of integrated luminosity. The plain histograms are from simulation or data-driven estimates. The shaded area describes the total uncertainty on the background contribution. The black dots correspond to data, with statistical uncertainty. From [1].

of uncertainty and the first CDF Run II results [139] after $197 \pm 12\,\text{pb}^{-1}$ by 36%. This result, at $3.1\,\text{pb}^{-1}$ of integrated luminosity, shows a good control of the detector and the tools used to extract the cross section, and paves the way for a more precise measurement at higher luminosity.

3.5 Latest results on cross-section determination

Since this first result, the CMS collaboration has continued its efforts. Following this early measurement, $t\bar{t}$ cross-section results have been published for the data harvested during the year 2010 ($35.9 \pm 1.4\,\text{pb}^{-1}$ of integrated luminosity) [17], using the dileptonic and the semileptonic channels.

The dileptonic results [128] are mainly based on an update from the baseline scenario technique described in this thesis. In addition to the baseline selection, two other selections have been defined, providing three selections in the three channels (ee, $e\mu$, $\mu\mu$). Using the simple counting method describes in Section 3.4.1, this lead to nine measurements, which are combined together used the best linear unbiased estimator technique (BLUE) [140].

The first selection is very similar to the one described in the baseline scenario. The triggers are of course adapted to the ones used for these data, but still provide really high efficiency for the dileptonic selection. The consistency with respect to triggers has been improved, requiring at least one lepton consistent with the trigger requirement. The lepton selection is slightly improved ($E_T^{\text{calo}} > 15\,\text{GeV}$ and no ECal cleaning for electrons, $|\eta| < 2.4$ for muons) and the threshold in the dileptonic invariant mass is slightly increased up to $12\,\text{GeV}/c^2$ (to reject Upsilon meson tail). The selected jets are reconstructed with the Particle Flow algorithm and benefits from residual corrections derived from data. The jet identification criteria have also been updated. No missing transverse energy cut is applied in the $e\mu$ channel.

The second selection is defined on top of the first selection and uses b-tagging algorithm, which has been found to be reliable in the measurements after $3.1\,\text{pb}^{-1}$ of integrated luminosity. The high efficiency track counting algorithm (cf. Section 2.5.4) has been used to select events with at least one b-tagged jet.

The third selection of events is similar to the first one, but selects events with exactly one jet. This leads to more background-contaminated samples, but can provide some gain in the combination. Due to the large background contamination, this selection also requires a missing transverse energy of at least 50 GeV for the ee and $\mu\mu$ channels and a sum of the transverse mass of the two leptons higher than 130 GeV for the $e\mu$ channel, the transverse mass being defined relative to the value and the direction of the missing energy transverse vector: $M_T^{\ell} = \sqrt{2p_T^{\ell}\not{E}_T[1 - cos(\phi_{\not{E}_T} - \phi_{\ell})]}$.

The data-driven methods used for estimate the backgrounds are identical to the ones presented in this thesis (Section 3.2). The main sources of systematic uncertainties

on the signal are the jet energy scale and the lepton selection model that results from the difference in the lepton isolation between a Drell-Yan and a $t\bar{t}$ environment. The selection with exactly one jet is also more affected by the Q^2 scale variation in the simulation of the signal process and the selection using the b-tagging contains the systematic uncertainty due to b-tagging. The uncertainty on the normalization of the integrated luminosity for 2010 data is 4% [141]. The final $t\bar{t}$ cross section for the dileptonic case is obtained from the nine cross sections combination: $\sigma_{t\bar{t}}(\text{dileptonic}) = 168 \pm 18(\text{stat.}) \pm 14(\text{syst.}) \pm 7(\text{lumi.})\text{pb}$.

The semileptonic study has been divided in two analyses, one [142] using a b-tagging algorithm to identify the b-jets and one [143] using the kinematical properties of the semi-leptonic $t\bar{t}$, using no b-jet identification. The semi-leptonic selection consists into selecting one energetic isolated lepton in the fiducial region of the detector ($p_T >$ 30 GeV/c and $|\eta| < 2.5$ for the electron, $p_T > 20$ GeV/c and $|\eta| < 2.1$ for the muon). The isolation is determined as in the dileptonic case. A selection is also applied on the jets multiplicity, based on the jets with corrected transverse momentum higher than 30 (25) GeV/c in $|\eta| < 2.4$ region for the no b-tags analysis (for the b-tag analysis). The b-tag analysis also requires a missing transverse energy of at least 20 GeV and the presence of at least one b-tagged jet using the simple secondary vertex reconstruction as b-tag discriminator (cf. Section 2.5.4).

In the b-tag analysis, the cross section is extracted using a simultaneous fit of the secondary vertex mass (from tracks associated with the vertex with a pion mass assumption) and the jets and b-tagged jets multiplicity. For each channel (lept) and each number of jets (jets) and b-tagged jets (tag), the predicted number of events for a process i is given by:

$$\mathcal{N}_i^{\text{pred}}(\text{lept}, \text{jets}, \text{tag}) = k_i \cdot N_i^{\text{MC}}(\text{lept}, \text{jets}, \text{tag}) \cdot \prod_X P_i^{\text{X}}(\text{lept}, \text{jets}, \text{tag}|\mathcal{R}_{\text{X}}) \quad (3.16)$$

where k is the scale factor parameter optimized by the fit, N^{MC} is the number of events predicted by simulation and corrected for discrepancies between data and simulated events, X is a systematic effect that can be the b-tag efficiency, the mistag rate, the jet energy scale or the Q^2 scale, $\mathcal{R}_{\text{X}}$ is the nuisance parameter corresponding to the systematic X and P^{X} is a polynomial function, obtained from simulation, describing the effect of the nuisance parameter.

The result of the fit, performed under constraints to guarantee a physically consistent result, returns the value of k and $\mathcal{R}_{\text{X}}$. The b-tag scale factor is evaluated to 97.5^{+5}_{-4}%, which is consistent with other b-tag studies. The jet energy scale factor is slightly harder than expected. The scale factors with respect to the NLO prediction for Wb and Wc processes are respectively $1.9^{+0.6}_{-0.5}$ and 1.4 ± 0.2, which is consistent with

recent observations at the TEVATRON [144][145][146]. The systematic uncertainties have been taken into account directly in the fit procedure or additionally on the cross-section results. The dominant systematic uncertainties are the jet energy scale, the b-tag efficiency and the Q^2 scale variation on the W+Jets background. Finally, the $t\bar{t}$ cross section is extracted: $\sigma_{t\bar{t}}$(semi-leptonic with b-tag) $= 150 \pm 9$(stat.)± 17(syst.)$\pm$ 6(lumi.)pb.

In the no b-tags analysis, the cross section is extracted from a fit on the missing transverse energy for events with exactly three jets and on the M_3 variable (invariant mass of the three jets of highest vectorial sum of the transverse momentum of their components) for events with at least four jets.

The fit is performed on template distributions, obtained from simulation or from data-driven method (for multijet QCD). Some constraints are applied to force the result of the fit in a region that is consistent with physics. The effect of a given systematic uncertainty on the result of the fit is evaluated by simulating templates with $\pm 1\sigma$ variation on the systematic uncertainty. The dominant systematic uncertainties are the jet energy scale and the Q^2 scale variation on the $W+$Jets background. The obtained $t\bar{t}$ cross section is: $\sigma_{t\bar{t}}$(semi-leptonic without b-tag) $= 173^{+39}_{-32}$(stat.+syst.)± 7(lumi.)pb.

The final $t\bar{t}$ cross section provided by the CMS collaboration is obtained by combining the results from the semi-leptonic analysis using the b-tagging algorithm and from the dileptonic analysis. This combination, using the BLUE technique, divides the uncertainties between the uncorrelated and the correlated ones. The uncorrelated uncertainties are the statistical uncertainties in the two analyses and the uncertainties from the background modelling in the dileptonic analysis. All the other uncertainties are considered correlated. The final $t\bar{t}$ cross section from the CMS Collaboration is:

$$\sigma_{t\bar{t}}(\text{CMS } 36\,\text{pb}^{-1}) = 158 \pm 10(\text{unc.}) \pm 15(\text{cor.}) \pm 6(\text{lumi.})\text{pb}$$

This result can be compared with theoretical cross sections, cf. Tab. 1.1, and shows a good agreement between the theory and the observed result.

More recently, a dileptonic analysis for $1.14\,\text{fb}^{-1}$ of integrated luminosity has been made public [147]. It consists mainly of an update of what has been realized in the $36\,\text{pb}^{-1}$ dileptonic analysis, with a selection very similar to the second selection described previously (two opposite sign isolated leptons, two jets, missing transverse energy in the ee and $\mu\mu$ channels and at least one b-tagged jet). The final number, obtained by combination of the cross sections obtained in the three channels, is:

$$\sigma_{t\bar{t}}(\text{CMS } 1.14\,\text{fb}^{-1}) = 169.9 \pm 3.9(\text{stat.}) \pm 16.3(\text{syst.}) \pm 7.6(\text{lumi.})\text{pb}$$

Conclusion

The measure of the inclusive top quark pair cross section plays an important role at several points of view. First, the precision on the normalization of the $t\bar{t}$ cross section can improve the study of many interesting signals where $t\bar{t}$ is a background. The presence of excess or deficit can point out new physics impacting the production or the decay of the top quark or mimicking its signature. Finally, the difference in $t\bar{t}$ production implies a complementary of the LHC with respect to the TEVATRON results, where some theoretical hypotheses can be constrained after a precision on the $t\bar{t}$ cross section of 20% [57].

The results of this thesis have been obtained from data harvested between March and September 2010 in proton-proton collisions at a centre-of-mass energy of 7 TeV, corresponding to an accumulated luminosity of $3.1\,\mathrm{pb}^{-1}$. Given the complexity of the topology and possible uncertainties related to the quality of early data, special care on the reconstruction techniques have been applied that guaranty the robustness of the analysis. Thus, two scenarios have been developed: The first, called the TrackJet scenario, is based on a minimal use of the CMS calorimeters and relies mainly on the CMS tracker subdetector information; The second, called the baseline scenario, makes use of the whole detector. In both cases, the data selection has been optimized to guaranty the robustness of the analysis. The events selection is based on the presence of two energetic isolated opposite sign leptons (electrons or muons) and the presence of at least two jets. The presence of transverse missing energy induced by the neutrinos is also requested in the baseline scenario.

It has been shown that the two scenarios are complementary and allows a direct cross-check, increasing the confidence on the results. In particular independent methods to estimate the amount of dominant backgrounds from data (the Drell-Yan process and the contribution from non-prompt or fake leptons) have been developed. The two cross section measurement uncertainties are dominated by the statistical uncertainty

around 37%, with systematic uncertainties of 15% and 12% for the TrackJets scenario and the baseline scenario, respectively.

The result for the baseline scenario is:

$$\sigma(\mathrm{pp} \to \mathrm{t\bar{t}}) = 194 \pm 72(\mathrm{stat.}) \pm 24(\mathrm{syst.}) \pm 21(\mathrm{lumi.})\mathrm{pb}$$

This analysis has proved that robust and early scenarios were able to provide remarkable results, mainly limited by statistical uncertainty. It has also demonstrated the potential of the LHC, considered as a top quark factory, and the outstanding performances of the CMS detector. This study has opened the way for top physics at CMS, currently in full activity. The baseline scenario developed and tested on real data in this study is indeed used as a basis for the dileptonic studies with high integrated luminosity. The TrackJets scenario has appeared to provide competitive results avoiding pile-up contamination and systematic uncertainties from the calorimeter detectors. This could be exploited in the context of a more and more challenging pile-up contamination due to the increase of the instantaneous luminosity at LHC.

With a better understanding of the reconstruction techniques and of the detector, several points of this early analysis should be developed, leading to a better resolution on the measured cross section. The statistical uncertainty will decrease with accumulated data, allowing more stringent selection enhancing the purity of the signal and decreasing the systematic uncertainty from the backgrounds. After ~ 1 fb^{-1}, the statistical uncertainty is already around 2% for the dileptonic channel. Improvement of the luminosity measurement techniques should also allow a reduction of the luminosity uncertainty. In the presented study, the other systematic uncertainties are dominated by the uncertainties on the background normalization, the lepton reconstruction and the jet energy scale. The background uncertainties can be reduced in several ways: they are conservative in this study and are based on limited data samples, and a more stringent selection will reduce the effect of the background. The lepton reconstruction and jet energy scale uncertainties should also decrease with a better understanding of the detector. Other uncertainties can appear in more complex studies at higher instantaneous luminosity, e.g., related to the b-tagging algorithm or the presence of pile-up. More advance combination methods will also improve the results, as well as the consideration of the other complementary channels. The possibility to measure the ratio between the $t\bar{t}$ cross section and the Drell-Yan cross section can also provide results where some systematic uncertainties cancel. Eventually, the LHC precision on the inclusive top quark pair cross-section measurement is expected to reach 5%, while the current approximated NNLO predictions have an uncertainty of $\sim 6\%$.

Appendix A

TrackJets Commissioning

A reconstruction of jets from tracker detector only has several advantages with respect to the usual jet reconstructions dominantly based on calorimeters deposits. It provides a simple reconstruction and an independent measurement with respect to the calorimeter based reconstructions and benefits from the good energy and direction resolutions of the tracker, especially for low energy. Thanks to the association of each tracks to a given vertex, it is also totally transparent to pile-up superposition effect. The counterpart is the absence of the contribution of neutral particles.

In this context, the TrackJets have been used in underlying event analysis [148], as validation for the calorimeter based jets [123] and in order to recover the benefits of a jet selection in conservative scenario which does not use the calorimeters, as it is the case in this thesis.

The jets are essential in the dileptonic $t\bar{t}$ selection, because it is a powerful discriminant with respect to the large Drell-Yan process. The TrackJets play therefore a central role in the scenario which makes a minimal use of the calorimeters. To validate this scenario, I have participated to the commissioning of the TrackJets which has involved the following study.

This appendix contains summary of the commissioning of TrackJets [119] and specific comparison with CaloJets in the dileptonic $t\bar{t}$ framework [10]. The goal of this appendix is to provide a complete definition of the TrackJets and to convince the reader that the different assertions made in the main chapters have been tested and are correct.

The first section, Section A.1, describes the reconstruction and the performance of the reconstruction of TrackJets. The second section, Section A.2, presents comparisons

between simulation and data for TrackJets, and a good agreement has been found in dijet and inclusive minimum bias events. The third section, Section A.4, illustrates the TrackJet robustness with respect to pile-up events. Finally, the last section, Section A.5, compares the usual CaloJets and the TrackJets in the dileptonic $t\bar{t}$ context, showing that the use of TrackJets is a satisfactory alternative to CaloJets in dileptonic $t\bar{t}$ environment.

A.1 TrackJet Reconstruction at CMS

TrackJets are reconstructed from tracks of charged particles. Since this study aims at validating TrackJets on a broad scale, the tracks used as input for the TrackJet reconstruction are selected with a minimal set of quality requirements, ensuring a low fake rate and at the same time sufficient efficiency to reconstruct TrackJets down to transverse momenta of a few GeV/c.

The TrackJets are reconstructed from tracks built by the standard CMS reconstruction, with the following requirement:

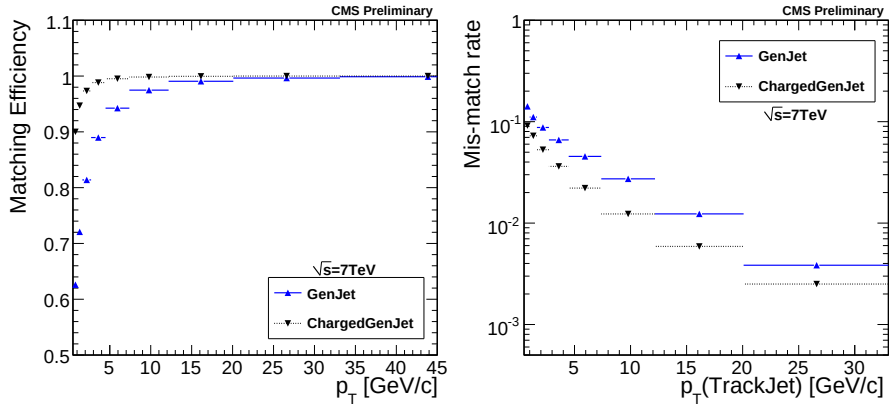
- in order to reject tracks not originating from hard interaction, a loose cut on the distance between the track and a primary vertex is applied ($|d_{xy}| < 0.2$ cm and $|d_z| < 1$ cm);
- a minimal transverse momentum is required ($p_T > 300$ MeV), which is kept very low to ensure an optimal reconstruction of low p_{T} TrackJets;
- in order to reject badly reconstructed tracks, a combined cut is applied on the momentum error and the track fit χ^2 ($\frac{\sigma(p_T)}{p_T}\max(1, \chi^2/n_{dof}) < 0.2$);
- the selected track should satisfy the "highPurity" flag which is defined during the reconstruction (cf. Section 2.2.2).

This strict track selection implemented for TrackJets ensures a straightforward association to the primary vertex, retaining very good angular resolution and very high jet finding efficiencies.

TrackJets are built by clustering the selected tracks using the anti-k$_{\mathrm{T}}$ algorithm [116] with jet cone size of $R = 0.5$.

All TrackJets are clustered separately for each primary vertex in the event, and the association of tracks to the primary vertex (PV) requires $|z(\mathrm{track})_{\mathrm{PV}} - z(\mathrm{PV})| < 0.5$ cm, where $z(\mathrm{track})_{\mathrm{PV}}$ is the z coordinate of the track helix at the point of closest approach to the primary vertex in the transverse (xy) plane.

One of the most important characteristics of jet reconstruction is the correspondence between the reconstructed jets and the "true" jets. I have studied this using simulated minimum bias events, where the "true" jets are built by applying the same clusterization algorithm to the simulated stable particles after the PYTHIA hadronization step ("GenJets" in the following) or to the charged ones only ("ChargedGenJets"). It is worth noting that the transverse momentum scale of the ChargedGenJets corresponds only to the charged particles, there is therefore a factor ~ 0.6 with respect to the transverse momentum of the GenJets. The same factor applies for TrackJets.



(a) (b)

Figure A.1: (a) The efficiency of matching GenJets and ChargedGenJets with to reconstructed TrackJets as a function of the GenJet and ChargedGenJet p_{T}; (b) the mismatch rate of reconstructed TrackJets to GenJets or ChargedGenJets, as a function of the TrackJet p_{T}.

The efficiency of matching GenJets and ChargedGenJets with $|\eta| < 2$ to reconstructed TrackJets within $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} < 0.5$ as a function of the GenJet and ChargedGenJets p_{T} is displayed in Fig. A.1 (a), and shows that TrackJets have efficiencies greater than 99% to correspond to true GenJets with $p_{\mathrm{T}} > 15\,\mathrm{GeV}/c$.

Fig. A.1 (b) illustrates the fraction of reconstructed TrackJets that are not matched (within $\Delta R < 0.3$) to GenJets or ChargedGenJets, as a function of the TrackJet p_{T}, and reveals that such mis-matched fraction of TrackJets that do not correspond to a true GenJet is lower than 1% for TrackJets with $p_{\mathrm{T}} > 10\,\mathrm{GeV}/c$.

A.2 Validation of TrackJets

TrackJets in data have been compared with simulation, for dijet and inclusive cases. The data sample is a minimum bias sample corresponding to $\sim 100\,\mu\text{b}^{-1}$ of integrated luminosity. The selected events correspond to LHC bunch crossing accompanied with activity in the beam scintillator on each side of the CMS detector. A veto on beam halo effect events has also been applied. The beam scraping events have been rejected and the presence of one good vertex ($|z| < 15\,\text{cm}$ and $N_{tracks} \geq 4$) has been requested. 4212310 events have been selected. No correction has been applied on the TrackJet momentum.

In order to compare the data with simulation, a multijet QCD sample produced by PYTHIA 8.1 and GEANT4 at $\sqrt{s} = 7\,\text{TeV}$ with simulation of realistic misalignment and miscalibration of the detector.

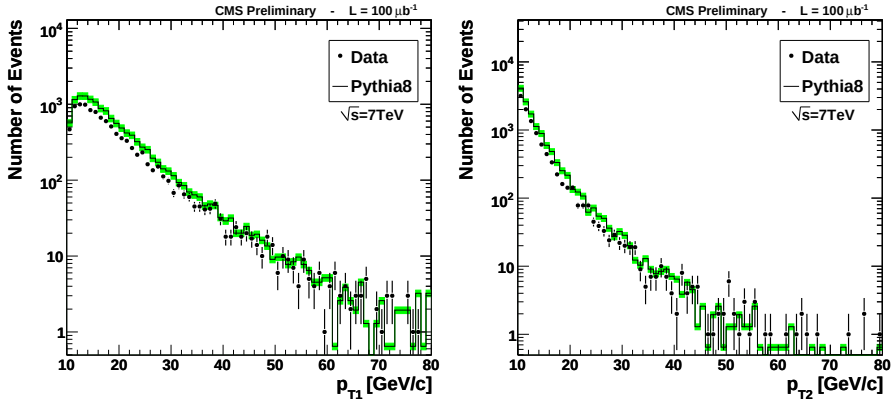
Events with high-p_T dijets provide a clean sample for the validation of jet reconstruction since they are expected to be reasonably well modelled by the leading order QCD as implemented in the PYTHIA event generator. Dijet events have been selected by requiring at least two reconstructed TrackJets with uncorrected $p_\text{T} > 10\,\text{GeV}/c$, which resulted in 10086 dijet events in data.

The p_T spectrum of the two leading jets are shown in Fig. A.2 where the simulation is normalized to the data luminosity estimate, and seems to overestimate the rate of dijets in the region between 10 and 30 GeV/c.

The distribution of the azimuthal angle between the two jets $\Delta\phi_{12}$ and of the momentum asymmetry of the two jets $(p_{\text{T}1} - p_{\text{T}2})/(p_{\text{T}1} + p_{\text{T}2})$ are shown in Fig. A.3 for selected dijet events. The data dijet events appear to be slightly more back-to-back than the simulation expectations. The momentum asymmetry observable of the dijets peaks towards zero which shows that the measured jets are balanced in p_T. Here the data reveals a slightly broader distribution with respect to the simulation.

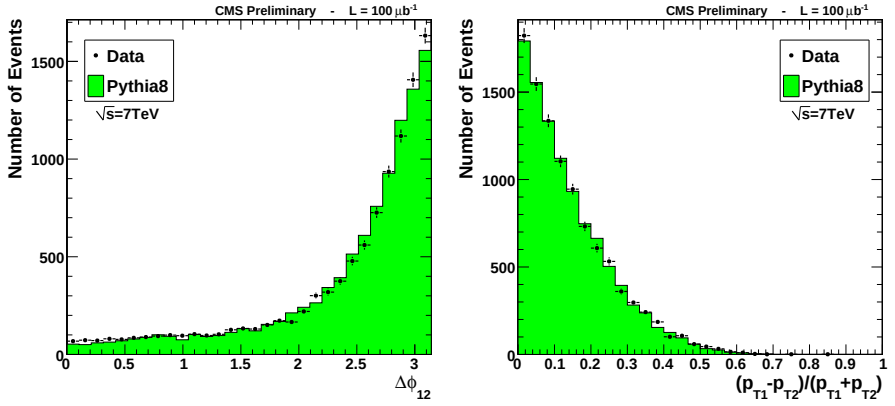
The inclusive jet analysis provides an overall validation for the performance of TrackJets in CMS. A minimum (uncorrected) $p_\text{T} > 10\,\text{GeV}/c$ is required on the TrackJets, in order to select a sample that is better modelled by the current simulation that is known to underestimate the production of low-p_T particles [148].

In Fig. A.4 the TrackJet multiplicity per event is shown, as well as the (uncorrected) p_T distribution for inclusive TrackJets, where simulation is normalized to the data luminosity estimate. The overall shape of both distributions is well described by the simulation. However, as for the dijet spectrum, the simulation overestimates the number of TrackJets in the region between 10 and 30 GeV/c. The η and ϕ distributions for



(a) (b)

Figure A.2: Transverse momentum (uncorrected) distribution of (a) the first and (b) the second leading TrackJets in events with at least two TrackJets. The simulations are normalized to the data luminosity estimate, and the green area reflects a 10% luminosity uncertainty.

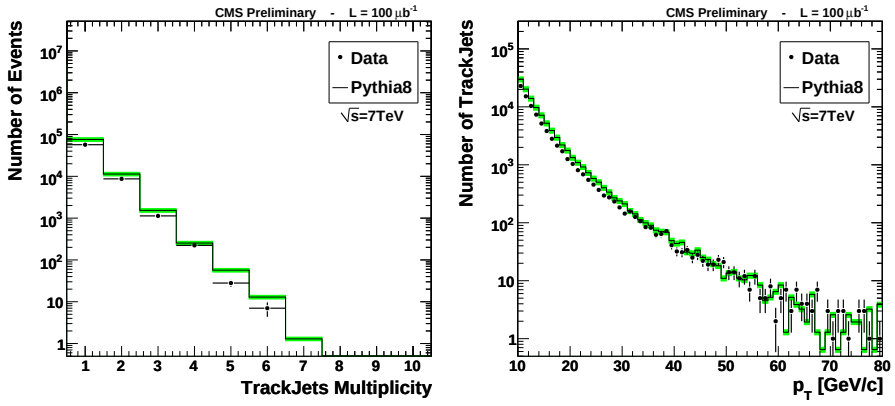


(a) (b)

Figure A.3: (a) Dijet azimuthal separation $\Delta\phi_{12}$ and (b) dijet momentum imbalance $(p_{\mathrm{T}1} - p_{\mathrm{T}2})/(p_{\mathrm{T}1} + p_{\mathrm{T}2})$ for di-TrackJet events with (uncorrected) $p_{\mathrm{T}} > 10\,\mathrm{GeV}/c$. The distributions are normalized to the number of events in each plot.

inclusive TrackJets, shown in Fig. A.5, exhibit an excellent agreement between data and simulation.

In Fig. A.6 constituent properties of TrackJets are compared between data and simulation. The general shape of these distributions is well described. There is however a trend that the number of tracks per jet is lower in the data than in the simulation, while the leading track p_{T} fraction is higher in the data. The excess of data at high leading-track p_{T} fraction corresponds to TrackJets consisting of only a few number of tracks. Tests with different MC tunes indicates that this can be attributed to the larger number of tracks per event in data compared to the underestimated simulation.



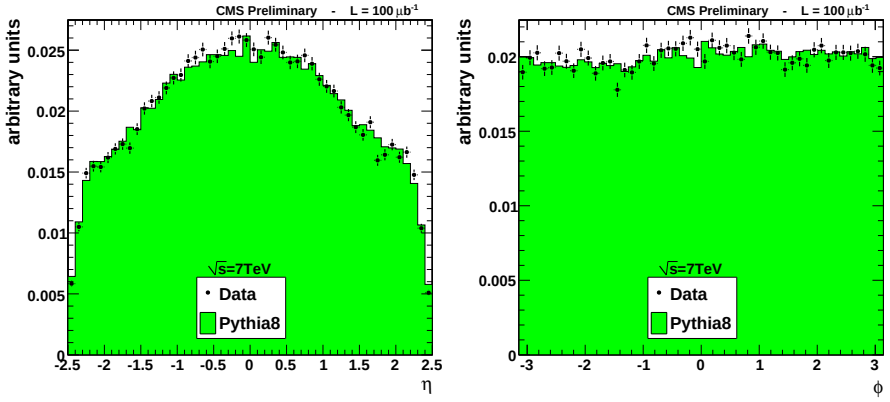
(a) (b)

Figure A.4: (a) TrackJet multiplicity per event and (b) p_{T} (uncorrected) for inclusive TrackJets. The simulations are normalized to the data luminosity estimate, and the green area reflects a 10% luminosity uncertainty.

A.3 Comparison of TrackJets to other jet reconstruction algorithms

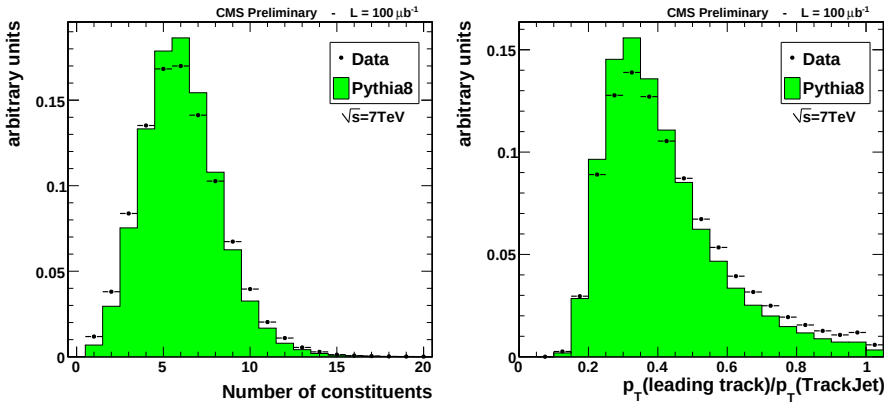
Given the very low mis-match rates and high jet finding efficiency of TrackJets, even at low transverse momenta, TrackJets can also be used to cross-check jets reconstructed using calorimetry (“CaloJets”) and particle-flow objects (“PFJets”).

A more robust jet measurement can be achieved if two or more jet reconstruction methods give the same results. In order to demonstrate this, I have realized a study comparing TrackJets on the one hand, and CaloJets or PFJets on the other hand.



(a) (b)

Figure A.5: Inclusive TrackJets (a) η and (b) ϕ distributions, normalized to the number of contributing TrackJets.



(a) (b)

Figure A.6: (a) Number of jet constituents and (b) leading track p_{T} fraction for inclusive TrackJets. The distributions are normalized to the number of TrackJets in each plot.

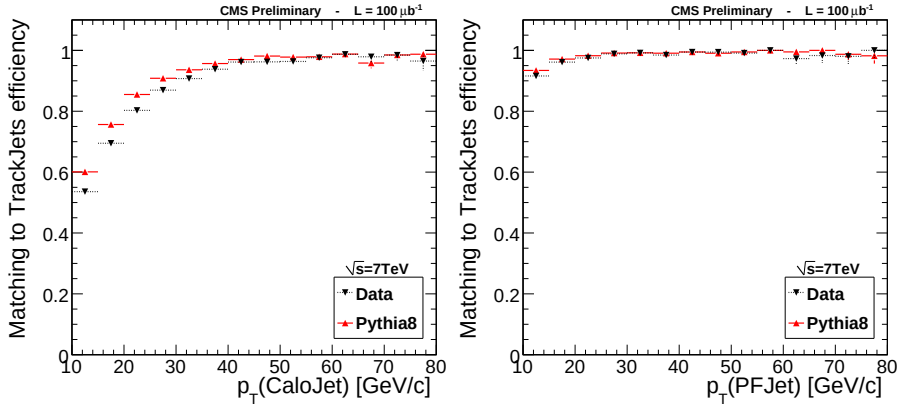
These results are obtained for the minimum bias sample described in the previous sections.

In the previous sections the TrackJet momenta are not corrected to the GenJet momentum scale, and only the measured tracking quantities are compared in the data and the simulation. In order to compare to "true" (GenJet) energies, a factor of roughly

1.7 needs to be applied to correct for the average fraction of neutral particles per jet and for the track selection inefficiencies. At the time of this study, the L2 and L3 corrections for TrackJets were not available and are therefore not applied.

In this section, for all jet types the corrected transverse momentum ($p_{\mathrm{T}}^{\mathrm{cor}} = 1.7 \times p_{\mathrm{T}}^{\mathrm{uncor}}$) is used to compare jets at the same momentum scale. The CaloJet and PFJet reconstructions follow the ones described in Ref. [123]. The two jets of dijet events ($|\Delta\phi_{12}| > \pi - 1$) with $p_{\mathrm{T}}^{\mathrm{cor}} > 10$ GeV/c are compared with the jets with $p_{\mathrm{T}}^{\mathrm{cor}} > 5$ GeV/c in the same event.

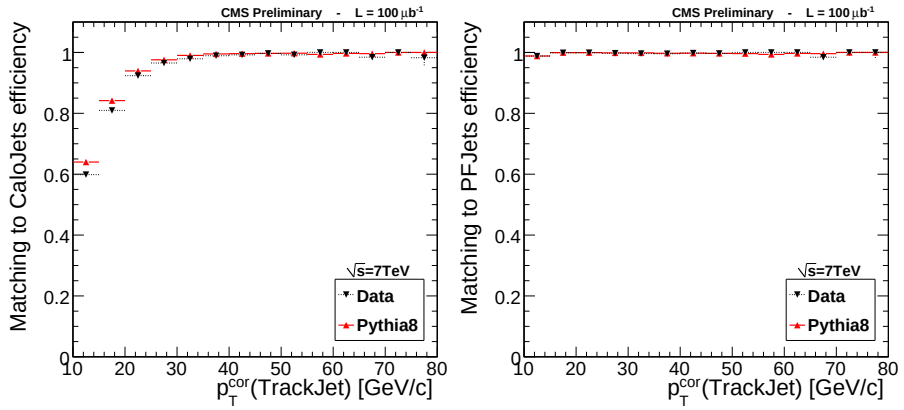
Based on the comparison of the jet directions, a matching efficiency is defined as the fraction of the matched jets within $\Delta R < 0.5$. This efficiency is first measured for CaloJets or PFJets with respect to TrackJets. Fig. A.7 shows these efficiencies as a function of CaloJet and PFJet transverse momenta, measured in data and simulation. A good agreement between the results in data and simulation is obtained and around 60% of the CaloJets at about 10 GeV of transverse momentum are also found as TrackJets. This fraction reaches up to 95% for CaloJets above 40 GeV. The matching efficiency between PFJets and TrackJets is better that 95% already just above 10 GeV/c.



(a) (b)

Figure A.7: Matching efficiency of CaloJets and PFJets to TrackJets in simulation and data.

In Fig. A.8, the efficiency of matching TrackJets to CaloJets or to PFJets is presented as a function of the TrackJet p_{T}. Just above 10 GeV/c, the cross-check between PFJets and TrackJets results in an efficiency close to 100% and is very consistent between data and simulation.



(a) (b)

Figure A.8: Matching efficiency of TrackJets with respect to CaloJets and PFJets in simulation and data.

Considering the high jet-finding efficiency of TrackJets, the low efficiency for CaloJets matching to TrackJets at low transverse momenta can be understood as mismeasurement of the CaloJet direction [125]. This is mostly because of charged particles deflecting in the strong 3.8 T magnetic field of the CMS detector.

In the case of PFJets the lower matching efficiency in the first bin is due to the $p_{\mathrm{T}}^{\mathrm{cor}} > 10$ GeV/c cut applied for PFJets, combined with the energy resolution of the TrackJet with $p_{\mathrm{T}}^{\mathrm{cor}} = 5$ GeV/c.

For corrected jet p_{T} above 40 GeV/c all three types of jet reconstruction find the jets to be in the same direction over 95% of the time, within the matching criteria.

A.4 Systematic effects and resilience against multiple interactions

Several instrumental effects can influence tracking efficiency and fake rates, affecting the performance for TrackJet reconstruction.

Two instrumental effects are considered that directly impact the data: the tracker noise and beam backgrounds.

The combined effect of electronics noise and beam induced noise has been assessed by analyzing events with single beam crossings with no minimum bias trigger activity. In

these events, the average of noisy cluster producing hits used for track reconstruction is 8 [98]. The probability to reconstruct a track from the noisy clusters is 10^{-5}.

The impact of physical beam backgrounds as beam-gas interactions have also been studied using both single beam events, with minimum bias activity, and double beam events without signals in the beam scintillator counters. In both cases no significant effect has been found that would disrupt the TrackJet reconstruction.

Two other systematic influences give rise to uncertainties in the knowledge of tracking efficiencies and fake rates. The tracker misalignment is found to have less than 0.1% impact on the TrackJet matching efficiency and mis-match rate. The material budget is estimated to have less than a 1% effect on tracking efficiency, and a negligible effect on the fake rate [148], witch leads to even smaller effects on the TrackJet performances. The effect of changing the track selection is studied, comparing the use of "loose" tracks [98] with the standard "high purity" ones. Though this is not a systematic uncertainty as such, the fact that this important change in track selection is found to only induce an effect on the TrackJet matching efficiency and mis-match rates below 1%, demonstrates the robustness of TrackJet reconstruction.

A very important effect influencing jet reconstruction in general at the LHC is the presence of multiple collisions in the same bunch-crossing, often referred to as pile-up. This effect will become increasingly important as LHC bunch intensities are increased and the beam is further reduced in transverse size.

To assess the importance of pile-up, the primary vertex reconstruction was redone using only tracks as selected throughout this note, and asking a longitudinal separation of the vertices to be larger than 1 cm.

In Fig. A.9 the distribution of H_{T}, the sum of the p_{T} of the jets in an event, is shown for TrackJets, where the sum runs over the TrackJets associated to each primary vertex separately, with each vertex contributing in the distribution. The fact that the measured H_{T} is the same for events with one or two vertices proves that TrackJets are robust in the presence of two pile-up interactions.

Given the primary vertex z resolution [98] and average expected separation, simulation studies [149] have shown that also in the presence of more pile-up interactions, TrackJets are expected to be transparent to pile-up effects even in higher LHC luminosity scenarios.

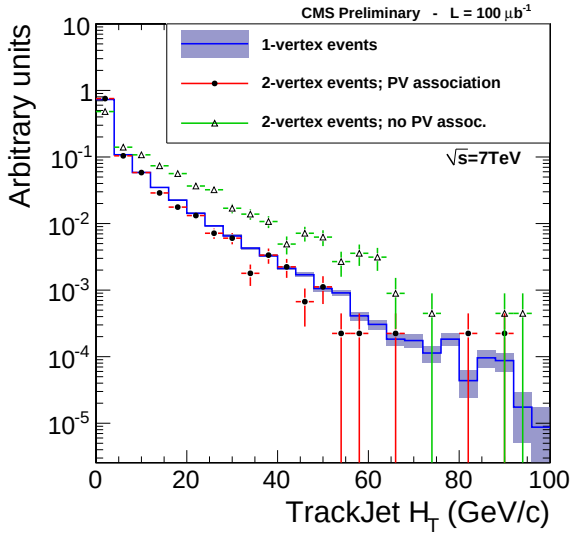


Figure A.9: TrackJet H_T for $p_{\mathrm{T}} > 3\,\mathrm{GeV}/c$ and $|\eta| < 2.5$, using jets associated to each primary vertex separately, for events with a single primary vertex and two primary vertices. The H_T distribution that is obtained from all TrackJets in two-vertex events, without associating the TrackJets to their vertex, is also shown.

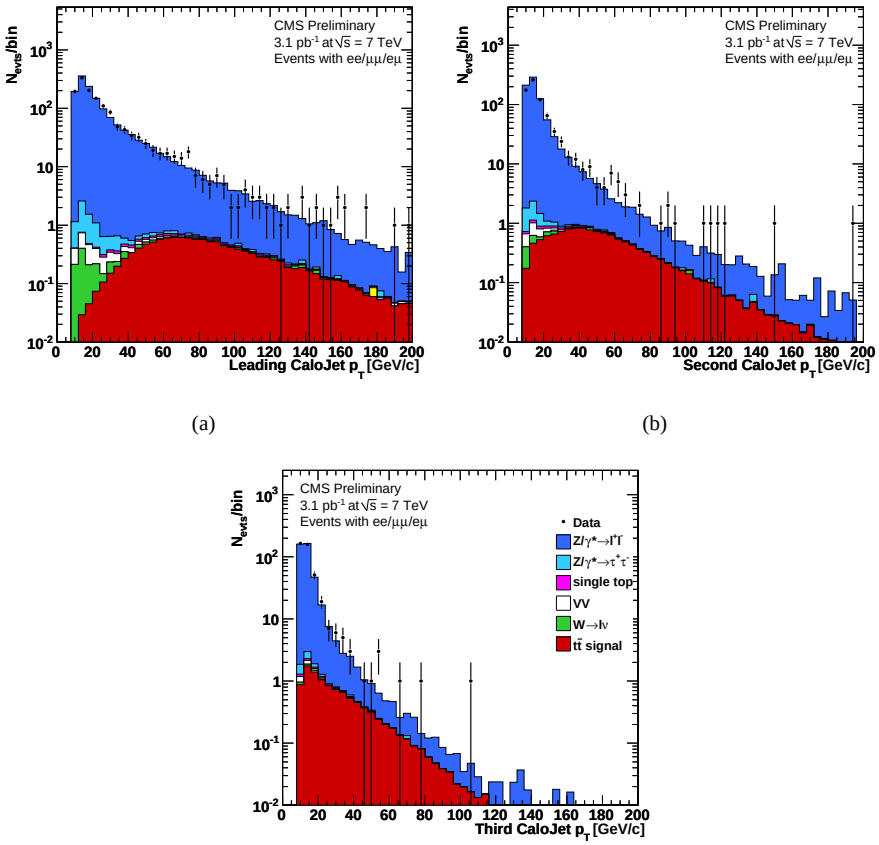
A.5 TrackJet and CaloJet comparison in the dileptonic $t\bar{t}$ framework

I have also tested the behaviour of the TrackJets in the dileptonic $t\bar{t}$ phase space, where both leptonic and hadronic activities are expected, and where the jets arise from heavy flavour fragmentation.

The definitions of dileptonic selection and the jet reconstruction and selection are the ones used in the study, described in Section 3.1.1. This implies that L2 and L3 correction are applied on both TrackJets and CaloJets. The TrackJets can be compared after dileptonic selection without Z-veto with CaloJets. The p_T spectrum of these jets are illustrated in Fig. 3.2 for TrackJets and Fig. A.10 for CaloJets. For each event, the two leading selected TrackJets (resp. CaloJets) can be found, and according to the presence or not of a CaloJets (resp. TrackJets), the matching efficiency can be obtained with respect to the TrackJets (resp. CaloJets) transverse momentum. These efficiencies are visible on Fig. A.11. Even if a large statistical uncertainty affects the results for data, these efficiencies are consistent between data and simulation expectation, especially for $p_T > 30\,\mathrm{GeV}/c$.

From each pair of matching jets, the two-dimensional plot comparing the TrackJets p_T and its matched CaloJets p_T can be obtained, after the dileptonic selection without the Z-veto, for data and for a Drell-Yan simulation, which is the dominant process in this selection, Fig. A.12. The correlation between the p_T of the paired jets while the matching depends only of the jet direction confirms that TrackJets and CaloJets are consistent.

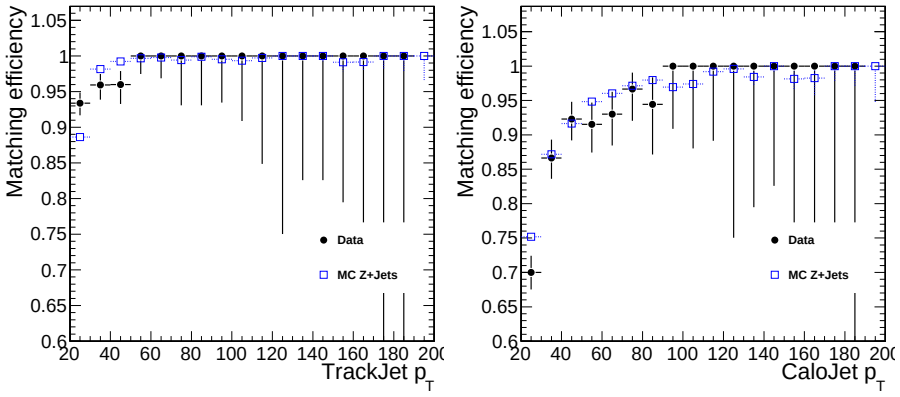
This also illustrates the effect of secondary vertex from the bottom quark fragmentation. I have indeed found that it has a negligible impact on the TrackJets reconstruction, the vertex reconstruction and primary vertex-track association parameters being not stringent enough to veto the tracks from this secondary vertex.



(a) (b)

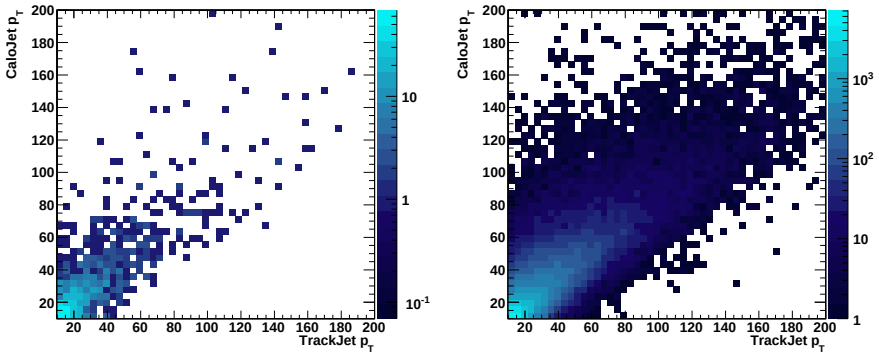
(c)

Figure A.10: Distributions of the transverse momentum of the first (a), second (b) and third (c) corrected CaloJets, after the dileptonic selection without the Z-veto, for all channels combined, after 3.1 pb^{-1} of integrated luminosity. The CaloJet is selected with the same condition that are applied to JPTJets in Section 3.1.2, excepted the p_T cut which has been relaxed up to $> 10\,\mathrm{GeV}/c$. The plain histograms are from simulation only, and the uncertainty from the number of simulated events are small enough to be neglected. The black dots correspond to data, with statistical uncertainty.



(a) (b)

Figure A.11: Matching efficiency for TrackJets to CaloJets (left) and for CaloJets to TrackJets (right), for the two leading jets in events passing the dileptonic selection without the Z-veto. The data is represented by the black dot and corresponds to $3.1\,\mathrm{pb}^{-1}$ of integrated luminosity. The simulation is represented by the blue square and is a Drell-Yan sample, which corresponds to the dominant process after this selection.



(a) (b)

Figure A.12: TrackJet p_T versus CaloJet p_T for the two leading jets in events passing the dileptonic selection without the Z-veto, for data (a) and simulation of the Drell-Yan process (b) which corresponds to the dominant process after this selection.

Bibliography

[1] V. Khachatryan et al., "First Measurement of the Cross Section for Top-Quark Pair Production in Proton-Proton Collisions at $\sqrt{s}$=7 TeV", *Phys.Lett.*, vol. B695, pp. 424–443, 2011, arXiv:1010.5994.

[2] Makoto Kobayashi and Toshihide Maskawa, "cp-violation in the renormalizable theory of weak interaction", *Progress of Theoretical Physics*, vol. 49, no. 2, pp. 652–657, 1973.

[3] M. L. Perl et al., "Evidence for anomalous lepton production in e^+-e^- annihilation", *Phys. Rev. Lett.*, vol. 35, no. 22, pp. 1489–1492, Dec 1975.

[4] S. W. Herb et al., "Observation of a dimuon resonance at 9.5 gev in 400-gev proton-nucleus collisions", *Phys. Rev. Lett.*, vol. 39, no. 5, pp. 252–255, Aug 1977.

[5] L. Montanet et al., "Review of particle properties. Particle Data Group", *Phys.Rev.*, vol. D50, pp. 1173–1823, 1994.

[6] F. Abe et al., "Observation of top quark production in $\bar{p}p$ collisions", *Phys.Rev.Lett.*, vol. 74, pp. 2626–2631, 1995, arXiv:hep-ex/9503002.

[7] S. Abachi et al., "Observation of the top quark", *Phys.Rev.Lett.*, vol. 74, pp. 2632–2637, 1995, arXiv:hep-ex/9503003.

[8] V.M. Abazov et al., "Evidence for production of single top quarks and first direct measurement of |Vtb|", *Phys.Rev.Lett.*, vol. 98, pp. 181802, 2007, arXiv:hep-ex/0612052.

[9] J. Caudron, A. Giammanco, and V. Lemaître, "Track-based measurement of top-quark pair events in the dileptonic channels with the first data of CMS", *CMS AN 2009/051*, 2009.

[10] J. Caudron, A. Giammanco, A. Pin, and V. Lemaître, "Measurement of the $t\bar{t}$ cross section in the dileptonic channels with track-jets at 7 TeV", *CMS AN 2010/195*, 2010.

[11] P. Azzurri, J. Caudron, D. Dobur, A. Hinzmann, T. Klimkovich, and S. Lowette, "Commissioning of Track-Jets with 900 GeV and 2360 GeV data ", *CMS AN 2010/015*, 2010.

[12] J. Caudron, A. Giammanco, A. Hinzmann, T. Klimkovich, P. Azzurri, S. Lowette, and D. Dobur, "Track-Jets Results with $\sqrt{s}$ =7 TeV pp Collisions ", *CMS AN 2010/079*, 2010.

[13] D. Barge et al., "A measurement of top quark pair production cross section in dilepton final states with early CMS data ", *CMS AN 2010/258*, 2010.

[14] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, "Parton distributions for the LHC", *Eur. Phys. J.*, vol. C63, pp. 189–285, 2009, arXiv:0901.0002.

[15] John M. Campbell, J. W. Huston, and W. J. Stirling, "Hard Interactions of Quarks and Gluons: A Primer for LHC Physics", *Rept. Prog. Phys.*, vol. 70, pp. 89, 2007, hep-ph/0611148.

[16] G.P. Korchemsky and A.V. Radyushkin, "Renormalization of the Wilson Loops Beyond the Leading Order", *Nucl.Phys.*, vol. B283, pp. 342–364, 1987.

[17] CMS Collaboration, "Combination of top pair production cross sections and comparisons with theory", *CMS Physics Analysis Summary, TOP-11-001*, 2011.

[18] M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer, et al., "HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR", *Comput.Phys.Commun.*, vol. 182, pp. 1034–1046, 2011, arXiv:1007.1327.

[19] M. Cacciari, S. Frixione, M.L. Mangano, P. Nason, and G. Ridolfi, "The t anti-t cross-section at 1.8 TeV and 1.96 TeV: A Study of the systematics due to parton densities and scale dependence", *JHEP*, vol. 0404, pp. 068, 2004, arXiv:hep-ph/0303085.

[20] S. Moch and P. Uwer, "Heavy-quark pair production at two loops in QCD", *Nucl.Phys.Proc.Suppl.*, vol. 183, pp. 75–80, 2008, arXiv:0807.2794.

[21] Nikolaos Kidonakis, "Top quark cross sections and differential distributions", 2011, arXiv:1105.3481.

[22] M. Cacciari, S. Frixione, M.L. Mangano, P. Nason, and G. Ridolfi, "Updated predictions for the total production cross sections of top and of heavier quark pairs at the Tevatron and at the LHC", *JHEP*, vol. 0809, pp. 127, 2008, arXiv:0804.2800.

[23] U. Langenfeld, S. Moch, and P. Uwer, "Measuring the running top-quark mass", *Phys.Rev.*, vol. D80, pp. 054009, 2009, arXiv:0906.5273.

[24] R. Kleiss and W.James Stirling, "Top quark production at hadron colliders: some useful formulae", *Z.Phys.*, vol. C40, pp. 419–423, 1988.

[25] Serguei Chatrchyan et al., "Measurement of the t-channel single top quark production cross section in pp collisions at $\sqrt{s}$ = 7 TeV", 2011, arXiv:1106.3052.

[26] John M. Campbell, Rikkert Frederix, Fabio Maltoni, and Francesco Tramontano, "NLO predictions for t-channel production of single top and fourth generation quarks at hadron colliders", *JHEP*, vol. 0910, pp. 042, 2009, arXiv:0907.3933.

[27] CMS Collaboration, "Measurement of the single-top t-channel cross section in pp collisions at $\sqrt{s}$=7 TeV", *CMS Physics Analysis Summary, TOP-10-008*, 2010.

[28] K. Nakamura et al., "Review of particle physics", *J.Phys.G*, vol. G37, pp. 075021, 2010.

[29] Johan Alwall, R. Frederix, J.-M. Gerard, A. Giammanco, M. Herquet, et al., "Is V(tb) = 1?", *Eur.Phys.J.*, vol. C49, pp. 791–801, 2007, arXiv:hep-ph/0607115.

[30] T. Aaltonen et al., "First Observation of Electroweak Single Top Quark Production", *Phys.Rev.Lett.*, vol. 103, pp. 092002, 2009, arXiv:0903.0885.

[31] V.M. Abazov et al., "Observation of Single Top Quark Production", *Phys.Rev.Lett.*, vol. 103, pp. 092001, 2009, arXiv:0903.0850.

[32] V.M. Abazov et al., "Simultaneous measurement of the ratio B($t \to Wb$) /B($t \to Wq$) and the top quark pair production cross section with the D0 detector at $\sqrt{s}$ = 1.96 TeV", *Phys.Rev.Lett.*, vol. 100, pp. 192003, 2008, arXiv:0801.1326.

[33] Ikaros I.Y. Bigi, Mikhail A. Shifman, N.G. Uraltsev, and A.I. Vainshtein, "The Pole mass of the heavy quark. Perturbation theory and beyond", *Phys.Rev.*, vol. D50, pp. 2234–2246, 1994, arXiv:hep-ph/9402360.

[34] M. Beneke and Vladimir M. Braun, "Heavy quark effective theory beyond perturbation theory: Renormalons, the pole mass and the residual mass term", *Nucl.Phys.*, vol. B426, pp. 301–343, 1994, arXiv:hep-ph/9402364.

[35] CDF, D0, et al., "Combination of CDF and D0 Results on the Mass of the Top Quark", 2010, arXiv:1007.3178.

[36] ALEPH, CDF, D0, DELPHI, L3, OPAL, SLD, LEP Electroweak Working Group, Tevatron Electroweak Working Group, SLD Electroweak Working Group, and Heavy Flavour Group, "Precision Electroweak Measurements and Constraints on the Standard Model", 2010, arXiv:1012.2367.

[37] V.M. Abazov et al., "Experimental discrimination between charge 2e/3 top quark and charge 4e/3 exotic quark production scenarios", *Phys.Rev.Lett.*, vol. 98, pp. 041801, 2007, arXiv:hep-ex/0608044.

[38] CDF Collaboration, "The CDF Measurement of the Top Quark Charge using the Top Decay Products in Lepton+Jet channel", *CDF conference note*, vol. 10460, 2011.

[39] CDF Collaboration, "Measurement of W Boson Helicity Fractions in Top Quark Decay to Lepton+Jets Events using a Matrix Element Analysis Technique with 2.7fb^{-1} of Data", *CDF conference note*, vol. 10004, 2009.

[40] V.M. Abazov et al., "First measurement of the forward-backward charge asymmetry in top quark pair production", *Phys.Rev.Lett.*, vol. 100, pp. 142002, 2008, arXiv:0712.0851.

[41] Johann H. Kuhn and German Rodrigo, "Charge asymmetry of heavy quarks at hadron colliders", *Phys.Rev.*, vol. D59, pp. 054017, 1999, arXiv:hep-ph/9807420.

[42] Oscar Antunano, Johann H. Kuhn, and German Rodrigo, "Top quarks, axigluons and charge asymmetries at hadron colliders", *Phys.Rev.*, vol. D77, pp. 014003, 2008, arXiv:0709.1652.

[43] Victor Mukhamedovich Abazov et al., "Forward-backward asymmetry in top quark-antiquark production", 2011, arXiv:1107.4995.

[44] T. Aaltonen et al., "Evidence for a Mass Dependent Forward-Backward Asymmetry in Top Quark Pair Production", *Phys.Rev.*, vol. D83, pp. 112003, 2011, arXiv:1101.0034.

[45] CMS Collaboration, "Measurement of the Charge Asymmetry in Top Quark Pair Production", *CMS Physics Analysis Summary, TOP-11-014*, 2011.

[46] ATLAS Collaboration, "Measurement of the charge asymmetry in top quark pair production in pp collisions at $\sqrt{s}$=7 tev using the atlas detector", *ATLAS Conference Note, ATLAS-CONF-2011-106*, 2011.

[47] T. Aaltonen et al., "Search for top-quark production via flavor-changing neutral currents in W+1 jet events at CDF", *Phys.Rev.Lett.*, vol. 102, pp. 151801, 2009, arXiv:0812.3400.

[48] T. Aaltonen et al., "Search for the Flavor Changing Neutral Current Decay $t \to Zq$ in p anti-p Collisions at s$^{(1/2)}$ = 1.96 TeV", *Phys.Rev.Lett.*, vol. 101, pp. 192002, 2008, arXiv:0805.2109.

[49] F. Abe et al., "Search for flavor-changing neutral current decays of the top quark in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV", *Phys.Rev.Lett.*, vol. 80, pp. 2525–2530, 1998.

[50] T L Cheng and P Teixeira-Dias, "Sensitivity of ATLAS to FCNC single top quark production", , no. ATL-PHYS-PUB-2006-029. ATL-COM-PHYS-2006-056, Aug 2006.

[51] C. Karafasoulis, A. Kyriakis, G. Vermisoglou, L. Benucci, A. Giammanco, and F. Palla, "Study of Flavour Changing Neutral Currents in Top Quark Decays with the CMS Detector.", *CMS Note 2006/093*, 2006.

[52] T. Stelzer and S. Willenbrock, "Spin correlation in top quark production at hadron colliders", *Phys.Lett.*, vol. B374, pp. 169–172, 1996, arXiv:hep-ph/9512292.

[53] CDF Collaboration, "Measurement of $t\bar{t}$ Helicity Fractions and Spin Correlation Using Reconstructed Lepton+Jets Events", *CDF conference note*, vol. 10211, 2010.

[54] Victor Mukhamedovich Abazov et al., "Measurement of spin correlation in ttbar production using dilepton final states", 2011, arXiv:1103.1871.

[55] H. Kovarik, "Eigenvalue bounds for two-dimensional magnetic Schroedinger operators", *ArXiv e-prints*, March 2011, arXiv:1103.5194.

[56] Rikkert Frederix and Fabio Maltoni, "Top pair invariant mass distribution: A Window on new physics", *JHEP*, vol. 0901, pp. 047, 2009, arXiv:0712.2355.

[57] Celine Degrande, Jean-Marc Gerard, Christophe Grojean, Fabio Maltoni, and Geraldine Servant, "Non-resonant New Physics in Top Pair Production at Hadron Colliders", *JHEP*, vol. 1103, pp. 125, 2011, arXiv:1010.6304.

[58] T. Aaltonen et al., "Search for Heavy Top-like Quarks Using Lepton Plus Jets Events in 1.96 TeV $p\bar{p}$ Collisions", *Phys.Rev.Lett.*, vol. 100, pp. 161803, 2008, arXiv:0801.3877.

[59] CMS Collaboration, "Search for Same-Sign Top-Quark Pair Production at $\sqrt{s}$ = 7 TeV and Limits on Flavour Changing Neutral Currents in the Top Sector", 2011, arXiv:1106.2142.

[60] CMS Collaboration, "Search for Heavy Stable Charged Particles in pp collisions at $\sqrt{s}$ = 7 TeV", *CMS Physics Analysis Summary, EXO-11-022*, 2011.

[61] CMS Collaboration, "Search for pair production of a fourth-generation t' quark in the lepton-plus-jets channel with the CMS experiment", *CMS Physics Analysis Summary, EXO-11-051*, 2011.

[62] CMS Collaboration, "Search for a Top-like Quark Decaying To A Top Quark And a Z boson in pp Collisions at $\sqrt{s}$ = 7 TeV", *CMS Physics Analysis Summary, EXO-11-005*, 2011.

[63] W. Braunschweig et al., "a study of bhabha scattering at Petra energies", *Z.Phys.*, vol. C37, pp. 171, 1988.

[64] H. Albrecht et al., "Observation of B0 - anti-B0 Mixing", *Phys.Lett.*, vol. B192, pp. 245, 1987.

[65] Gordon L. Kane and Michael E. Peskin, "A Constraint from B Decay on Models with No T Quark", *Nucl.Phys.*, vol. B195, pp. 29, 1982.

[66] A. Bean et al., "Improved Upper Limit on Flavor Changing Neutral Current Decays of the B Quark", *Phys.Rev.*, vol. D35, pp. 3533, 1987.

[67] D.P. Roy and S. Uma Sankar, "B(d)0 - anti-B(d)0 mixing as the evdience for the existence of the top quark", *Phys.Lett.*, vol. B243, pp. 296–300, 1990.

[68] Sandip Pakvasa, D.P. Roy, and S. Uma Sankar, "phenomenological analysis of a topless left-right model", *Phys.Rev.*, vol. D42, pp. 3160–3168, 1990.

[69] F. Abe et al., "The CDF Detector: An Overview", *Nucl.Instrum.Meth.*, vol. A271, pp. 387–403, 1988.

[70] R. Blair et al., "The CDF-II detector: Technical design report", 1996.

[71] D. Acosta et al., "Measurement of the J/ψ meson and $b-$hadron production cross sections in $p\bar{p}$ collisions at $\sqrt{s} = 1960$ GeV", *Phys.Rev.*, vol. D71, pp. 032001, 2005, arXiv:hep-ex/0412071.

[72] S. Abachi et al., "The D0 Detector", *Nucl.Instrum.Meth.*, vol. A338, pp. 185–253, 1994.

[73] V.M. Abazov et al., "The Upgraded D0 detector", *Nucl.Instrum.Meth.*, vol. A565, pp. 463–537, 2006, arXiv:physics/0507191.

[74] (ed.) Evans, Lyndon and (ed.) Bryant, Philip, "LHC Machine", *JINST*, vol. 3, pp. S08001, 2008.

[75] G. Aad et al., "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST*, vol. 3, pp. S08003, 2008.

[76] K. Aamodt et al., "The ALICE experiment at the CERN LHC", *JINST*, vol. 3, pp. S08002, 2008.

[77] R. Adolphi et al., "The CMS experiment at the CERN LHC", *JINST*, vol. 3, pp. S08004, 2008.

[78] A.Augusto Alves et al., "The LHCb Detector at the LHC", *JINST*, vol. 3, pp. S08005, 2008.

[79] John M. Campbell and R. Keith Ellis, "An Update on vector boson pair production at hadron colliders", *Phys.Rev.*, vol. D60, pp. 113006, 1999, arXiv:hep-ph/9905386.

[80] John M. Campbell and R.K. Ellis, "MCFM for the Tevatron and the LHC", *Nucl.Phys.Proc.Suppl.*, vol. 205-206, pp. 10–15, 2010, arXiv:1007.3492.

[81] CMS Collaboration, "Rates of Jets Produced in Association with W and Z Bosons", *CMS Physics Analysis Summary, EWK-10-012*, 2010.

[82] Johan Alwall, Pavel Demin, Simon de Visscher, Rikkert Frederix, Michel Herquet, et al., "MadGraph/MadEvent v4: The New Web Generation", *JHEP*, vol. 0709, pp. 028, 2007, arXiv:0706.2334.

[83] F. Abe et al., "Evidence for top quark production in $\bar{p}p$ collisions at $\sqrt{s} = 1.8$ TeV", *Phys.Rev.*, vol. D50, pp. 2966–3026, 1994.

[84] F. Abe et al., "Evidence for top quark production in $\bar{p}p$ collisions at $\sqrt{s} = 1.8$ TeV", *Phys.Rev.Lett.*, vol. 73, pp. 225–231, 1994, arXiv:hep-ex/9405005.

[85] S. Abachi et al., "Search for high mass top quark production in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV", *Phys.Rev.Lett.*, vol. 74, pp. 2422–2426, 1995, arXiv:hep-ex/9411001.

[86] S. Abachi et al., "Search for the top quark in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV", *Phys.Rev.Lett.*, vol. 72, pp. 2138–2142, 1994.

[87] "$t\bar{t}$ discovery Fermilab News release", http://www.fnal.gov/pub/inquiring/physics/discoveries/pr/top_news_release.html.

[88] "D0 Top physics results", http://www-d0.fnal.gov/Run2Physics/WWW/results/top.htm.

[89] "CDF Top physics results", http://www-cdf.fnal.gov/physics/new/top/top.html.

[90] CMS Collaboration, "Measurement of CMS Luminosity", *CMS Physics Analysis Summary, EWK-10-004*, 2010.

[91] S. Dasu et al., "CMS. The TriDAS project. Technical design report, vol. 1: The trigger systems", *CERN-LHCC-2000-038*, 2000.

[92] (ed.) Sphicas, P., "CMS: The TriDAS project. Technical design report, Vol. 2: Data acquisition and high-level trigger", *CERN-LHCC-2002-026*, 2002.

[93] A. Abulencia et al., "Measurements of inclusive W and Z cross sections in p anti-p collisions at s$^{(1/2)}$ = 1.96 TeV", *J.Phys.G*, vol. G34, pp. 2457–2544, 2007, arXiv:hep-ex/0508029.

[94] M. Asai, "Geant4-a simulation toolkit", *Trans.Amer.Nucl.Soc.*, vol. 95, pp. 757, 2006.

[95] W. Adam, V. Adler, B. Hegner, L. Lista, S. Lowette, et al., "PAT: The CMS physics analysis toolkit", *J.Phys.Conf.Ser.*, vol. 219, pp. 032017, 2010.

[96] Fabio Maltoni and Tim Stelzer, "MadEvent: Automatic event generation with MadGraph", *JHEP*, vol. 0302, pp. 027, 2003, arXiv:hep-ph/0208156.

[97] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands, "PYTHIA 6.4 Physics and Manual", *JHEP*, vol. 0605, pp. 026, 2006, arXiv:hep-ph/0603175.

[98] Vardan Khachatryan et al., "CMS Tracking Performance Results from early LHC Operation", *Eur.Phys.J.*, vol. C70, pp. 1165–1192, 2010, arXiv:1007.1988.

[99] "CMS, tracker technical design report", *CERN-LHCC-98-06*, 1998.

[100] "Addendum to the CMS tracker TDR", *CERN-LHCC-2000-016*, 2000.

[101] CMS Collaboration, "Track reconstruction in the CMS Tracker", *CMS Physics Analysis Summary, TRK-09-001*, 2009.

[102] CMS Collaboration, "Tracking and Primary Vertex Results in First 7 TeV Collisions", *CMS Physics Analysis Summary, TRK-10-005*, 2010.

[103] Wolfgang Waltenberger, "Adaptive vertex reconstruction", *CMS Note 2008/033*, 2008.

[104] CMS Collaboration, "Measurement of Tracking Efficiency", *CMS Physics Analysis Summary, TRK-10-002*, 2010.

[105] CMS Collaboration, "Measurement of Momentum Scale and Resolution using Low-mass Resonances and Cosmic Ray Muons", *CMS Physics Analysis Summary, TRK-10-004*, 2010.

[106] "CMS, the Compact Muon Solenoid. Muon technical design report", *CERN-LHCC-97-32*, 1997.

[107] Serguei Chatrchyan et al., "Performance of CMS Muon Reconstruction in Cosmic-Ray Events", *JINST*, vol. 5, pp. T03022, 2010, arXiv:0911.4994.

[108] "CMS: The electromagnetic calorimeter. Technical design report", *CERN-LHCC-97-33*, 1997.

[109] S. Baffioni, C. Charlot, F. Ferri, D. Futyan, P. Meridiani, et al., "Electron reconstruction in CMS", *Eur.Phys.J.*, vol. C49, pp. 1099–1116, 2007.

[110] CMS Collaboration, "Particle flow reconstruction of 0.9 TeV and 2.36 TeV collision events in CMS", *CMS Physics Analysis Summary, PFT-10-001*, 2010.

[111] 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 Physics*, vol. 31, pp. 9–+, September 2005, arXiv:physics/0306087.

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

[113] CMS Collaboration, "Electron reconstruction and identification at $\sqrt{s}$ = 7 TeV", *CMS Physics Analysis Summary, EGM-10-004*, 2010.

[114] "CMS: The hadron calorimeter. Technical design report", *CERN-LHCC-97-31*, 1997.

[115] Matteo Cacciari and Gavin P. Salam, "Dispelling the N^3 myth for the k_t jet-finder", *Phys.Lett.*, vol. B641, pp. 57–61, 2006, arXiv:hep-ph/0512210.

[116] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez, "The Anti-k(t) jet clustering algorithm", *JHEP*, vol. 0804, pp. 063, 2008, arXiv:0802.1189.

[117] Serguei Chatrchyan et al., "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS.", 2011, arXiv:1107.4277.

[118] A. Heister, Olga Kodolova, V. Konoplyanikov, S. Petrushanko, J. Rohlf, et al., "Measurement of jets with the CMS detector at the LHC", *CMS Note 2006/036*, 2006.

[119] CMS Collaboration, "Commissioning of TrackJets in pp Collisions at $\sqrt{s}$=7 TeV", *CMS Physics Analysis Summary, JME-10-006*, 2010.

[120] CMS Collaboration, "The Jet Plus Tracks Algorithm", *CMS Physics Analysis Summary, JME-09-002*, 2009.

[121] CMS Collaboration, "Particle Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", *CMS Physics Analysis Summary, PFT-09-001*, 2009.

[122] CMS Collaboration, "Measurement of the Jet Energy Resolutions and Jet Reconstruction Efficiency at CMS", *CMS Physics Analysis Summary, JME-09-007*, 2009.

[123] CMS Collaboration, "Jet Performance in pp Collisions at $\sqrt{s}$=7 TeV", *CMS Physics Analysis Summary, JME-10-003*, 2010.

[124] Serguei Chatrchyan et al., "Missing transverse energy performance of the CMS detector.", 2011, arXiv:1106.5048.

[125] CMS Collaboration, "Performance of Jet Algorithms in CMS", *CMS Physics Analysis Summary, JME-07-003*, 2007.

[126] CMS Collaboration, "Performance of Track Corrected Missing ET in CMS", *CMS Physics Analysis Summary, JME-09-010*, 2009.

[127] CMS Collaboration, "Commissioning of b-jet identification with pp collisions at $\sqrt{s}$ = 7TeV", *CMS Physics Analysis Summary, BTV-10-001*, 2010.

[128] Serguei Chatrchyan et al., "Measurement of the t t-bar production cross section and the top quark mass in the dilepton channel in pp collisions at $\sqrt{s}$ =7 TeV", *JHEP*, vol. 1107, pp. 049, 2011, arXiv:1105.5661.

[129] A. Abulencia et al., "Measurement of the top quark mass using template methods on dilepton events in proton antiproton collisions at $\sqrt{s} = 1.96$TeV.", *Phys.Rev.*, vol. D73, pp. 112006, 2006, arXiv:hep-ex/0602008.

[130] Victor Mukhamedovich Abazov et al., "Measurement of the top quark mass in final states with two leptons", *Phys.Rev.*, vol. D80, pp. 092006, 2009, arXiv:0904.3195.

[131] CMS Collaboration, "Early ttbar cross section in the dilepton channel at $\sqrt{s}$ = 10 TeV", *CMS Physics Analysis Summary, TOP-09-002*, 2009.

[132] W. Andrews et al., "Data-driven methods to estimate the electron and muon fake contributions to lepton analyses", *CMS AN 2009/041*, 2009.

[133] D.J. Lange, "The EvtGen particle decay simulation package", *Nucl.Instrum.Meth.*, vol. A462, pp. 152–155, 2001.

[134] S. Jadach, Z. Was, R. Decker, and Johann H. Kuhn, "The tau decay library TAUOLA: Version 2.4", *Comput.Phys.Commun.*, vol. 76, pp. 361–380, 1993.

[135] Daniel Stump, "Uncertainties of parton distribution functions", Prepared for PHYSTAT2003: Statistical Problems in Particle Physics, Astrophysics, and Cosmology, Menlo Park, California, 8-11 Sep 2003.

[136] Federico Demartin, Stefano Forte, Elisa Mariani, Juan Rojo, and Alessandro Vicini, "The impact of PDF and alphas uncertainties on Higgs Production in gluon fusion at hadron colliders", *Phys.Rev.*, vol. D82, pp. 014002, 2010, arXiv:1004.0962.

[137] Hung-Liang Lai, Joey Huston, Zhao Li, Pavel Nadolsky, Jon Pumplin, et al., "Uncertainty induced by QCD coupling in the CTEQ global analysis of parton distributions", *Phys.Rev.*, vol. D82, pp. 054021, 2010, arXiv:1004.4624.

[138] F. Abe et al., "Measurement of the top quark mass and $t\bar{t}$ production cross section from dilepton events at the Collider Detector at Fermilab", *Phys.Rev.Lett.*, vol. 80, pp. 2779–2784, 1998, arXiv:hep-ex/9802017.

[139] Darin E. Acosta et al., "Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV using dilepton events", *Phys.Rev.Lett.*, vol. 93, pp. 142001, 2004, arXiv:hep-ex/0404036.

[140] Louis Lyons, Duncan Gibaut, and Peter Clifford, "How to combine correlated estimates of a single physical quantity", *Nucl.Instrum.Meth.*, vol. A270, pp. 110, 1988.

[141] CMS Collaboration, "Absolute luminosity normalization", *CMS Detector Performance Summary, DP-11-003*, 2011.

[142] Serguei Chatrchyan et al., "Measurement of the t t-bar Production Cross Section in pp Collisions at 7 TeV in Lepton + Jets Events Using b-quark Jet Identification", 2011, arXiv:1108.3773.

[143] Serguei Chatrchyan et al., "Measurement of the Top-antitop Production Cross Section in pp Collisions at $\sqrt{s}$=7 TeV using the Kinematic Properties of Events with Leptons and Jets", 2011, arXiv:1106.0902.

[144] Victor Mukhamedovich Abazov et al., "Measurement of the top quark pair production cross section in the lepton+jets channel in proton-antiproton collisions at $\sqrt{s}$=1.96 TeV", 2011, arXiv:1101.0124.

[145] T. Aaltonen et al., "First Measurement of the Ratio $\sigma(t-\bar{t})/\sigma(Z/\gamma^* \rightarrow \ell\ell)$ and Precise Extraction of the t-tbar Cross Section", *Phys.Rev.Lett.*, vol. 105, pp. 012001, 2010, arXiv:1004.3224.

[146] T. Aaltonen et al., "First Measurement of the b-jet Cross Section in Events with a W Boson in p anti-p Collisions at s$^{(1/2)}$ = 1.96 TeV", *Phys.Rev.Lett.*, vol. 104, pp. 131801, 2010, arXiv:0909.1505.

[147] CMS Collaboration, "Measurement of the $t\bar{t}$ production cross section in the dilepton channel in pp collisions at $\sqrt{s}$ = 7 TeV with a luminosity of 1.14 fb^{-1}", *CMS Physics Analysis Summary, TOP-11-005*, 2011.

[148] Vardan Khachatryan et al., "First Measurement of the Underlying Event Activity at the LHC with $\sqrt{s}$ = 0.9 TeV", *Eur.Phys.J.*, vol. C70, pp. 555–572, 2010, arXiv:1006.2083.

[149] CMS Collaboration, "Performance of Jet Reconstruction with Charged Tracks only", *CMS Physics Analysis Summary, JME-08-001*, 2008.

List of Tables

List of Figures