Università degli Studi di Pavia,
Dipartimento di Fisica

Istituto Nazionale di Fisica Nucleare,
Sezione di Pavia

DOTTORATO DI RICERCA IN FISICA – XXVI CICLO

Search for baryon number violation in top-quark decays with the CMS Experiment

by Michele Gabusi

Submitted to the Graduate School in Physics
in Partial Fulfillment of the Requirements
for the Degree of

DOTTORE DI RICERCA IN FISICA
DOCTOR OF PHILOSOPHY IN PHYSICS

at the
University of Pavia

Advisors: Prof. P. Vitulo (Università di Pavia)
Prof. G. Bruno (Université Catholique de Louvain)

Referees: Prof. A. Colaleo (Università di Bari)
Prof. D. Pedrini (Università di Milano Bicocca)

Cover: A simulated event displaying a pair production of top quarks at the CMS Experiment, where one top quark undergoes a Standard Model decay into three jets and the other a baryon number violating decay.

Search for baryon number violation in top-quark decays with the CMS Experiment

Michele Gabusi
PhD thesis – University of Pavia
Printed in Pavia, Italy, October 2013
ISBN 978-88-95767-63-5

Questo omo ha una somma pazzia:
cioè che sempre stenta per non istentare, e la vita se li fugge
sotto speranza di godere i beni con somma fatica acquistati.
(*Leonardo da Vinci*)

And this man has an eternal madness:
he struggles not to struggle, while the life goes by,
wishing to enjoy the goods so hardly achieved.

Contents

List of Figures

List of Tables

Introduction

Our understanding of matter and energy dynamics lies on a set of fundamental theories grown up during 1960s and 1970s.
This description, called "Standard Model" (SM), reduces all the known phenomena (except gravity) to simple interactions between elementary particles. The SM originally incorporated just quantum electrodynamics, which was afterwards extended with the theory of electroweak processes by Glashow (1960)[1]; Salam and Weinberg then inserted into Glashow's theory [2][3] the Higgs Mechanism (1967), introduced in 1964 by Higgs and, independently, by Brout and Englert (who were working on the idea proposed by Philip Anderson), Guarlnik, Hagen, and Kibble. The aim of their works was to give an explanation [5] to the Electroweak Symmetry Breaking (EWSB); finally Wilczek and Gross during the first 1970s gave to quantum chromodynamics its actual form.
The goal of the SM is to provide a global, self-consistent theory which could provide an accurate description of the behavior of energy and matter.
In fact the full consistency of SM was recognized in 1978 [4]. A huge experimental effort has been carried out to carefully check the predictions of the Theory. The milestones of this long (and not yet accomplished) process started in the first years of 1960s, though the first big success came in 1974, when J/Ψ meson was simultaneously discovered at Brookhaven and SLAC [6]. The success of SM was confirmed by the discovery of W and Z bosons (1982), of top quark (1995) and finally by the finding of a boson (2012), identified tentatively as the "Higgs Boson".
However the development of the theory has been also endorsed by an impressive technological development which has allowed the production of more and more powerful accelerators. The Large Hadron Collider (LHC) is just the last, most powerful accelerator of an outstanding tradition of colliders, whose leading actors were Tevatron (proton-antiproton collider, at Fermilab) and LEP (electron-positron collider, at CERN). The LHC is the circular accelerator built in the same 27 km long tunnel where LEP was hosted, near Geneva. Designed to reach a center of mass energy of 14 TeV, it can provide an instantaneous luminosity of about $10^{34}\ /(s \cdot cm^2)$. The collisions between the

two proton beams occur in four points along the ring, where the four main detectors are placed: ATLAS, ALICE, CMS, LHCb. ATLAS and CMS have been conceived as general purposes detectors, mainly aiming for the discovery of new physics and test the SM. LHCb and ALICE are born to be employed to study B-physics and quark-gluon plasma studies, respectively.
Indeed SM met almost every experimental test so far, despite a number of open questions and pitfalls must be still clarified. The discovery of the Higgs boson can be considered the most remarkable achievement of the whole LHC scientific program so far, even though a full characterization of Higgs boson is still missing.
Many physicists are confident that a new Physics could emerge beyond the TeV scale (where SM might become unsatisfactory). A number of hypothesis are being tested in fact. Supersimmetry (SUSY) is one of the best known and most accepted extensions of SM and - even if new predicted particles are not observed yet - many efforts are currently addressed to test its reliability. The origin and constituents of Dark matter are still far from being understood. Much should be still learnt about CP violation, which could be one of the basic ingredients of asymmetry between matter and antimatter; no new generation of quarks, or new families of particles (related to Technicolor theories) have been observed yet; GUT theories, which provide a frame to link weak, electromagnetic and strong interactions have not been confirmed yet, even though the SM fails meeting running gauge coupling constant at the GUT scale.
In such an exciting and diverse framework of new and old theories, baryon number conservation still remains an important object of debate.
In fact baryon number has been introduced for the first time in the 1920s [24], before the discovery of the positron, as a kind of particles "heavy charge" [4]: the existence of a symmetry law that granted the conservation of baryon number appeared very soon as a natural statement in the physicists community. Such conservation was needed to assure, for instance, the stability of proton (and of the known matter), even though it could not be used to justify the observation of antimatter [27] in the universe.
Whereas the difficulty to prove the existence of proton decay from an experimental point of view became immediately evident [26], in few years the number of arguments in favor of a (global) symmetry breaking increased dramatically. Huge experimental efforts have been made during the second half of nineteenth century to check experimentally the possibility of baryon number violation in several different processes. Both direct and indirect measurements have been performed: proton lifetime has been measured with increasing precision for years [42][43]; high precision limits to heavy mesons [44], Z boson and τ decays [47][46] have been set. Tighter and tighter constraints to baryon number violation have been introduced but no clues of symmetry breaking have been found so far. On the other hand, a number of theories still rely on baryon (and lepton) number violation.
In this complex scenario, LHC represents a unique opportunity to test con-

servation laws at TeV scale, both in production and decay channels. This work of thesis, based on the effective model presented in [7], aims to look for any evidence of baryon number violation (BNV), looking at top quark decays produced in pp collisions at LHC. A new point-like, four-fermion particle interaction has been introduced, without making any preliminary assumptions on the existence of a possible (heavy) mediator. When a top quark undergoes such process, it transforms into one lepton and two jets without any neutrinos in the final state.
Thanks to their relatively high production cross section, $t\bar{t}$ pairs are the most suitable objects to observe such rare decay. Indeed due to the low expected branching ratio ($BR \ll 1$), BNV process in single top channel is almost impossible to observe; on the other hand, also $t\bar{t}$ decays in dileptonic channel are expected to be strongly suppressed. Hence we look for any evidence of new signal in semileptonic $t\bar{t}$ decays, where one top decays hadronically and the other undergoes a BNV process.
In 2011 about 5 fb^{-1} of p-p collision collected by CMS Experiment at $\sqrt{s} = 7$ have been analyzed [68]. In 2012 a new search has been performed using 19.6 fb^{-1} of data at $\sqrt{s} = 8$ TeV [69]. Some improvements have been implemented thanks to the experience of the previous analysis.
Our analysis strategy (based on a counting experiment) has been conceived to be as model independent as possible and robust against a number of systematic uncertainties; a data-driven approach has been adopted to estimate QCD; a customized, template-morphing oriented statistical treatment has been implemented to give an estimate of decay BR. All the cuts have been tuned maximizing the expected limit for the searched signal. However, both in 2011 and 2012 no significant excess of data over SM expectations has been found, and an upper limit on BNV decay has been set.
This work of thesis has been structured as follows: in Chapter 1 a quick review of SM formulation, with some references to top Physics, will be presented; after few historical remarks about the debate on baryon number conservation, the model which has inspired this thesis will be described in detail in Section 1.2.3, giving also some further perspective for future searches; finally a not exhaustive review of the most relevant experimental results collected so far will be presented. A picture of Large Hadron Collider and CMS Experiment will be given in Chapter 2; in Chapter 3 some more detail about the object reconstruction techniques which have been used in this analysis will be examined. Chapters 4 and 5 will deal more specifically with the core of this work: in the former, the analysis strategy will be presented, focusing on the differences between 2011 and 2012 approaches; in the latter, 2011 and 2012 results will be reported, as well as their statistical interpretation. In Chapter 6, some cross checks, which have been used to demonstrate the robustness of the analysis strategy and to support our results, will be presented.

Chapter 1

The Standard Model of particle physics

1.1 Standard Model

1.1.1 Elementary particles and fundamental interactions

According to the SM description all the visible matter is made up by *fermions*, particles with half spin, which interact through the exchange of *bosons*, particles with integer spin, as a result of gauge-invariant theories.
The fundamental fermions are divided in three generations, further classified in leptons and quarks.
Each lepton carries an electric unitary charge q, and a "lepton number" which identifies the generation which the lepton belongs to (Table 1.1). Six flavors of quarks also exist which can be split in three generations, too. One can associate each quark to an electric charge and to a common baryon number B.

All the "stable" matter around us is made by fermions (leptons and quarks) which belong to the first generation. Actually due to their high masses, second and third generation quarks can be produced just in extreme conditions,

Generation	Charge	L_e	L_μ	L_τ
e, (ν_e)	-1 (0)	1	0	0
μ, (ν_μ)	-1 (0)	0	1	0
τ, (ν_τ)	-1 (0)	0	0	1

Table 1.1: Six leptons (and six anti-leptons), namely electron, muon, tau and all the related neutrinos, can be classified. For each generation a particle and a neutrino (ν) can also be identified. Lepton number of neutrinos matches the lepton number of the lepton of the same generation. Lepton number change sign for charge conjugation.

Generation	Charge	B
d	-1/3	1/3
u	2/3	1/3
s	-1/3	1/3
c	2/3	1/3
b	-1/3	1/3
t	2/3	1/3

Table 1.2: List of all quarks that make up the Standard model; the charge are expressed as fraction of electron charge; charge conjugation flip the charge of baryon number B.

in accelerators or in the primordial universe.
All the atomic nuclei are made by neutrons and protons, which are also called *baryons*. Baryons are fermions made up by three quarks, having integer baryon number and integer charge; on the contrary bound states of a quark and an antiquark are called *mesons* and due to their integer spin, they are bosons. Since baryon and mesons are both bounded by strong forces, they are also called "hadrons".
Four kinds of interactions between quarks and leptons are known, respectively mediated by massless photons (electromagnetic field), massive $W^{\pm}$ and Z^0 bosons (weak interactions), massless gluons for strong interactions and graviton (not discovered yet) for gravity.
Remarkably the asymmetry between massless photons and massive weak bosons can be solved invoking the Higgs Mechanism, which spontaneously breaks gauge symmetry. This process leaves the lagrangian invariant under gauge symmetries, while makes the ground state not invariant.
Finally, since strong interactions introduce a further charge, called "color", quarks and gluons are also identified by a quantum number which can assume three states: *red*, *green* and *blue*. In fact Hadrons exist just in uncolored states and any attempt to isolate single quarks give rise to a new quark pair. This phenomenon, known as "quark confinement", it is due to the asymptotic freedom of strong interactions, whose intensity decreases with the energy.

1.1.2 Gauge invariant theories

The SM is a quantum field theory based on the symmetry groups $SU(3)_C \times SU(2)_L \times U(1)_Y$, being C the color quantum number, L the chirality and Y the hypercharge. Actually electroweak (EW) interactions are described by $SU(2) \times U(1)$ gauge symmetry, while quantum chromodynamics is based on $SU(3)_C$ gauge symmetry.

Gauge symmetry

Gauge symmetry description [8] mostly relies on the work of Yang and Mills, who generalized electromagnetism to electroweak theory. The starting point is the lagrangian of a free particle (Dirac lagrangian), associated to the field ψ

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi \tag{1.1}$$

Dirac equation turns out from Equation 1.1 applying Eulero equations:

$$i\partial_\mu \bar{\psi}\gamma^\mu + m\bar{\psi} = 0 \tag{1.2}$$

It's easy to show that the Lagrangian (Equation 1.1) is invariant under the global gauge transformation

$$\psi \to e^{i\theta}\psi \tag{1.3}$$

If one postulate that every Lagrangian should be symmetric also under the local gauge transformation

$$\psi \to e^{i\theta(x)}\psi \tag{1.4}$$

some workaround is required to keep Expression 1.1 invariant. For U(1) gauge group, one can introduce a new vector field A_μ: when ψ undergoes the transformation reported in Equation 1.4, A_μ transforms accordingly:

$$A_\mu \to \partial_\mu - igA_\mu \tag{1.5}$$

A new local-gauge invariant Lagrangian is then obtained:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{1.6}$$

being $F_{\mu\nu}$ the electromagnetic field tensor

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{1.7}$$

Generalizing Expression 1.5 to higher dimension gauge groups, the covariant derivative can be written:

$$D_\mu = \partial_\mu - igT^a A^a_\mu \tag{1.8}$$

T^a are the generators of the transformation group and g is a coupling constant. The field strength tensor reads

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + gf^{abc}A^b_\mu A^c_\nu \tag{1.9}$$

where f^{abc} are the structure constants of the symmetry group.
In the Yang-Mills approach indeed the reasoning is extended to include weak interaction through the $SU(2)_L$ non-abelian gauge group. In this case the new covariant derivative then reads:

$$D_\mu = (\partial_\mu - ig_1 A^a_\mu \tau^a - \frac{i}{2}g_2 Y \mathbf{B}_\mu) \tag{1.10}$$

being g_i the coupling constants to the corresponding $U(1)_Y$ and $SU(2)_L$ gauge group, τ and Y are the generators of transformations (where τ corresponds to $\hat{\sigma}/2$, $\hat{\sigma}$ being Pauli matrices), and A_μ and $\mathbf{B}_\mu$ are the new gauge boson fields for $U(1)$ and $SU(2)$ group respectively.
In fact to extend the theory to the QCD gauge group $SU(3)_C$ a further term

$$D'_\mu = -ig_s\frac{\lambda}{2}\mathbf{G}_\mu \tag{1.11}$$

must be be added to the covariant derivative, accounting for the $SU(3)_C$ gauge bosons, the gluons.
In this picture the conserved charge, associated to the gauge boson A_μ through $U(1)_Y$ symmetry, is called the hypercharge Y; the $SU(2)_L$ gauge group, represented in the Equation 1.10 by B^a_μ boson (a ranging between 1 and 3), conserves the weak isospin T; color quantum numbers represents the "charge" related to $SU(3)_C$ symmetry, denoted by the gauge boson G^a_μ (a ranging between 1 and 8).

Higgs mechanism

A further trick [5] is needed to account for massive, short-range weak bosons ($W^\pm$ and Z^0) discovered in 1982 by UA1 and UA2 experiments at CERN. This phenomenon, called *spontaneous electroweak symmetry breaking*, can be described introducing a new $SU(2)$ doublet of scalars through a complex "Higgs" field ϕ

$$< \phi >= \frac{1}{\sqrt{2}}\begin{pmatrix} \phi_1 - i\phi_2 \\ \phi_3 - i\phi_4 \end{pmatrix} \tag{1.12}$$

The gauge invariant Lagrangian for this field reads

$$L = (D^\mu\phi)^\dagger(D_\mu\phi) - V(\phi) = (D^\mu\phi)^\dagger(D_\mu\phi) - \mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2 \tag{1.13}$$

A quadratic potential has been introduced with mass parameter $\mu^2 > 0$ and self-coupling constant λ. In fact this choice assures the spontaneous breaking of $U(1)$ symmetry, V(ϕ) having a non zero minimum value:

$$\phi^\dagger\phi = -\frac{\mu^2}{2\lambda} = \frac{v^2}{2} \tag{1.14}$$

This "Mexican hat" potential $V(\phi)$ can be chosen in such a way that:

$$< \phi >= \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ \phi^0 \end{pmatrix} \tag{1.15}$$

where [1]

$$\phi^0 = \frac{v}{\sqrt{2}} \tag{1.16}$$

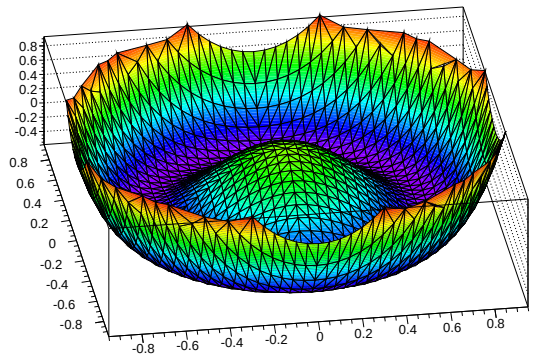


Figure 1.1: Scalar Higgs potential from Equation 1.13 as a function of components of scalar fields.

being

$$M_H^2 = 2\lambda v^2 = -2\mu^2 \tag{1.18}$$

the new Higgs boson mass. The following terms appear in the Lagrangian evaluated on the ground state:

$$\Delta\mathcal{L} = \frac{1}{2}\frac{v^2}{2}[g_1^2(a_\mu^1)^2 + g_1^2(A_\mu^2)^2 + (-g_1 A_m u^3 + g_2 B_\mu)^2] \tag{1.19}$$

When matrix products are evaluated explicitly, expression 1.19 three new gauge bosons can be identified:

$$W^\pm = \frac{1}{\sqrt{2}}(A_\mu^1 \mp iA_\mu^2), \qquad m_W = q\frac{v}{2} \tag{1.20}$$

$$Z_\mu^0 = \frac{1}{g_1^2 + g_2^2}(g_1 A_\mu^3 - g_2 B_\mu), \qquad m_Z = \sqrt{g_1^2 + g_2^2}\frac{v}{2} \tag{1.21}$$

$$A_\mu = \frac{1}{g_1^2 + g_2^2}(g_1 A_\mu^3 + g_2 B_\mu), \qquad m_A = 0 \tag{1.22}$$

Using such a definition the new entities $W^\pm$ and Z get a mass, while the photon field remains massless.
From Equation 1.20 it can be shown also that the Gell Man-Nishijima formula holds for the three generators of the groups:

$$Q = T_3 + \frac{Y}{2} \tag{1.23}$$

[1]The value of parameter v can be determined using the Fermi coupling constant G_F, measured from muon decay, from the following expression:

$$v = \frac{1}{2^{-1/4}G_F^{-1/2}} \simeq 246 \quad \text{GeV} \tag{1.17}$$

where T_3 is the third component of weak isospin. Equation 1.20 can be simplified if one considers:

$$\begin{pmatrix} Z^0 \\ A \end{pmatrix} = \begin{bmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{bmatrix} \begin{pmatrix} A^3 \\ B \end{pmatrix} \tag{1.24}$$

where θ_W (called *electroweak mixing angle* or *Weinberg angle*) can be expressed as a function of coupling constants g_1 and g_2

$$\cos\theta_W = \frac{g_1}{\sqrt{g_1+g_2}}, \qquad \sin\theta_W = \frac{g_2}{\sqrt{g_1+g_2}}, \qquad \tan\theta_W = \frac{g_2}{g_1} \tag{1.25}$$

The W mass can be now expressed as function of Z mass and Weinberg angle:

$$m_W = m_Z \cos(\theta_W) \tag{1.26}$$

Remarkably, experimental values of Z^0 and $W^{\pm}$ mass are in excellent agreement with SM expectation being [9]:

$$\begin{aligned} m_W &= 80.385 \pm 0.015 \quad \mathrm{GeV/c^2} \\ m_Z &= 91.18765 \pm 0.0021 \quad \mathrm{GeV/c^2} \end{aligned} \tag{1.27}$$

On 2012, July the 4th the discovery of a new boson, tentatively identified with a neutral scalar boson (spin and charge equal to zero), was announced by CMS and ATLAS experiments. The boson mass, measured independently by the two experiment, is

$$\begin{aligned} M_H^{ATLAS} &= 125.3 \pm 0.4(\mathrm{stat}) \pm 0.5(\mathrm{sys}) \quad \mathrm{GeV/c^2}[10] \\ M_H^{CMS} &= 126 \pm 0.4(\mathrm{stat}) \pm 0.4(\mathrm{sys}) \quad \mathrm{GeV/c^2}[11] \end{aligned} \tag{1.28}$$

After few months, in March 2013 this boson was temporarily confirmed to be the Higgs Boson, even if further tests and measurements are expected for the next few years.

Fermion masses

Experimentally weak interactions are known to violate parity: actually W boson do not interact with right-handed electrons. Left-handed particles form a doublets in $SU(2)$ group, while right-handed particles form a singlet, which is invariant under $SU(2)$ transformation. This makes the right-handed particles not interacting with gauge boson B^a_μ.
Quarks and fermions masses can be included assuming that the scalar field ϕ acquires a vacuum expectation value. A further term (named *Yukawa term*) can be added to lagrangian density for this aim. For example, in the electron case it reads:

$$\Delta L_{Yukawa} = -\lambda_e \bar{E}_L \phi e_r \tag{1.29}$$

where hermitian conjugate terms have been omitted, λ_e represents the coupling parameter, e_R (e_L) the right-handed (left-handed) electron field, and $\bar{E}_L$ ($\bar{E}_R$) the left-handed (right handed) SU(2) doublet $(\nu_e, e^-)_L$ ($(\nu_e, e^-)_R$). Now electron mass can be written as

$$m_e = \frac{1}{\sqrt{2}}\lambda_e v \tag{1.30}$$

From this expression, which has the same structure for all the fermions, mass turns out to be proportional to the vacuum expectation value through the coupling parameter λ_e which accounts for the difference between different fermion masses.

1.1.3 The top quark

Top Quark has been discovered in 1995 at the Tevatron collider (Fermilab) by CDF and D$\emptyset$, analyzing $p\bar{p}$ collisions data. In fact many of the precision measurements done at Tevatron were constrained by statistical uncertainties, because of the relatively small size of the collected dataset [12].
LHC has been designed as a real "Top quark factory", being the top production cross section at $\sqrt{s} = 8$ TeV about two order of magnitude larger with respect to the Tevatron one. That's why LHC is an ideal tool to study top Physics, producing tens of millions of top quark per year.
Top is the heaviest known quark, with a mass of 173.07 $\pm$ 0.52 (stat) $\pm$ 0.72 (sys) GeV/c^2 [9]. In fact a very good knowledge of the top mass m_t is mandatory to explore all its SM implications: for instance it plays an important role in the B and K Physics, since many observables depends on $(m_t/m_W)^2$ ratio, being m_W the W boson mass; moreover even a precise determination of m_W depends on its quantum corrections, proportional to m_t^2 and $\ln(M_H)$. Hence, the mass makes top quark a unique probe to test (Beyond) Standard Model Physics, but it is not its only feature.
Top quark decays weakly, but it's lifetime is one order of magnitude shorter ($\tau \simeq 3 \times 10^{-25} s$) than the characteristic hadronization life of colored particles ($10^{-26} s$). This makes the top quark decay products originating $\mathcal{O}(10^{-16})$ m far from decay vertex, with a good angular separation and mostly distributed in the central region (low η values).
SM predicts a branching fraction $\mathcal{B}(t \to bW) > 0.998$ and through measuring $\Gamma(t \to Wb)$ experiments at both LHC and Tevatron [13] found $|V_{tb}|$ consistent with 1 with a precision in the order of 10%. More precisely

$$\frac{\mathcal{B}(t \to bW)}{\mathcal{B}(t \to qW)} = \frac{|V_{tb}|^2}{|V_{tb}|^2 + |V_{ts}|^2 + |V_{td}|^2} \tag{1.31}$$

turns out to be 0.91 ± 0.04 [9]. Due to the $\mathcal{B}$ value close to one, the $t \to bW$ decay is often assumed to be 100%.

Top quark decay width strictly depends on the CKM matrix element V_{tb}; neglecting b quark mass, the bW decay channel partial width reads

$$\Gamma(t \to Wb) = \frac{\hbar}{\tau} \frac{G_F}{8\pi\sqrt{2}} m_T^3 |V_{tb}|^2 \left(1 - 3\frac{M_W^4}{m_t^4} + 2\frac{M_W}{m_t^6}\right) \tag{1.32}$$

resulting in 1.56 GeV at Leading Order (LO) and 1.45 GeV at Next to Leading Order (NLO).
Another interesting feature of top quark is given by its Yukawa coupling, which relates the matter content of SM to the Higgs sector. The lagrangian term related to the top quark reads

$$\mathcal{L} = Y_t \bar{t}_L \phi t_R + h.c. \tag{1.33}$$

When the Higgs field ϕ acquires its vacuum expectation value (i.e. $\phi \to \frac{1}{\sqrt{2}(v+H)}$) the Equation 1.33 represents the interaction between a top quark and the Higgs boson. Since one finds

$$m_t = \frac{Y_t v}{\sqrt{2}} \qquad v = 246 \quad \text{GeV}, \qquad m_t = 173.07 \quad \text{GeV} \tag{1.34}$$

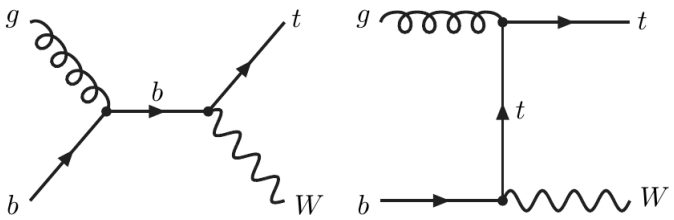
Y_t results to be exactly 1. Many people believe that Y_t value is not accidental and many speculations rose about the possibility that new Physics beyond the SM could be accessed studying top properties.

Top quark production and final states

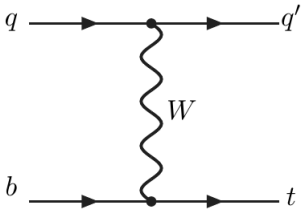
Top quark can be produced via $q\bar{q}$ interactions or gluon fusion according to the diagrams in Figure 1.2. The production cross section in Born approximation reads:

$$\sigma(s, m_T^2) = \sum_{i,j} \int_0^1 dx_1 \int_0^1 dx_2 f_i\left(x_i, \mu_f^2\right) f_j\left(x_j, \mu_f^2\right) \sigma_{ij}^2\left(\hat{s}, m_t, \alpha_s\left(\mu_s^2\right)\right) \tag{1.35}$$

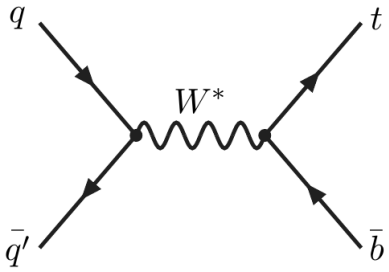
where i, j represents all the possible combination of ingoing quarks; $f(x_i, \mu_f)$ are the PDF evaluated at μ_f, x_i, the former matching the problem scale, the latter being the momentum fraction carried by protons; σ is the cross section, that depends on the quark mass m_T and on the strong coupling constant α_s, evaluated at μ_r. For $t\bar{t}$ production one can set $\mu_f = \mu_r = m_t$, even though the scales are not necessary the same.
In fact at p-p colliders like LHC, production cross section is dominated by systematic uncertainties coming from the uncertainty on the total momentum carried by colliding partons through the parton distribution function (PDF) (actually gluons fusion $gg \to t\bar{t}$ contribution represents about 90% (86.2%) of the total cross section at 14 TeV (8 TeV) [14]).
The $t\bar{t}$ final states depend on the W boson decay modes, and four channels are



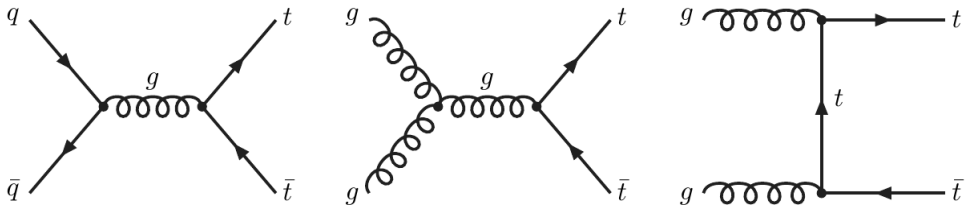
(a) Top production associated with W boson.



(b) t-channel.



(c) s-channel.



(d) $t\bar{t}$ production

Figure 1.2: Non exhaustive review of leading order top quark production modes, in Single top mode (Figure 1.2a, 1.2b, 1.2c) and $t\bar{t}$ pairs (Figure 1.2d). [12]

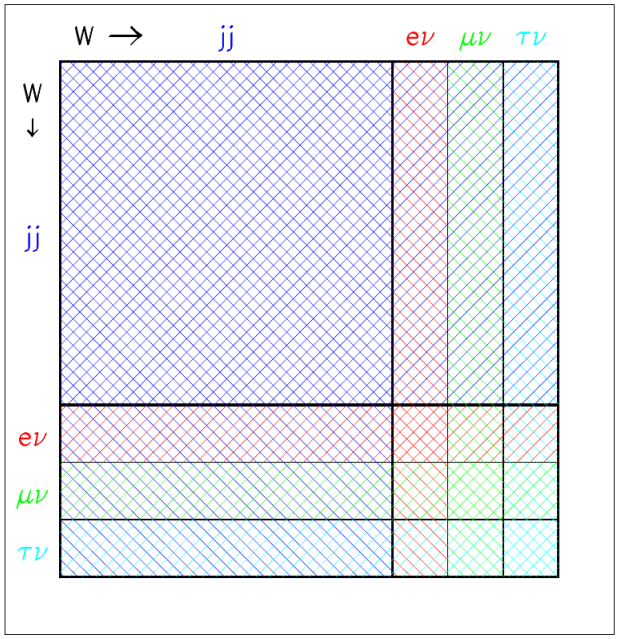


Figure 1.3: Top quark decay modes are shown, where: $BR(t\bar{t} \to b\bar{b}q\bar{q}q'\bar{q'}) = 44\%$, $BR(t\bar{t} \to b\bar{b}qql\nu_l) = 30\%$ and $BR(t\bar{t} \to b\bar{b}l'\nu'_l l\nu_l) = 5\%$. [12]

usually identified: dileptonic, when both W bosons decay in one lepton and a neutrino (BR $\sim$ 5%); single lepton, when one W boson decays in two quarks, while the other decays leptonically (BR $\sim$ 30%); fully hadronic, when both W bosons decays hadronically (BR $\sim$ 44%); in a SM framework usually one consider the complement to 1 given by top decaying in τ lepton (BR $\sim$ 21%), neglecting other possible decay modes. A pictorial view of top decay branching fraction is given in Figure 1.3. At LHC top quark can also be produced in Single top channel, through the following production modes:

- **s-channel**: $qq \to t\bar{b}$ the rarest production process, (approx. NNLO cross section: 5.55 pb$^-$1 at $\sqrt{8}$ TeV)
- **t-channel**: $gq \to t\bar{b}$, which has bigger PDF uncertainties due to the presence of gluons, but it has a larger cross section with respect to s-channel (approx. NNLO cross section: 87.1 pb$^-$1 at $\sqrt{8}$ TeV).
- **tW**: $gb \to t\bar{b}$ it starts with one gluon and one b-quark, and it represents the 20% of the total cross section in single top production. (approx. NNLO cross section: 22.2 pb$^-$1 at $\sqrt{8}$ TeV)

Flavor Changing Neutral Currents (FCNC) are strongly suppressed by GIM mechanism, hence are often considered possible signatures for new Physics Beyond the Standard Model. However, no FCNC have been observed so far. Among the most sensitive measurements performed so far, ATLAS experiment

(2.05 fb^{-1} at $\sqrt{s} = 7$), set a limit on two-quark decays [15]

$$\begin{aligned} \mathcal{B}(t \to ug) &< 5.7 \times 10^{-5} \\ \mathcal{B}(t \to cg) &< 2.7 \times 10^{-4} \end{aligned} \tag{1.36}$$

On the other hand CMS set a branching ratio upper limit of 0.21% for the direct search of FCNC $\mathcal{B}(t \to Zb)$, using a statistics of 4.6 fb^{-1} at $\sqrt{s} = 7$ TeV [16].

Sensitivity to new physics

As already mentioned one of the favorite channels for the discovery of new physics beyond the SM (BSM) relates to the top quark. A number of theories predicts the existence of particles that interact with third generation quarks, and could bring some changes in the top production cross section or angular distribution.
The $t\bar{t}$ invariant mass spectrum is one of the most sensitive monitors of new $t\bar{t}$ production processes [12]: SUSY [17] and models of "composite quarks" [19][20] are expected to change the event yield at high $m_{t\bar{t}}$; in case of existence of new bosons (like Z' [22]), further resonances are expected; finally the $t\bar{t}$ invariant mass spectrum analysis can also be used to probe the existence of extra dimension or ultra heavy intermediate states [21].
The SM also predicts that the decay branching ratio $\mathcal{B}(t \to bW)$ is greater than 0.998, but BSM theories often predict the existence of new decays. Theories with an extended Higgs sector predicts $t \to H^{\pm}b$ [23]; new $EWSB$ mechanisms, like topcolor assisted Technicolor [22] and SUSY are just some of the theories that could introduce alternative top quark decays [12].

1.2 Baryon number violation

Before introducing the model which has inspired this work of thesis, a brief historical review on the baryon number debate will be presented (Section 1.2.1). A detailed description of the theoretical effective approach which underlies baryon number violating top decays will be presented in Section 1.2.3.

1.2.1 Historical remarks

A "surrogate" of baryon number was first proposed as the fundamental ingredient of a conservation principle in 1929 for the electron and proton by Herman Weyl [24], when proton was not discovered yet [25]. In 1938 Stuckelberg [4] proposed the conservation of the "Heavy Charge", starting from the observation that no heavy particles could change into light ones. However the very first definition of "Baryons" for nucleons and Hyperons appeared in 1953 with Pais [4] and the same year Marx, Zeldovich Konopinski and Mahnoud proposed the conservation of lepton number, even though Zeldovich only considered the

violation of lepton number through the $\beta\beta$ decay. Indeed a large literature covered this topic in the following years. It's worth citing the work by Goldhaber, Cowan and Reines [26] that put for the first time a lower limit on the mean lifetime of unbound proton ($\tau \simeq 10^{20}$ years), giving the first lower constraint on future experiments. Interestingly, they stated that even if one was not able to "conceive an experiment which would prove the absolute stability of nucleons, judging from the demonstrated practical stability of nucleons", such a conservation law could be used "with considerable confidence in discussions of *practically observable* nuclear reactions".
In 1959 Pontecorvo theorized violation of lepton flavor through neutrino oscillations, while in 1967 Kuzmin and Sakharov considered the baryon number violation as a mandatory ingredient of baryon-antibaryon asymmetry in the universe.
Indeed with the advent of Grand Unified Theories and unification of electromagnetic, weak and strong interactions, the belief that the Baryon and Lepton number could be violated became stronger and stronger.

1.2.2 Baryon number violation

Baryon number (B) is defined as

$$B = \sum_q (N_q B_q - N_{\bar{q}} B_{\bar{q}}) = \sum_q \frac{1}{3}(N_q - N_{\bar{q}}) \tag{1.37}$$

where q runs on flavors, $N_q(N_{\bar{q}})$ is the number of quarks (antiquarks), and $B_{\bar{q}} = B_q = 1/3$. Obviously the baryon number of proton (neutron), composed by uud (ddu) quarks combination reads

$$B_p = B_{[uud]} = B_q \times 3 = 1 \qquad B_n = B_{[ddu]} = 1 \tag{1.38}$$

The lepton number (L) is then defined as

$$L = \sum_l L(l, \nu_l) = \sum_l (N_l + N_{\nu l} + N_{\bar{l}} + N_{\nu\bar{l}}) \tag{1.39}$$

In the SM both of these numbers are accidentally conserved, even if this effect is founded on experimental results, more than theoretical ground.
Actually while electric charge conservation has its strong theoretical support in $U(1)_{em}$ gauge local symmetry and remains alive even in the electroweak group $G_{ew} = SU(2)_L \times U(1)_Y$, B/L conservation comes from a global (not local) symmetry.
In principle, if one would agree with the "axiom" which states that "just local gauge symmetries are exact in nature", Baryon and Lepton number conservation laws are exposed to possible violations. On the other hand if B/L conservation came from a local symmetry some difference between gravitational

and inertial mass should be observed even in the Eotvos experiment (except for very tiny coupling constants)[28].
However, there is a number of hints that could suggest the non conservation of Baryon number:

- The baryon-antibaryon asymmetry in the universe [27]
- An extension of the SM can incorporate the non conservation of baryon number asymmetry without loosing the renormalizability of the theory
- The non conservation of Baryon/Lepton Number come from (very small) SM non perturbative effects, suppressed by a factor $e^{-2\pi/\alpha_w} \simeq 10^{-86}$ [29]
- Baryon and Lepton Number Violation are included in many existing BSM theories: in SUSY, by R-parity violating interactions and dimension 5 operators ([30]); in Grand Unified Theories ([31]); in Black Holes Physics [32].

Baryon and lepton number conservation and R-Parity

In SUSY the new additive quantum number R is defined [33]:

$$R = (-1)^{2S}(-1)^{3B+L} \begin{cases} +1 & \text{for ordinary particles (even parity)} \\ -1 & \text{for superpartners (odd parity)} \end{cases} \tag{1.40}$$

being S the particle's spin, B and L baryon and lepton numbers.
R is associated with a Z_2 subgroup of the group of $U(1)$ R-Symmetry transformations, acting on the gauge superfields and on the two chiral doublets Higgs superfields H_d and H_u, responsible for electroweak symmetry breaking.
It's evident that R-parity is closely related with conservation of baryon and lepton numbers: R-parity is conserved even if L and B are separately violated, as long as $B - L$ remains conserved [33].
Conservation of R-Parity grants that $spin - 0$ squark and slepton cannot be exchanged between ordinary quark and leptons; it ensures that R-odd sparticles can be produced only in pairs and that the decay of an R-odd sparticle should lead to another one (or an odd number of them); R-Parity causes also the stability of "Lightest Supersymmetric Particles", like neutralinos (one of the best dark matter candidates).
On the other hand a violation of R-Parity (necessarily accompanied by a violation of B and/or L) may results in a proton decay with a very short lifetime (unless R-Parity violating interactions were sufficiently small). Yet it may also provide a source of Majorana masses for neutrinos, and it leads to decay of superpartners into ordinary particles.
In fact remarkable efforts have been conveyed so far looking for possible R-Parity violations.

1.2.3 Effective fields and baryon number violation

The effective approach

There are two ways to look for new physics: one consists of introducing new models, with new sets of particles; otherwise one can just introduce new interactions among known particles [35]. In the latter case, using a model-independent approach, new physics can be searched without involving any particular extension of SM.
Such an approach can be pursued employing an *effective quantum field theory* (EQFT), whose features can be appreciated especially when calculation related to the full theory cannot be easily performed (because of strongly coupled terms or higher order terms for instance).
An *effective quantum fields* satisfies the following requirements:

- Every possible extension of SM meets the requirements of unitary, analyticity etc. of S-matrix
- All the SM symmetries are conserved (namely Lorentz invariance, and $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge invariance)
- SM is restored in the appropriate limit
- The new theory is general enough to describe any physics BSM, but some guidelines must be given about the best conditions to detect effects of new Physics
- Radiative corrections can be re-calculated at any order for the (non-)SM interactions in the effective theory

In the SM all the the operators (the product of the fields) in the Lagrangian are of mass dimension four or less. However other operators of higher dimension can be added. These operators have coefficients proportional to the inverse power of mass: when the mass is much larger than the experimental energy, their contribution becomes negligible and the dominant terms are those with the lowest dimensionality. An operator of dimension five for instance is responsible for generating Majorana neutrino masses: while a number of operator of dimension six exist, only few of them contribute to a given physical process.
The mass scale (i.e. the scale of new physics) can be represented with Λ. It usually acquires much larger values than the experimentally-accessible energies, ranging from few TeV up to the Planck Scale.
The effective lagrangian reads

$$\mathcal{L} = \mathcal{L}_{SM} + \sum_i \frac{c_i}{\Lambda^2}\mathcal{O}_i^2 + \cdots \tag{1.41}$$

being $\mathcal{O}_i$ the dimension-six operators, ellipses stands for higher-order operators and c_i are dimensionless coefficients and parameterize the strength of coupling

between new physics and SM particles.
For instance a typical effective model [35] can be represented by heavy Z' boson. At low energy (less than Z' mass) one cannot observe Z' directly, but could describe the exchange of a W boson as a four-fermion interaction (something very similar to Fermi theory of weak interaction at energies lower than W boson mass). This is an operator having dimension six. In this case one could set $\Lambda = m_{Z'}$. Indeed this interaction is suppressed by the propagator through the inverse power of Z' mass and SM description is restored when $\Lambda \to \text{inf}$.
When one specify the effective operators, all the involved fields in operator description must be completely specified (one can assume they are just the SM particles, eventually with Higgs doublet). Then EQFT can be used to compare specific theory predictions or to make phenomenological predictions on the base of experimental data. When a new particle is revealed, one should include the associated field and revise the theory. If the experimental energy is approximately Λ, contribution of higher order operators becomes not negligible, and one should include them into Lagrangian calculation. Since higher order operators are infinite, in this scenario an effective description is no more useful, and a new theory, including the new particle, should be created.
If the new Physics does not appear, one can just quantify the accuracy of exclusion of that process.

An effective Baryon-number violating model for LHC

In this paragraph the theoretical model which has inspired this work of thesis [7] will be described in detail.
One should first note that the LHC represents a unique opportunity to probe Baryon number violation beyond the TeV scale. Thanks to the high production yield, the very clear signature (through a charged isolated lepton) and the negligible impact of hadronization, top quarks look very clearly identifiable candidates in the search for new physics.
As a first step a new dimension six operator which keeps $SU(3)_C \times SU(2)_L \times U(1)_Y$ invariance is introduced:

$$\mathcal{L} = \sum_i \frac{c_i}{\Lambda^2} \mathcal{O}_i^2 \tag{1.42}$$

where Λ represents the mass scale related to new Physics, c_i the dimensionless effective coefficients and $O^1..O^5$ is the basis of five, independent, dimension-six operators which conserve B-L symmetry [36]:

$$O^{(1)}_{abcd} \equiv (\bar{d^c})^{\alpha}_{a}(u)^{\beta}_{b}(\bar{q^c})^{i\gamma}_{c}(l)^{j}_{d} \quad \epsilon_{\alpha\beta\gamma}\epsilon_{ij} \tag{1.43}$$

$$O^{(2)}_{abcd} \equiv (\bar{q^c})^{i\alpha}_{a}(u)^{j\beta}_{b}(\bar{u^c})^{\gamma}_{c}(l)_{d} \quad \epsilon_{\alpha\beta\gamma}\epsilon_{ij} \tag{1.44}$$

$$O^{(3)}_{abcd} \equiv (\bar{q^c})^{i\alpha}_{a}(u)^{j\beta}_{b}(\bar{q^c})^{k\gamma}_{c}(l)^{l}_{d} \quad \epsilon_{\alpha\beta\gamma}\epsilon_{ij}\epsilon_{kl} \tag{1.45}$$

$$O^{(4)}_{abcd} \equiv (\bar{q^c})^{i\alpha}_{a}(u)^{j\beta}_{b}(\bar{q^c})^{k\gamma}_{c}(l)^{l}_{d} \quad \epsilon_{\alpha\beta\gamma}[\epsilon\tau]_{ij}\cdot[\epsilon\tau]_{kl} \tag{1.46}$$

$$O^{(5)}_{abcd} \equiv (\bar{d^c})^{\alpha}_{a}(u)^{\beta}_{b}(\bar{u^c})^{\gamma}_{c}(e)_{d} \quad \epsilon_{\alpha\beta\gamma}\epsilon_{ij} \tag{1.47}$$

being a, b, c, d the flavors, i, j, k, l the $SU(2)_L$ indices and α, β, γ the colors.

The effective operators in Equation 1.42 does not include any neutrino interactions and it can be expanded on the basis of two operator P_L and P_R, where $P_{L,R} = 1 \pm \gamma^5$:

$$\begin{aligned} O^{(s)} &\equiv \epsilon^{\alpha\beta\gamma}[\bar{t^c_\alpha}(aP_L + bP_R)D_\gamma][\bar{U^c_\beta}(cP_L + dP_R)E] \\ O^{(t)} &\equiv \epsilon^{\alpha\beta\gamma}[\bar{t^c_\alpha}(a'P_L + b'P_R)E][\bar{U^c_\beta}(c'P_L + d'P_R)D_\gamma)] \end{aligned} \tag{1.48}$$

D,U,E stands respectively for down, up quark and charged lepton; all the fermions in Equation 4.1 are taken as eigenstates; Greek indices run over color charges; $a, a', ...$ are fermion-flavor dependent, dimensionless parameters. In this base effective lagrangian reads:

$$\mathcal{L}_{eff} = \mathcal{L}_{\mathcal{SM}} + \frac{1}{\Lambda^2}\big(O^{(s)} + O^{(t)} + h.c.\big) \tag{1.49}$$

If the Λ scale refers to the mass of an heavy mediator exchanged in s or in t channel, coefficients $a, a', ..$ can be safely set to a value in the order of the unity. Heavy gauge mediators give rise to $O^{(1,2)}$, with $a = d = 0$ and $b = c = 0$. When two top quarks are involved, one can substitute t with U in Equation 4.1, even though the two operators are no longer independent. Such operators have a role at LHC in processes like $gd \to \bar{t}te^+$, or $e^-d \to t\bar{t}$ in (future) e-,p colliders.

Top decay with one lepton in final state without any neutrinos causes a change in lepton number $\Delta L = 1$. However, due to the angular momentum conservation, the following statement must always be true [7]

$$\Delta(L + 3B) \in 2\mathbb{Z} \tag{1.50}$$

Therefore from the non-conservation of lepton number, the violation of baryon number conservation law follows.

Two LHC processes can be described by the model presented in

$$\begin{aligned} t &\to \bar{u}\bar{d}e^+ \\ ud &\to \bar{t}e^+ \end{aligned} \tag{1.51}$$

In both cases baryon number is not conserved. The most interesting feature of such processes is the presence of a charged lepton without neutrino; on the

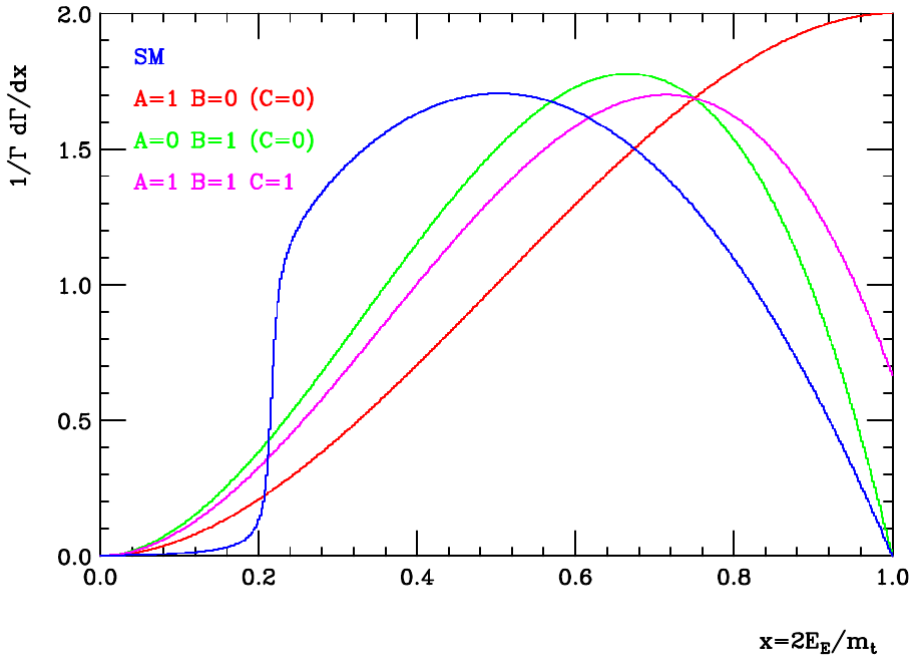


Figure 1.4: Energy spectrum of charged lepton in the SM decay: $t \to bE^+\nu_E$ and in BNV decay $t \to \bar{U}DE$ for different values of A, B, C constants.[7]

other hand the presence of a top quark makes the process clearly identifiable through the tagging of a c- or b- quark. Neglecting all the fermion masses except the Top's one, the squared amplitude of this process can be written:

$$\begin{aligned}\sum_{spin,colors} |M|^2 = \frac{24}{\Lambda^4}[&(p_T \cdot p_D)(p_U \cdot p_E)(A+C) - \\ &- (p_T \cdot p_U)(p_D \cdot p_E)C + \\ &+ (p_T \cdot p_E)(p_D \cdot p_U)(B+C)]\end{aligned} \tag{1.52}$$

The following coefficients

$$\begin{aligned}A &\equiv (|a|^2 + |b|^2)(|c|^2 + |d|^2) \\ B &\equiv (|a'|^2 + |b'|^2)(|c'|^2 + |d'|^2) \\ C &\equiv \Re(a^* c^* a' c' + b^* d^* b' d')\end{aligned} \tag{1.53}$$

come from the square of $O^{(s)}$ and $O^{(t)}$ and their interference.
Hence decay amplitude of BNV process reads

$$\Gamma_t^{BNV} = \int_0^{m_t/2} \mathrm{d}E_E \frac{m_t^2 E_E^2}{32\pi^3\Lambda^4}\left[\left(\frac{A}{3} + B + C\right)\left(1 - \frac{2E_E}{m_t}\right) + \frac{A}{6}\right] \tag{1.54}$$

$$= \frac{m_t^5}{192\pi^3}\frac{1}{16\Lambda^4}[A + B + C] \tag{1.55}$$

being E_E the lepton energy in the top rest frame.
If one assume a top quark width of 1.4 GeV, the BNV branching ratio now results

$$\begin{aligned}\mathcal{B}_t^{BNV} &= 1.2\times10^{-6}\left(\frac{m_t}{173\text{GeV}}\right)^5\left(\frac{1\text{TeV}}{\Lambda^4}[A+B+C]\right)\\ &\simeq 1.2\times10^{-6}\frac{1\text{TeV}}{\Lambda^4}[A+B+C]\end{aligned} \tag{1.56}$$

Assuming a $t\bar{t}$ production cross section of 150 (950) pb, at $\sqrt{s} = 7(14)$ TeV we can expect 0.35 (2.2)/fb^{-1} BNV top decays when A+B+C = 1 for each combination.
On the other hand, the partonic cross section for BNV production reads:

$$\hat{\sigma}_t^{\text{BNV}} = \frac{1}{96\pi\Lambda^4}\int_{m_t^2-\hat{s}}^{0}\mathrm{d}\hat{t}\left[A\frac{\hat{t}(\hat{t}-m_t^2)}{\hat{s}^2}+B\frac{(\hat{s}-m_t^2)}{\hat{s}}+2C\frac{\hat{t}}{\hat{s}}\right] \tag{1.57}$$

$$= \frac{\hat{s}}{96\pi\Lambda^4}\left(1-\frac{m_t^2}{\hat{s}}\right)^2\left[\left(\frac{A}{3}+B+C\right)+\frac{m_t^2}{\hat{s}}\frac{A}{6}\right] \tag{1.58}$$

being the Mandelstam variables $\hat{s} \equiv (p_U + p_D)^2$ and $\hat{t} \equiv (p_U - p_E)^2$. As expected the cross section induced by six-dimension operator grows as $\hat{s}/\Lambda^4$ (where $\hat{s} \ll \Lambda$ to keep the effective model valid).
Out of the six possible initial flavor assignments in production channel, three process can be considered relevant for analysis, namely

1. $ud \to \bar{t}E^+$
2. $ub \to \bar{t}e^+$
3. $cb \to \bar{t}\mu^+$

being the first the most PDF-favored, the second the (possibly) flavor-unsuppressed, while the third the most PDF-suppressed yet (possibly) flavor unsuppressed.

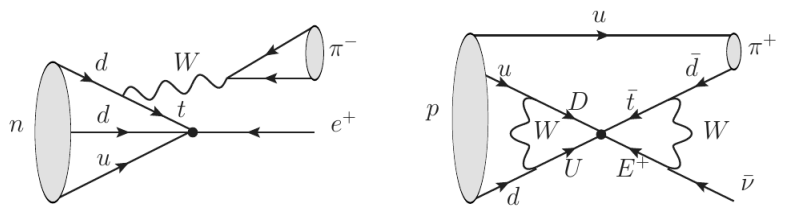


Figure 1.5: Representative tree-level and two-loop level diagrams involving BNV operators and leading, in principle, to nucleon decay [7].

Even if the operator in Equation 1.42 contributes to tree and/or loop diagrams of nucleon decay, an higher limit on Λ is given by the production process

in Figure 1.5 (Left) when leptons are e, μ (the constraint is relaxed when lepton it's τ). When second and third quark generations are involved, three emissions of W bosons are required, and the process rate is suppressed at a level consistent with data. This considerations tend to favor PDF-suppressed processes. In fact if one consider a single operator contribution at time with fixed flavors as reported in Figure 1.5 (Right) small upper bounds on the effective parameters are obtained. However, when summing over all the possible UDUE virtual contribution, due to some cancellation effective parameters can also be large and this kind of mechanism worth a separate discussion.

Three generations B and L violation and future perspectives

In fact as a future perspective beyond the model that has just been described, in the following section a more recent study on baryon number violation in top quark channel will be presented.
Actually in the previous treatment a minimal effective theory without flavors requirements has been presented. In fact it can be shown [38][37] that when flavor symmetries present in gauge sector are broken through SM-like flavor mixing, baryon and lepton number are violated at TeV scale and they satisfy at the same time all the constraints that preserve proton stability. Actually the three generation lepton and baryon number violation takes place breaking the flavor symmetry group, already explicitly broken in Yukawa sector (though this breaking is well constrained by experimental results). In such a model, which does not rely on effective theories, a significant suppression of some flavor structures occurs (like those restricted to light generations only, whose primary importance is due to low energy processes) and at high energy resonant processes violating lepton and Baryon number could be seen in processes like $\bar{t}\mu^+e^+$ or $t\bar{t}$+jets at LHC.
Even though an accurate description is not easy to be given in a model independent framework (non local process cannot be modeled by effective theories), the flavor mixing suppression is no longer present. Thus heavy fermions are directly produced and - in principle - transitions involving all the three generations are allowed (in the limit where all Yukawa couplings are diagonal).

If one introduces slight deviations from SM, assuming flavor and Lorentz symmetries and conservation of electric charge, a number of processes could be observed [37].

Interestingly, all the processes are characterized by the presence of same sign fermions (either quark or leptons). Moreover, a dilepton charged asymmetry can be seen. At LHC processes initiated by u, d are strongly favored, and they happen much likely than those involving $\bar{u}\bar{d}$. Transitions like $uc \to \bar{t}e^+\mu^+\bar{\nu}_\tau$ and $dd \to \overline{t}\overline{t}\bar{s}\bar{s}$ give higher positively charged lepton yield and anti-top pairs than negatively charged lepton and top pairs. A charged symmetry in

same sign dilepton BLV production can be defined as

$$A_{ll'} \equiv \frac{\sigma^{BLV}(l^+l'^+) - \sigma^{BLV}(l^-l'^-)}{\sigma^{BLV}(l^+l'^+) + \sigma^{BLV}(l^-l'^-)} \tag{1.59}$$

being $A_{ll'} > 0$ for $(\Delta B, \Delta L) = (\pm 1, \pm 3)$ and $A_{ll'} < 0$ for $(\Delta B, \Delta L) = (\pm 2, 0)$. It's remarkable that these BLV processes are among the very few non-SM processes that feature both positive and negative charge asymmetry.
Such signatures may still represents interesting signatures for LHC future searches.

1.2.4 Experimental searches

The motivations supporting the search of Baryon number violation have much increased with time. At the beginning they were just aiming to the confirm some conservation laws; then they became a cosmological need, and a powerful way to test Grand Unification Theories. Since a huge amount of data has been gathered to test Baryon number violation in several research channels so far, just a quick (but not exhaustive) review of some experimental tests will be given in the next paragraphs.

Nucleon stability

One of the most straightforward ways to check the reliability of the $B-$ and $L-$ conservation laws is just to probe nucleon stability: nucleon decay may for instance occur when quarks are close enough to exchange heavy-number-violating particles of mass M_X.
A rough scale for the lifetime τ_p of nucleon decay is based on a simple dimensional analysis being [39]

$$\tau_p \propto \frac{M_X^4}{\alpha^2 m_p^5} \tag{1.60}$$

Even if a detailed model is needed to get an accurate estimate of branching ratio, one may expect to be $p \to e^+\pi^0$ the most likely decay (as long as the transition happens inside one generation only). On the other hands SUSY models favor transition across generations [40], making $p \to \nu K$ decay modes quite likely as well (but more difficult to detect). In addition to GUT, that preserve $B - L$ quantum numbers and allows $\Delta B = 1$, in some models even $\Delta B = 2$ transitions are allowed, which may lead to $n - \bar{n}$ oscillations or $n \to \nu\nu\nu$ decays.
The mainstream to perform this kind of measurements is the "direct counting approach". An instrument filled with several thousands of sensitive volume ($\sim 10^{34}$ nucleons) can be monitored to look for the desired decay and to set a lower limit on nucleon lifetime. Indeed such a detector has been set at Super-Kamiokande experiment[41]. The setup is based on a huge tank (39

m of diameter, 42 m high) filled with 50 kton of pure water and surrounded by about 11100 photomultiplier tubes, designed to separate electromagnetic showers from Cherenkov light. The search for $e^+\pi^0$ consist in detecting electron shower and back-to-back two-ring showers due to $\pi^0 \to \gamma\gamma$.
The most recent measurement gives a lower limit 90% C.L. $\tau_p \geq 1.3 \times 10^{34} y$ for $p \to e^+\pi^0$ and $\tau_p = 1.1 \times 10^{34} y$ for $p \to \mu^+\pi^0$, and no evidence of proton decay.[43]
Other several channel have been investigated, such as $p \to K^+\bar{\nu}$, $p \to \mu^+ K^0$, $n \to e^- K$, [41] [42], that could play a role in Supersymmetric models. However no evidence of nucleon decay has been found so far, and a lower limit on proton lifetime has been been set to an order of $\tau_n \sim 10^{33}$ years.

Mesons, τ and Z decays

At intermediate energies observation of baryon number violation can be approached also studying mesons and τ decay.
In 2011 BaBar Collaboration made public results of searches for $B^0 \to \Lambda_c^+ l^-$, $B^- \to \Lambda l^-$ and $B^- \to \bar{\Lambda} l^-$, being l^- a muon or an electron [44]. This was the first measurements of such decays. The data sample consisted of 490 fb^{-1} (471 $\pm$ 3 $B\bar{B}$ pairs, recorded at $-$(4S) resonance, obtained at PEP-II asymmetry energy e^+e^- storage ring at $\sqrt{s} = 10.58$ GeV/c^2). In fact neither lepton number nor baryon number violation was detected, and 90% CL upper limits were set. The upper limit on branching ratio for $B^0 \to \Lambda_c^+ \mu^-$ was 180×10^{-8}. More details are reported in [44].
A search for B- and L- violation in τ decay was performed repeatedly for different channels. In fact, given the limits set on proton lifetime, the probability to observe such a violation in τ decay channel was vanishingly small [45]. However, CLEO and Belle collaboration had the opportunity to improve existing limits exploiting a much improved sensitivity with respect to the past measurements.
In 2005 Belle Collaboration [46] explored the possibility of $\tau^- \to \bar{\Lambda}\pi^-$ and $\tau^- \to \Lambda\pi^-$ decays, with a data sample of 154 fb^{-1} collected at $\Gamma(4S)$ resonance with Belle detector at KEBK e^+e^- asymmetric energy collider. No signal was found and the upper limit on branching fractions at 90% C.L. was set for the first time at $B(\tau^- \to \bar{\Lambda}\pi^-) < 1.4 \times 10^{-7}$ and $B(\tau^- \to \Lambda\pi^-) < 0.72 \times 10^{-7}$. Some details can be found in [46].
In 1999 CLEO Collaboration published results collected at CLEO II detector, at Cornell Electron Storage Ring (CESAR)[47]. The integrated luminosity of data sample was 4.7 fb^{-1} (4.3×10^6 $\tau^+\tau^-$ events). Five decay modes have been analyzed: $\tau^- \to \bar{p}\gamma$, $\tau^- \to \bar{p}\pi^0$, $\tau^- \to \bar{p}\eta$, $\tau^- \to \bar{p}2\pi^0$ and $\tau^- \to \bar{p}\pi^0\eta$. No evidence were found, even if the existing upper limit on branching fraction for the decays in anti-proton and a photon, or a π^0 or an η meson, was strongly improved. The 90% C.L upper limit set for such decays was in the order of 10×10^{-6} (3.5×10^{-6} for $\tau \to \bar{p}\gamma$). Further details can be found in [47].
In 1999 OPAL Collaboration made public the result of a research at LEP,

aiming to observe a Z^0 decay violating simultaneously Baryon and Lepton number. The responsible (resonant) processes were $e^+e^- \rightarrow Z^0 \rightarrow pe^-/p\mu^-$ (and its charge-conjugate final state). Even in this case no candidates violating baryon and lepton symmetry were found, and an upper limit on Z^0 width was set both in muon and electron channel (4.6 keV for $\Gamma_{Z^0 \rightarrow p\mu^-}$ and 4.4 keV for $Z^0 \rightarrow pe^-$) at 95% CL. That was also the first time that limits on the Z^0 partial width have been obtained for processes violating simultaneously baryon, lepton and fermion number conservation.

The LHC era

With the LHC, the possibility to check for Baryon and Lepton Number violation beyond TeV scale. The first measurements of τ decay at LHC has been reported by LHCb collaboration in [48], where upper limits on $\tau \rightarrow \mu^-\mu^+\mu^-$ lepton flavor violating decay, $\tau^- \rightarrow \bar{p}\mu^+\mu^-$ and $\tau^- \rightarrow p\mu^-\mu^-$ have been set, based on a data sample of 1.0 fb^{-1} at $\sqrt{7}$ TeV. Observed upper limits on branching fraction span over the range of $0.83 - 5.4 \times 10^{-7}$. While branching fraction for the process $\tau \rightarrow \mu^-\mu^+\mu^-$ is compatible with previous BaBar and Belle limits, a first limit on BNV decay was set. An increased sensitivity was expected adding further integrated luminosity. A detailed description of this analysis can be found in [48].
Indeed many analyses for SUSY searches have been performed at LHC, aiming to check the validity of $R-parity$ conservation laws. As an illustrative example, in [34] $R-Parity$ violating process are searched, looking for squarks and gluinos yielding anomalous multi-lepton productions. At CMS Experiment for instance a data sample of 2.1 fb^{-1} at $\sqrt{s} = 7$ TeV was used and R-Parity violation coupling parameters λ_{122}, λ_{123}, λ_{233} and H-RPV were excluded for squark and gluino masses in the first 1 TeV/c^2 range (95% CL). In 2012 [49] the upper limit has been further improved, extending exclusion range up to about 2 TeV (with a statistics of 9.2 fb^{-1} at $\sqrt{8}$ TeV).

Chapter 2

Experimental setup

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [50] is the largest particle accelerator ever built. It is located at CERN and buried near the city of Geneva, at 100 m underground. It has been installed in the 27 km-long tunnel of the Large Electron Positron (LEP) accelerator [51], constructed between 1984 and 1989. The LHC has been designed to be a proton-proton and heavy ions superconducting collider: its main goal is at the same time to explore the existence of new Physics at TeV scale and to probe limits and predictions of the SM, providing a center of mass energy up to 14 TeV.

2.1.1 Collider physics

Collider accelerators offer at least one important advantage with respect to fixed target ones. When two beams of particles with the same energy E and the same mass m collide, the energy provided in the center of mass reads:

$$\sqrt{s} = E_{cm} = 2E \tag{2.1}$$

On the contrary, in fixed target accelerators beams impinge with energy E on massive blocks, made up of target particles having mass m_p. The energy in the center of mass reads:

$$\sqrt{s} = E_{cm} \simeq \sqrt{2Em_p} \tag{2.2}$$

From Equation 2.1 and 2.2 it is evident that for a given beam momentum the colliding configuration allows to maximize the amount of energy in the center of mass, with respect to fixed-target experiments.
The availability of high energy translates into the possibility to produce new particles, provided the energy is at least equal to the particle mass.
At the same time, cross sections are energy dependent quantities: for many interesting processes, such as Higgs production, cross section may increase as a

function of energy (while background cross section remains constant), making more convenient to have higher energy in the center of mass to increase particle production rate.
The LHC is a hadron collider. For proton colliders, the energy loss due to synchrotron radiation is smaller with respect to electron colliders. Actually the energy loss is proportional to $(E/m)^4/R$, where E and m represent respectively particle energy and mass and R the curvature radius.
On the other hand whereas in the electron colliders colliding particles are point-like leptons, in proton-based machines particles have an underlying substructure that makes the initial state difficult to predict. Actually the colliding partons carry a momentum fraction x_1 and x_2, while the effective energy in the center of mass is:

$$\sqrt{s_{eff}} = \sqrt{sx_1x_2} \tag{2.3}$$

Due to the non-uniform distribution of momentum over interacting partons, a wide energy range turns out to be available for the production of new particles. The beam remnants might continue their flight in a parallel direction without being detected: as a consequence a sizeable portion of the event energy becomes undetectable, and just the transverse missing energy can be estimated event-by-event, thanks to the sum-zero balance of linear momentum in transverse plane (being the center of mass at rest in the transverse plane and in the laboratory frame).
Of course the high energy in the center of mass is not the only interesting feature at the LHC, which is also able to provide an unprecedented luminosity (namely the number of crossing partons per time and surface units). Colliding protons are grouped in bunches. The number of *events* produced at the LHC per second $\dot{N}_{evt}$ can be written as:

$$\dot{N}_{evt} = \sigma_{tot} \cdot L \tag{2.4}$$

being L the machine luminosity and σ_{tot} the interaction cross section (dominated by inelastic contributions) between particle pairs which reads:

$$\sigma_{tot} = \sum_{i,j} \int dx_i \int dx_j f_i(x_i, Q^2) f_j(x_j, Q^2) \hat{\sigma}_{ij}(Q^2) \tag{2.5}$$

In Equation 2.5 $f_i(x_i, Q^2)$ ($f_j(x_j, Q^2)$) represents the parton density function (PDF), namely the probability to find a parton i (j) carrying a momentum fraction x_i, at a scale equal to Q^2 (see section 3.2.1 for more details); σ_{ij} is the partonic cross section.
In Figure 2.1 the cross section and event rate are shown as a function of center-of-mass energy.

PDFs are different for gluons and u, d valence quarks, as well as for low-momentum quark-antiquarks pairs, produced (or annihilated) as virtual particles in the sea quarks. Since for high momenta the contribution of sea quarks and gluons increases, $f_i(x_i, Q^2)$ is also functions of exchanged momentum Q^2.

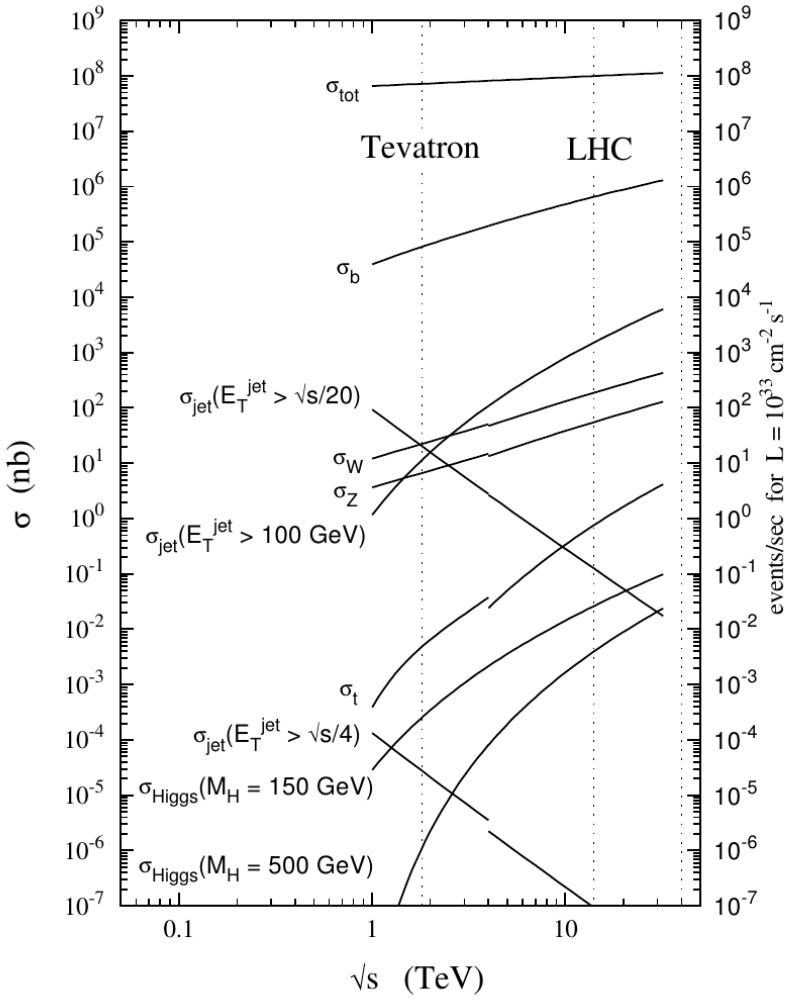


Figure 2.1: Expected cross section and event rate for various process at the LHC as a function of the center of mass energy. [52]

Beam luminosity depends only on the beam parameters, and can be written as a gaussian variable:

$$L = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta^*} \tag{2.6}$$

where N_b is the number of particle per bunch, n_b the number of bunches per beam, f_{rev} the revolution frequency, γ_r the relativistic gamma factor, ϵ_n the normalized transverse beam emittance [54], β^* the beta function [54] at the collision point and F the geometric luminosity reduction factor due to the crossing angle at the interaction point (IP) [50]:

$$F = \left(1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*}\right)^2\right)^{-1/2} \tag{2.7}$$

θ_c is the full crossing angle at the IP, σ_z the RMS bunch length, and σ^* the

transverse RMS beam size at the IP (assuming round beams).
Due to the degradation of both beam intensities and circulating beams emittance, the luminosity in the LHC is not constant over a physics run. The main cause of luminosity reduction is given by collisions at IP, according to the formula [50]

$$L(t) = \frac{L_0}{(1 + t/\tau)^2} \tag{2.8}$$

being τ

$$\tau = \frac{N_{tot,0}}{L\sigma_{tot}k} \tag{2.9}$$

In Equation 2.9 $N_{tot,0}$ represents the initial beam intensity, while L the luminosity and k the number of IPs. As a first approximation, a decay time of about 29 hours is requested to yield a luminosity decay of $1/e$ (assuming an inelastic cross section $\sigma = 10^{25} cm^{-2}$ at 14 TeV, two high luminosity experiments and a peak luminosity of 10^{34} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$. [50]
In fact, there are several other sources of degradation that contribute to decrease beam luminosity (beam losses, slow emittance blow up due to the scattering of particles on residual gas, etc.), reducing the net estimated luminosity lifetime to about 10-15 hours in the aforementioned conditions. [50]

2.1.2 LHC components

The LHC is just the last stage of an accelerating sequence designed to accelerate proton beams, as depicted in Figure 2.2 (in fact, the LHC has been designed to accelerate both protons and heavy ions, like Pb).
First, protons are produced: hydrogen is extracted from a tank and atoms are ionized using a plasmatron. A linear accelerator (LINAC) take over and boost them to an energy of 50 MeV. Proton are then injected into four Proton Synchrotron Boosters (PSB) rings, which squeeze bunches to a length of 90 ns and accelerate them to 1.4 GeV. Two batches, each containing three bunches, are filled into the Proton Synchrotron (PS) reaching an energy of 25 GeV. There bunches are also split in batches composed by 72 bunches, with a 25 ns spacing. Protons are then fed into the Super Proton Synchrotron (SPS), which boost them to an energy of 450 GeV, and then injected in the LHC, where they reach the final desired energy (in the order of the TeV scale) using Radio-Frequency Cavities (RF). The whole procedure can take about one hour. When the beams are completely filled and accelerated, the proton current is 0.582 A, with an energy stored per beam of 362 MJ. In fact to reach design luminosity 2808 proton bunches are required, at intervals of 25 ns.
Protons are kept along their trajectory by 1232 superconducting dipole magnets, using a state-of-art technology: superfluid helium cools down the NbTi cables, as done at Tevatron, DESY and RHIC, to a temperature of 2 K. Thus an 8.4 T field, needed to perform 7 TeV operations, can be maintained. A "twin-core" design allows two pipes share a common cold mass and cryostat,

each one mounting an own core, with magnetic flux circulating in opposite sense through the two channels.

2.1.3 LHC design conditions

At design conditions the LHC will run at a luminosity of $L = 10^{34} cm^{-2} s^{-1}$, with two proton beams having a center of mass energy of about 14 TeV. Four IPs are placed along the 27 km long circumference, in correspondence of six particle detectors: ATLAS [55] and CMS [56] are the two multi-purpose, high-luminosity experiments, built around two proton-proton IPs, where collisions are expected to happen every 25 ns. Both detectors have a wide physics program, aiming to study SM and look for new Physics phenomena, such as Higgs mechanism, Supersymmetry and extra dimensions.
TOTEM [57] and LHCf [58], which respectively share the same IP of CMS and ATLAS, require a lower luminosity, namely $L = 10^{29} cm^{-2} s^{-1}$ for TOTEM and $L = 10^{32} cm^{-2} s^{-1}$ for LHCf. The former provides data respectively for the elastic scattering experiment, the latter is mainly intended to study neutral particles in the very forward region.
In correspondence of a third IP, the LHCb experiment [59] is hosted, mainly to study b-physics and CP violation processes. It is a single-arm spectrometer, with a good forward angular coverage, dedicated to the study of heavy flavor Physics.
Finally, as already mentioned, LHC is able to accelerate also heavy ions at 2.60 TeV per nucleon, providing heavy ion collisions with a total energy in the center of mass $\sqrt{s}$ of 1150 TeV: in correspondence of the fourth IP, ALICE detector [60] is dedicated to probe heavy ion physics at peak luminosity $L = 10^{29} cm^{-2} s^{-1}$.

Parameter	p-p	$^{208}Pb^{82+}$
Center of mass energy (TeV)	14	1148
Number of particles per bunch	1.1×10^{11}	$\sim 8 \times 10^{7}$
Number of bunches	2808	608
Design Luminosity (cm^{-2} s^{-1})	10^{34}	2×10^{27}
Luminosity lifetime (h)	10	4.2
Bunch length (mm)	53	75
Beam radius at IP (μm)	15	15
Time between collisions (ns)	24.95	124.75×10^{3}
Bunch crossing rate (MHz)	40.08	0.008
Circumference	2659	2659
Dipole Field	8.3	8.3

Table 2.1: The LHC design parameters for p-p and Pb-Pb collisions

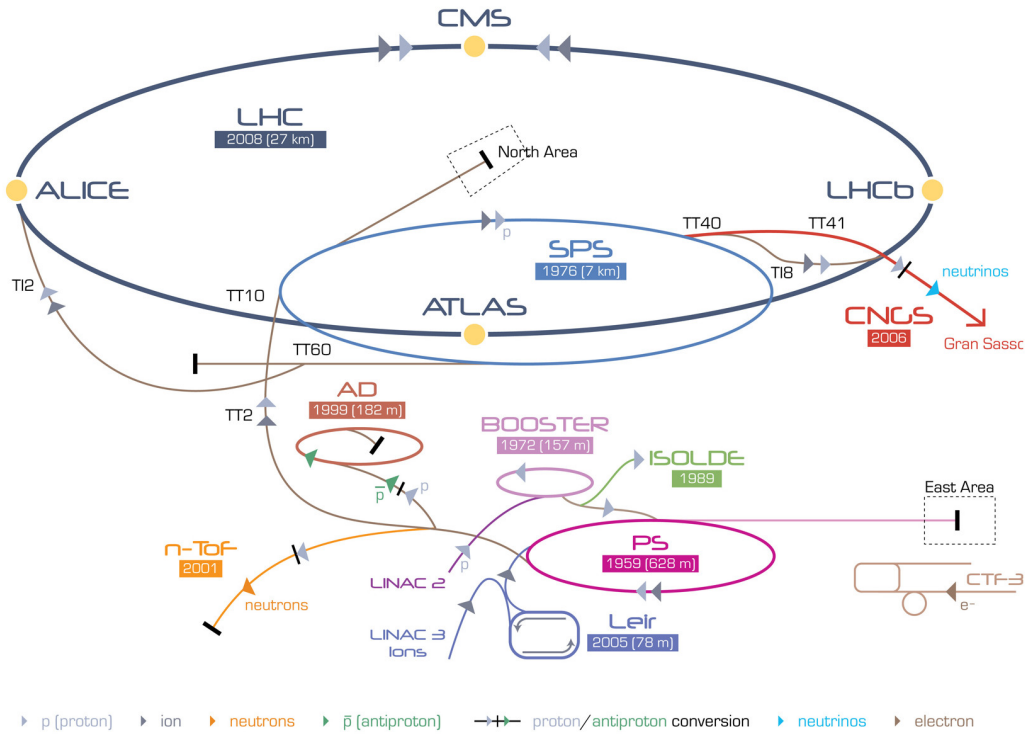


Figure 2.2: Schematic layout of LHC injection system. Protons are first accelerated in the linear accelerator (LINAC) and injected in the Booster and accelerated to 1.4 GeV. Then they are transferred in the Proton Synchrotron (PS), where the beam is split in bunches, reaching an energy of 25 GeV. Protons are then transferred to Super Proton Synchrotron (SPS) which inject them into the LHC at 450 GeV. The LHC may accelerate bunches to the nominal energy (14 TeV). Main LHC experiments are also shown. [53]

2.1.4 Operating conditions

The LHC has been designed to work with a center of mass energy of $\sqrt{s} = 14$ TeV, at a luminosity of $10^{34} cm^{-2} s^{-1}$, but it has never operated at its design condition so far. Actually, after the first beam injection in September 2008, a faulty electrical connection between two of the accelerator dipole magnets caused a mechanical damage which resulted in a helium loss into the tunnel. Several magnets heated up and substantial damages to accelerator infrastructure occurred.
After a stop of more than one year, in the next three years the LHC have run well below its possibilities. [62] The first proton-proton collision took place in March 2010, with $\sqrt{7}$ TeV, namely half of the LHC design energy. During the first year an integrated luminosity of 44.2 pb^{-1} in 2010 was gathered. Actually the first beam ran with a number of low luminosity bunches ranging from 2 to

13, reaching 368 bunches at the end of the year.
During 2011 6.1 fb^{-1} of integrated luminosity have been collected and they were added to further 23.3 fb^{-1}, gathered in 2012. New run conditions, making use of 1380 bunches, with a spacing of 50 ns, were adopted. Due to the Higgs search results, during 2012 beam energy rose again to $\sqrt{8}$ TeV. During 2011 a peak luminosity of $0.4 \times 10^{34} cm^{-2} s^{-1}$ was reached, while in 2012 the record of $0.7 \times 10^{34} cm^{-2} s^{-1}$ was hit. A history plot reporting the cumulative luminosity as function of time can be found in Figure 2.3.
The LHC has now just started the first long shutdown (LS1), which will prepare the machine to start up again in 2015 presumably with an energy of 13 or 14 TeV in the center of mass. After three years of activity, a full check of the collider has been scheduled: several maintenance operations are foreseen, especially aiming to consolidate magnet interconnections, collider injectors and cryogenic system, allowing the LHC to operate at design energy in the center of mass [63]. At the same time, LHC Experiments are undergoing a new upgrade phase to get ready to collect new data at higher energy and luminosity.

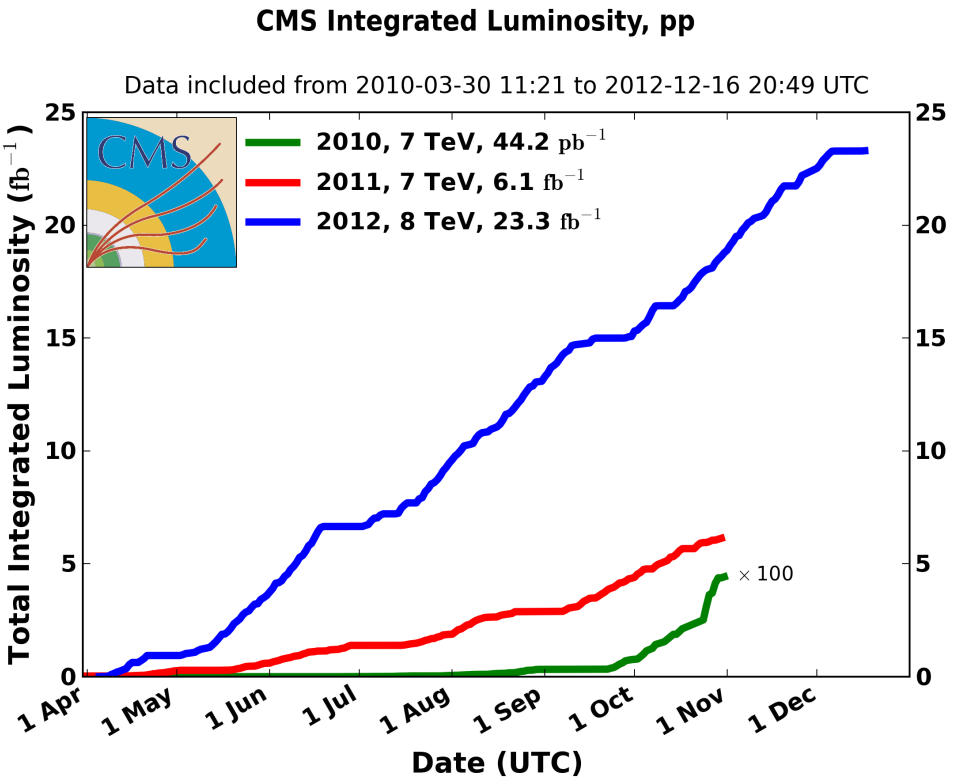


Figure 2.3: Cumulative luminosity versus day delivered to CMS during stable beams and for p-p collisions. This is shown for 2010 (green), 2011 (red) and 2012 (blue) data-taking. [61]

2.2 The Compact Muon Solenoid

The Compact Muon Solenoid (CMS) is one of the two multi-purpose experiments which operate at the LHC. The CMS Collaboration consists in about 3600 scientists, engineers and students from 38 countries. The CMS primary goal was to study the Electroweak Symmetry Breaking mechanism, explained by the existence of Higgs boson. However, apart the discovery of Higgs boson, found in 2012, CMS has been conceived as a tool to check the consistency of the SM beyond TeV scale and, possibly, shed light on the existence of new forces and symmetries, whose implications would be visible at the TeV scale.
Being the inelastic cross section of proton-proton collisions at $\sqrt{s} = 14$ TeV approximately 100 mb, one expects to observe a rate of 10^9 inelastic events per second at $10^{34} cm^{-2} s^{-1}$. A mean of 20 collisions per events are thus expected, with a production of 1000 particles for each event.
Hence severe time and space resolution issues arise: a quick detector response is compulsory (possibly less than 25 ns), and a high spacial granularity is mandatory to identify emitted particles. Unfortunately a high granularity requires also a large number of electronic channels, which must be synchronized in time. Moreover the strong radiative activity at the collision point results in an intense radiation exposure which requires radiation-hard detectors and front end electronics.
CMS implements a solution for all the mentioned issues; in addition the Experiment has been requested to meet all the basic requirements of the LHC Physics program:

- A good muon identification, namely a high momentum resolution and a good charge determination capability (at least for p $<$ 1 TeV), as well as precise dimuon mass determination ($\sim$ 1% at 100 GeV).
- A good charged-particle momentum resolution, reconstruction efficiency through the inner tracker, accompanied by an efficient triggering and offline tagging of $\tau-$ and $b-$ jets
- Good electromagnetic energy resolution using a wide geometric coverage ($|\eta|$<2.5); precise diphoton (dielectron) mass determination ($\approx$ 1% at 100 GeV/c), and π^0 rejection, as well as the capability to efficiently determine photon and lepton isolation
- Good resolution on transverse missing energy and di-jet mass, with large geometric coverage ($|\eta| < 5$) and fine lateral segmentation ($\Delta\eta \times \Delta\phi < 0.1 \times 0.1$)

Hence CMS detector, a 12500-tons heavy detector, takes shape as a cylinder with a diameter of 15 m and 21 m long.
Usually one distinguishes two main detector regions according to their eta coverage: the former is named *barrel*, extended to a maximum pseudorapidity

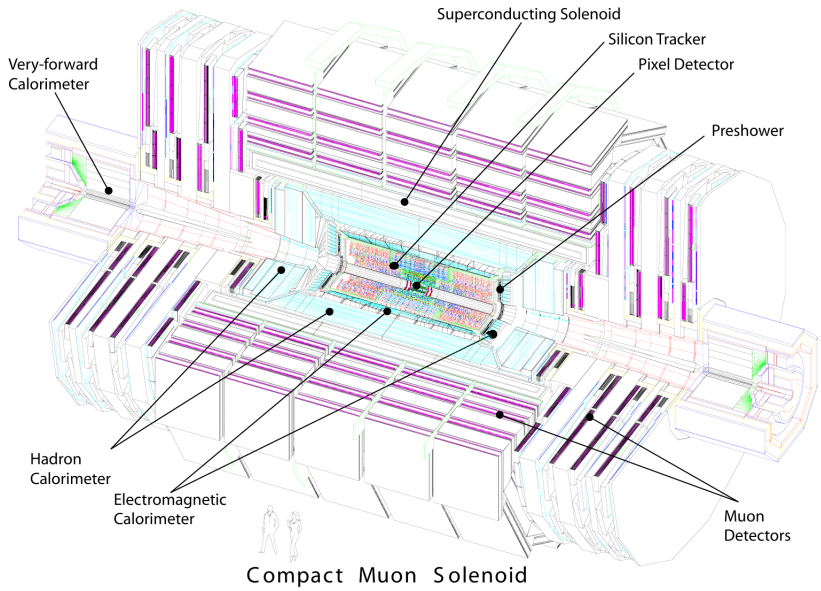


Figure 2.4: Schematic layout of Compact Muon Solenoid detector [56].

$|\eta| < 1.2$ (in the muon system); the latter is called *endcap*, covering higher η regions ($0.9 < |\eta| < 2.4$ for muon chambers).
Particle identification is performed processing the information coming from the detector components:

- The tracker, namely a silicon cylinder, providing the needed granularity and precision, allowing to precisely measure impact parameter (and the position of secondary vertices) of charged particles

- The electromagnetic calorimeter (ECAL), homogeneously made by lead tungstate ($PbWO_4$), read by silicon avalanche photodiodes in the barrel region and by vacuum phototriodes in the endcap region (coverage up to $|\eta| < 3$)

- The hadronic calorimeter, a brass/scintillator sampling hadron calorimeter (coverage up to $|\eta| < 3$) which is connected via wavelength shifters to photodetectors equipped with hybrid photodiodes that can operate in high magnetic fields. A forward calorimeter made up of scintillating fibers assure a coverage up to $|\eta| < 5$.

- The muon system, made up Drift Tubes detectors (DT) and Resistive Plate Chambers (RPC) in the barrel region and by Cathode Strip Chambers (CSC) and RPC in the endcap region.

The measurement of muon momentum is performed through an intense magnetic field, produced by a 3.8 Tesla superconducting solenoid 13-m-long,

with a 6-m diameter, providing a bending power of 12 T ·m. The CMS apparatus design aims to provide maximum acceptance over the 4π solid angle. In this section the main CMS components, shown in Figure 2.4 will be described,

2.2.1 Coordinate system

In CMS a right-handed coordinate system having origin in the IP is defined. The x-axis points toward the center of the LHC ring, y-axis is perpendicular to the LHC plane, while z-axis is oriented along the beam pipe, in anti-clockwise direction. The polar angle θ is measured from the positive z-axis and the azimuthal angle ϕ is measured in the x-y plane. Pseudorapidity is defined with respect to the θ angle as follows:

$$\eta = -\ln \tan(\theta/2) \tag{2.10}$$

Combining all the information from all subsystem, CMS covers the pseudorapidity range $-6.6 < \eta < +5.2$.

2.3 Superconducting magnet

The magnetic field is one of the most important features of the CMS apparatus. It is generated by a superconducting magnet, with a diameter of 6 m and a length of 12.5 m, generating a 3.8 T magnetic field by four layers of NbTi superconducting coils. The superconductor is cooled down to 4.5 K and about 2.6 GJ are stored in the fields during operation.
In fact the requirements given by physics, namely to achieve a good momentum resolution and high efficiency in muon detection - without loosing a compact design and a good spacial resolution in the alignment of muon chambers, have driven magnet design. The choice of a solenoidal field (instead of a toroidal one) allows a bending of muon tracks in the transverse plane, improving the accuracy in the determination of muon vertex and facilitating the task of triggers based on vertex reconstruction.
The magnetic flux is returned through a 1.5 m thick saturated iron yoke, instrumented with four stations of muon chambers. The iron yoke is made up of five wheels in the barrel region, and six disks in endcap region, three for positive and three for negative z coordinate. Whereas in the core a 3.8 T magnetic field is provided, in the return yoke a 2 T reversed magnetic fields (variable in intensity and direction) is used to measure muon momentum. Of course the multiple scattering experienced by penetrating particles (e.g. muons) must be taken into account in particle reconstruction and identification. In fact this is one of the main drawbacks of such a design, as well as the space limitation for detector components placed inside the coil (which may have some consequence especially for particle calorimetry).

2.4 Tracker

The CMS tracking system aims to provide an accurate and efficient reconstruction system of charged particles tracks emerging from primary and secondary vertices. A good vertex reconstruction is one of the Physics requirements at TeV scale, especially for $b-$jets and τ physics.
It operates in an extremely high radiation density environment: at the LHC design conditions a hit rate density of 1 MHz/mm^2 at a radius of 4 cm is expected, falling to 3 kHz/mm^2 at a radius of 115 cm. It has a cylinder shape 5.8 m long, with the diameter of 2.5 m, surrounding the IP. The high granularity of the inner tracking system, as well as the fast time response needed to operate in dense pile-up environments, require multi layer geometries, a high power density for the on-detector electronics and consequently an efficient cooling systems.
The need for reducing the amount of surrounding material to limit particle interaction and energy losses, has inevitably lead to a compromise for the tracker design. Moreover, due to the intense flux of radiation in proximity of IP, electrical components with an expected lifetime of 10 years have been employed.
The CMS tracker is entirely based on silicon detector technology. It is composed of two sub-systems: the inner one is a pixel detector, the outer one is a strip detector. A cooling to -10° C (achieved through C_6F_{14} gas flow) is needed in order to reduce radiation damage, largely prevent mobility defects in the semiconductor.
With a total surface of 200 m^2, the CMS tracker extends its acceptance up to $|\eta| < 2.5$. It has been designed to have a total momentum resolution in the central region of $\frac{\Delta p_T}{p_T} = 0.15 \times p_T[\text{TeV}] \bigoplus 0.5\%$ which degrades to $\frac{\Delta p_T}{p_T} = 0.6 \times p_T[\text{TeV}] \bigoplus 0.5)\%$ as $|\eta|$ approaches to 2.5 [64]. A track reconstruction efficiency of about 98% for muons and 90% for electrons with $p_T > 1$ GeV has been measured, while tracking efficiency is greater than 95% for particles in jets with $p_T > 10$ GeV, and greater than 85% for particles in jets with $p_T > 1$ GeV.
The nominal impact parameter resolution for high momentum tracks is about 10 μm [65].

2.4.1 Pixel tracker

Pixel detector geometry (shown in Figure 2.5 as PIXEL) has been conceived to keep the occupancy below 1% up to a radius of about 10 cm and to provide high-precision three dimensional points along track trajectories. It is split into three layers, 4.4, 7.3 and 10.2 cm far from IP, typically providing three track measurements. Two endcaps at $z = \pm 34.5$ and 46.5 cm enclose barrel layers at the two opposite sides, extending the total acceptance up to $|\eta| < 2.5$ (providing a very efficient three-hit coverage until $|\eta| < 2.2$). Silicon pixels size ($100 \times 150 \mu m^2$ in $r - \phi$ and z) allows to reach the desired resolution on impact

parameter and to keep the occupancy in the order of 10^{-4} per pixel and bunch crossing.
The CMS silicon pixel tracker hosts about 66 million pixels over 1440 pixel modules.

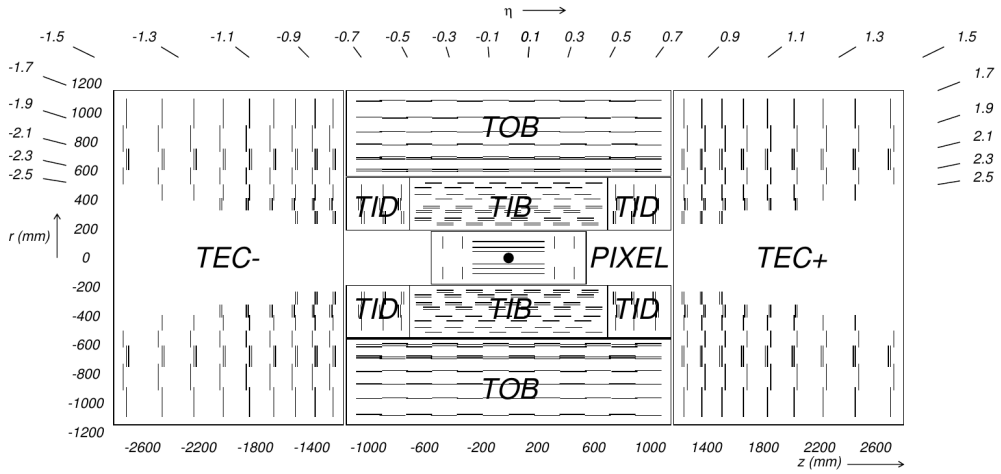


Figure 2.5: Schematic layout of CMS tracker; lines represents detector modules. The components are: Pixel Detector (PIXEL), Tracker Inner barrel (TIB), Tracker Outer barrel (TOB), Tracker Inner Disk (TID), Tracker endcap (TEC). [56]

2.4.2 Strip tracker

The strip tracker region is shown in Figure 2.5 as composed by TIB / TID / TOB / TEC+ / TEC- (standing for Tracker Inner Barrel and Disks, Outer Barrel and Endcap). TIB is made up of 4 layers extending up to a radius of 55 cm (corresponding to maximum 4 $r-\phi$ measurements) and it is confined by 3 disks (TID) at each end. The inner trackers are surrounded by TOB, which consists of 6 barrel layers (providing 6 additional measurements on $r-\phi$ plane), enclosed by 9 endcap disks for each side (providing 9 additional measurements of ϕ coordinate). The silicon strip tracker extends over a total area of 198 m^2. The need to keep low detector occupancy beyond 10 cm from IP is less severe and allows tracking operations with a cheaper solution than the pixel technology. Hence pixel detectors at larger distance (between 20 and 110 cm) are replaced by micro strip, starting with the size of 10 cm $\times$ 80 μm (occupancy 2-3% per strip), with a variable strip pitch, function of the radial distance (it ranges from 80 μm in the inner pair of TIB layers, to 183μm in the TOB.
Beyond a radial distance of 55 cm from IP, strip size is further increased to 25 cm $\times$ 180 μm thus reducing the number of readout channels. Being the electronic noise a function of the strip length, in order to keep the signal to noise ratio well above 10, the silicon thickness (approximately 320 μm in the

inner tracker) is increased to 500 μm in the outer tracker.
The CMS silicon strip tracker contains 9.3 million of strips over 15148 strip modules.

2.5 Calorimeters

Whereas just charged particles can be detected by tracker detectors, calorimeters are designed to measure the energy of charged and uncharged particles. They are concentric detectors placed just outside the tracker, inside the core of iron solenoid, designed to measure properties of both electromagnetic and hadronic radiation.
Since electron and gamma showers have usually a shorter longitudinal extension than those produced by hadrons, hadron calorimeter is placed on the outer side of an electromagnetic one.
Calorimeters are based on different detection principles [66]: when a particle impinges on the detector surface, it may interact with some active material. In fact detectors can be made up of light-emitting materials (like scintillators, noble gases or other substances suitable for Cherenkov detection), semiconductors or gaseous mixtures, interleaved by passive slabs aiming for particles showering. An interaction with the active material provide an emission of light, or charged particles via ionization processes. The total light or charge yield is proportional to the energy loss, matching radiation energy if no particle escapes from the detector.
Electrons and photons are usually promptly stopped by homogeneous crystals (more refined solutions are required to confine hadronic radiation, as shown in the next paragraphs). High energy muons ($p_T > 10$ GeV) and neutrinos escape from calorimeters, the former being detected by the external muon system, the latter just escaping with high probability from detector acceptance. At high energy as well as for high pileup running conditions, a good separation among energy deposits in η and ϕ is fundamental for an effective particle reconstruction. To meet such requirements, a trade off between costs and granularity has been carried out as well.

2.5.1 Electromagnetic calorimeter

Detector features

The electromagnetic calorimeter is a homogeneous calorimeter conceived to detect mainly electron and photons and entirely made up of lead tungstate ($PbWO_4$) crystals (Figure 2.7). The choice of such a material is motivated by several reasons: 80% of the scintillation light is emitted in 25 ns, the same bunch crossing time of the LHC; crystal is optically clear and can be considered quite radiation-hard. The light yield is driven by a wavelength distribution peaked in the range of 420-430 nm, being the number of photons emitted per

MeV approximately 4.5 at 18°C (the variation of the response to incident electrons with temperature has been measured to be $(-3.8 \pm 0.4)\%^o C^{-1}$. Moreover the high density (8.28 g/cm^3), the short radiation length (0.89 cm) and the small Molière radius (2.2 cm) allowed to build a compact device, with a high granularity. In fact light collection is not fully uniform inside the crystal due to its shape and its high refraction index (n = 2.29 in the peak range), especially in the endcap, where crystal faces are not parallel.
Aging effects, due to the high absorption rate of radiation, can give rise to impurities in the lattice and to a loss of light transmission capability. The presence of damages is monitored making use of laser light, through the so-called ECAL laser monitoring system. Even though the rise of impurities depends on the absorbed dose, is well counterbalanced by recovery processes at the operational temperature of 18°C.

Calorimeter structure

Light output in the active material is collected by avalanches photodiodes (APD), which can operate even inside strong magnetic fields. The ECAL is divided into three main parts: barrel, endcap and Preshower regions. ECAL barrel region extends up to $|\eta| < 1.479$. 62100 crystals are divided in 360 sectors in ϕ and 2×85 in η, and they are enclosed in alveolar submodules, made of 0.1 mm thick aluminum walls. Their axis form a small angle of 3° with the vector originating from the nominal IP, just to avoid the presence of cracks aligned with particle trajectories. The crystal cross section is 0.0174×0.0174 in $\eta - \phi$ plane (22×22 mm^2 at the front face of crystal and 26×26 mm^2 at the rear face). The total crystal length is 230 mm, corresponding to a radiation length of 25.8 X_0 (being X_0 the distance needed to attenuate incident radiation of a factor $1/e$). ECAL endcaps cover the range $1.479 < \eta < 3.0$, being the distance between endcap and IP of 315.4 cm. Identically shaped crystals are grouped in units of 5×5 (supercrystals) consisting of a carbon-fibre alveolar structure. Each endcap is divided into 2 halves, each one holding 3662 crystals, mounted in 138 standard super crystals and 18 partial supercrystals. Both crystals and supercrystals are placed inside an xy grid; their axis points at a distance of 1300 mm far from the IP, covering an angle ranging from 2 to 8 degrees. The crystals have a front face cross section of 28.62×28.62 mm^2 and they have a total length of 220 mm (corresponding to 24.7 X_0).
The performance of ECAL measured during test beams in terms of energy resolution are shown in Figure 2.6, where electron energy resolution as function of the initial electron energy is shown.

ECAL energy resolution can be modeled according to the following functional form:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \tag{2.11}$$

In Equation 2.11:

- S represents a stochastic term, driven mainly by three contribution: the event-by-event fluctuations in the lateral shower containment, a "photo-statistics" contribution of 2.1% and the fluctuations due to energy deposition in the Preshower absorber, that can affect the actual measurement in the Preshower detector;
- N represents the noise term, which includes electronic noise, digitization noise and pileup noise contributes;
- C indicates a constant term resulting by intercalibration errors, non uniformity of the longitudinal light collection and leakage of energy from the back of the crystals.

ECAL energy resolution has been measured [67] using electron beams between with energies ranging between 20 and 250 GeV, hitting the crystal around the point of maximum containment. The typical energy resolution is found to be

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{0.12}{E}\right)^2 + (0.30\%)^2 \tag{2.12}$$

where E is measured in GeV. The energy resolution as a function of incident electron energy is shown in Figure 2.6 for a central impact of the beam into a crystal.

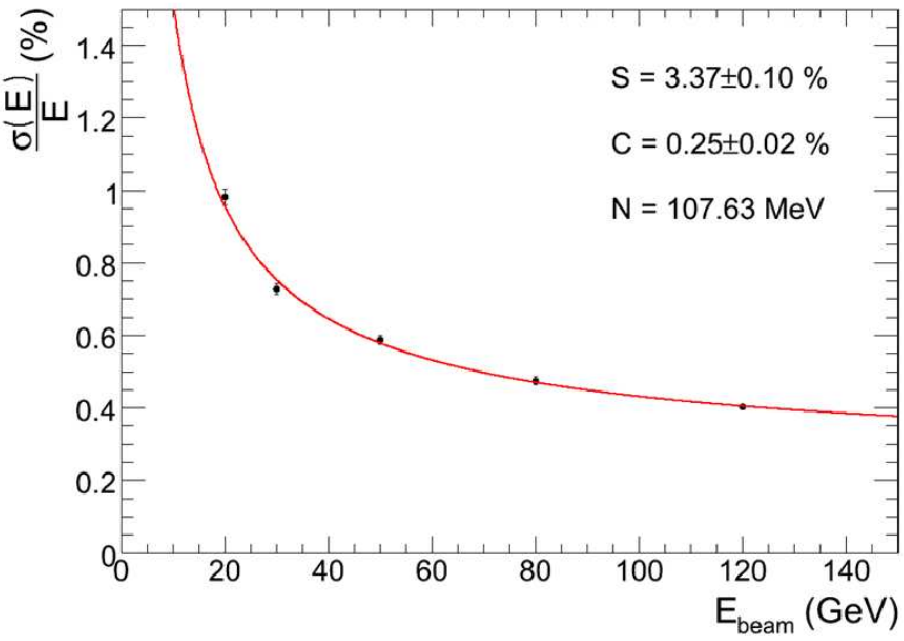


Figure 2.6: Energy resolution of a matrix of 3×3 crystals. The energy is reconstructed by summing the signals of the 3×3 crystals, for electrons entering the central (4×4) mm^2 area [67].

Preshower

In front of the ECAL endcap crystals, a Preshower detector has been placed, covering the fiducial region defined by $1.653 < |\eta| < 2.6$. Indeed it has been design to discriminate genuine photons from neutral pions, to improve electrons identification and to improve the accuracy in determining electrons and photons position with high granularity.
The Preshower is in fact a sampling calorimeter composed by two layers of lead radiator, designed to initiate electromagnetic showers. Behind each radiator silicon strip sensors, oriented according to two orthogonal planes, are placed to measure the deposited energy and the transverse shower profiles. The first sensor plane is 2 X_0 long, while the second one is 1 X_0 long. In total the Preshower is long about 20 cm.

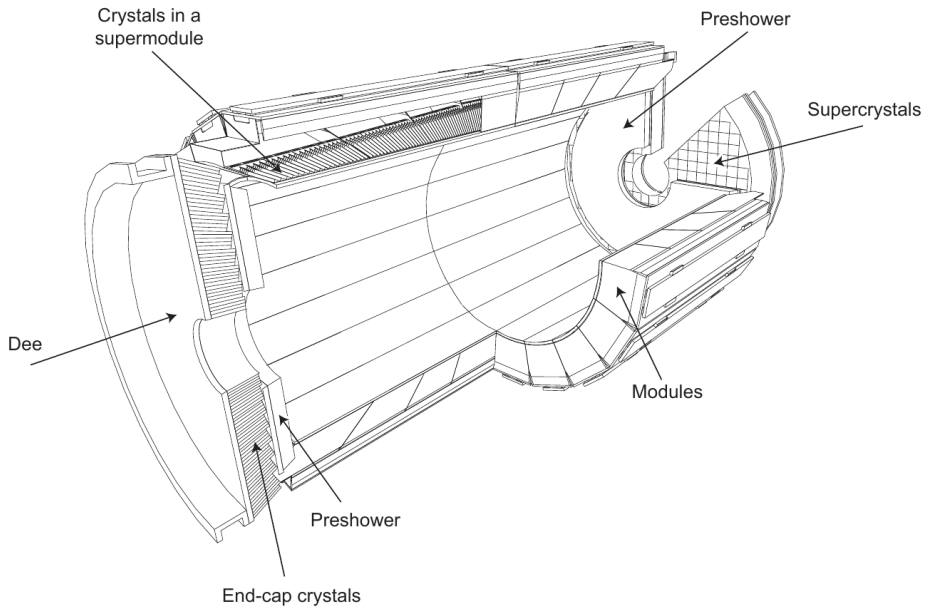


Figure 2.7: Schematic of CMS electromagnetic calorimeter, showing the arrangement of crystal modules, supermodules and endcaps, with the Preshower in front. [56]

2.5.2 Hadronic calorimeter

The hadronic calorimeter (HCAL) provides a complementary information measuring hadron jets energy. Therefore, merging data from calorimeters, even the existence of neutrinos and other exotic particles can be inferred measuring transverse missing energy for each event.
HCAL is a sampling calorimeter, meaning that it uses alternating layers of dens "absorber" material, aiming to produce particle cascades from both charged and uncharged particles, plus scintillator materials, producing light pulses when

crossed by charges. Signal is then conveyed by optical fibers into readout boxes, where is amplified by photodetectors.
A good containment of hadronic shower is fundamental to perform good energy measurements: actually HCAL ranges from the outer edge of electromagnetic calorimeter (R = 1.77 m) to the magnetic coil (R = 2.95), this latter limiting the total amount of material that can be used to absorb showers. However a *tail catcher* is placed outside the coil extending up to $|\eta < 1.3|$, in order to improve the absorption of escaping particles (Hadron Outer, HO). HCAL itself is divided in barrel (HB) and endcap (HE) region, which assure a coverage until $|\eta| < 3$; a very forward calorimeter (HF) is also placed at 11.2 m from the IP, extending coverage to $|\eta| = 5.2$ using a Cherenkov-based detector (Figure 2.8). The HB covers the pseudorapidity range within $|\eta| < 1.3$. It is divided into two half-barrel section, inserted from both end of superconducting solenoid barrel cryostat. The calorimeter is composed by 36 identical azimuthal wedges (Figure 2.9), which form two half-barrels. Wedges, segmented into 4 azimuthal angle ϕ-sectors, are made of flat brass absorbers, aligned with beam axis. The plates are staggered in such a way that no projective dead material is present for the full radial extent of the wedge. Plastic scintillator is further divided into 16 η sectors, providing a granularity in cells $(\Delta\eta, \Delta\phi) = (0.087, 0.087)$. The wedges are then mounted in such a way the cracks between them is less than 2 mm.

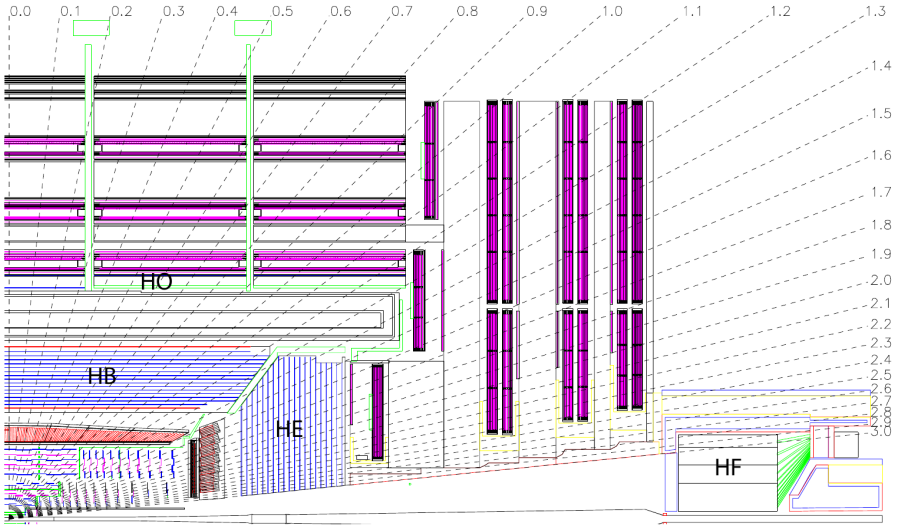


Figure 2.8: Longitudinal view of the CMS detector, showing the locations of the Hadron Barrel (HB), Endcap (HE), Outer (HO) and Forward (HF) calorimeters. [56]

The absorber is made up by eight 50.5 mm-thick and six 56.5-mm-thick brass plates enclosed in steel plates. The total absorber thickness at 90° is 5.82 interaction lengths (λ_I); however the effective thickness increases as a

function of the $1/\sin(\theta)$ being θ the polar angle, resulting in 10.6 λ_l at $|\eta| = 1.3$ (in addiction 1.1 λ_l must be added, accounting for the bulk material due to the electromagnetic crystal in front of HB) .
The active material made of 3.7 mm-thick Kururay SCSN81 plastic scintillator featuring a long-term stability and a moderate radiation hardness. Just the first layer, designed to sample showers developing in inert material between EB and HB is made of 9-mm-thick Bicron BC408. The light yield is conveyed to readout electronics through optical fibers.
Since at $|\eta| = 0$ just 5 λ_l are provided by HB, as mentioned before a Hadronic Outer Calorimeter is placed just outside the coil, to measure shower energy inside inert material. In such a way about 11.8 radiation length can be achieved (including EB, HB and iron yoke). The mean fraction of energy deposited in HO can increase up to 4.3% for 300 GeV pions.
The HE extends over the rapidity range $1.3 < |\eta| < 3$. It covers 13.2% of the solid angle, accounting for about 34% of the final state particles. A high radiation tolerance is requested: therefore a trade off among costs, the requirement to handle non-magnetic material, to have good mechanical properties and a large equivalent thickness, has lead to the choice of C26000 cartridge brass as the absorber material. The design of the absorber minimizes the cracks between HB and HE and provides a self-supporting hermetic construction. Brass plates have a thickness of 79 mm, interleaved with 9-mm gaps where scintillators are mounted. The outer layers are fixed to a 10-cm thick stainless steel support plate. The total length of the calorimeter, including electromagnetic crystals, is about 10 interaction lengths (λ_l). The total energy resolution can be modeled according to the following expression:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{120\%}{\sqrt{E}}\right)^2 + (6.9\%)^2 \tag{2.13}$$

being E measured in GeV.
Forward hadron calorimeter has been conceived to be exposed to very high particle fluxes. On average 760 GeV per p-p interaction are absorbed by the two forward calorimeters, being the distribution peaked at very high pseudo-rapidities.
After 10 years of operations at design luminosity, a dose of 10 MGy is expected to be absorbed at $|\eta| = 5$. The need to build a stable detector in so hard conditions for at least 10 years, has pushed quartz fibers to be used as active medium, providing about 10 λ_l.
The HF signal is generated by showering charged particle overcoming Cherenkov threshold (190 keV for electrons), making the calorimeter mostly sensitive to shower electromagnetic fraction. The light output is finally conveyed through fibers to readout electronics.

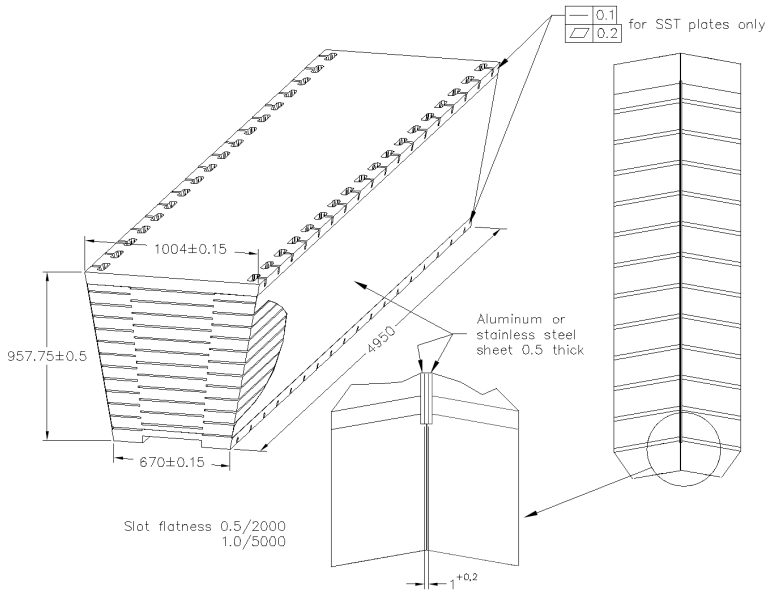


Figure 2.9: Isometric view of HB wedges, showing the hermetic design of the scintillator sampling. [56]

2.5.3 The muon system

The muon system has a primary importance in CMS design: actually many interesting signatures are characterized by the presence of the muon in the final state. The most famous one, $H \to ZZ \to 4l$, which has been called "gold plated channel" in case the four leptons are muons, has recently led to the discovery of the Higgs boson. Indeed in the CMS physics program a long list of exotic and SM processes require an excellent muon identification on a wide coverage angle.
The crucial importance played by muon detection, has led to the design of the muon system, which has three function: muon identification, momentum measurement and trigger. Actually due to the huge amount of information collected, Tracker data cannot be used for online trigger purposes; on the other hand, tracker hits alone would be indistinguishable without any other information. Thus muon tracks reconstruction is performed interpolating muon chambers hits and then propagating back muon trajectories to the tracker hits. After the hits have been assigned, the track is fully reconstructed using reconstruction algorithms described later (Sec. 3.3.1) .
As the Muon System covers a huge surface (about 25000 m^2), the employment of gaseous detectors looks to be a natural solution to perform muon identification, being relatively inexpensive, reliable and robust.
The working principle is the following: when a charged particle crosses a sen-

sitive volume containing suitable gas mixture, it ionizes gas molecules and produces free electrons. An ideal mixture should have a high gain, a low working voltage and a high rate capability. Therefore features like low ionization potential, unflammability, as well as quenching properties[1] are preferable.
Free electrons are accelerated by an external electric field towards an anode, triggering an avalanche multiplication process. The electric field can be either uniform or shaped with more complex geometries, according to the detector technology used. In some cases electric field intensity may reach several tens of kV/m, especially in the electrodes neighborhood. It can be shown that the motion of charged particles (either ions or electrons, depending on the detector technology) induces on the cathode a charge, which is amplified and measured.
However three different technologies are used, whose features make them appropriate to play different roles. In the barrel a combination of Drift Tube Chambers (DT) and Resistive Plate Chambers (RPC) are mounted on five wheels. In the endcap, composed by three parallel disks, Cathode Strip Chambers (CSC) and Resistive Plate Chambers are employed instead. DT and CSC are mainly used for tracking purposes due to their good space resolution and timing properties, the former taking benefit from the relatively low muon flux in the barrel, the latter thanks to its reliability even at high particle fluxes. RPC are mostly used as trigger detectors instead, because of their good time resolution and their high rate capabilities.
In the barrel region, both RPCs and DT provide an η coverage extending up to 1.2; in the endcap CSC cover the pseudorapidity region $0.9 < |\eta| < 2.4$, while endcap RPCs cover $0.9 < \eta < 1.6$
Muon chambers are embedded in the iron yoke, taking advantage by the 2 T magnetic field which provides a good momentum resolution and trigger capabilities. Iron yoke serves also as absorber of punch through hadrons, though it provides a smearing on muon p_T which dominates momentum resolution up to about 300 GeV, because of the muon multiple scattering inside inert material.

Drift tube chambers

Drift Tube chambers (DTs) are mounted in the barrel region ($|\eta| < 1.2$) (Figure 2.11), where a relatively low intensity of magnetic field and particle flux is expected. They are made up of a a 50 μm-thick anode wire, surrounded by a stainless steel case, having a rectangular section (42 × 13 mm). An aluminum strip on the shortest side of the cell serves as cathode: it is kept at a potential of about +3600 V, while the anode at a potential of about -1800 V. Each chamber is filled with a gas mixture (85% CO_2, 15 % Ar) chosen for

[1]Excited atoms often produce high energy photons which can cause further avalanches and a continuous discharges inside the gas.
Quenchers are polyatomic gases (except for few inorganic cases like CO_2, BF_3) which are able to absorb radiated photons, dissipating energy through molecules collision or dissociation. When these gases are added to detector gas mixture, further avalanches triggered by high energy photons can be avoided and much higher gains can be achieved.

its quenching properties and since it is not inflammable. A typical chamber unit is sketched in Figure 2.10.

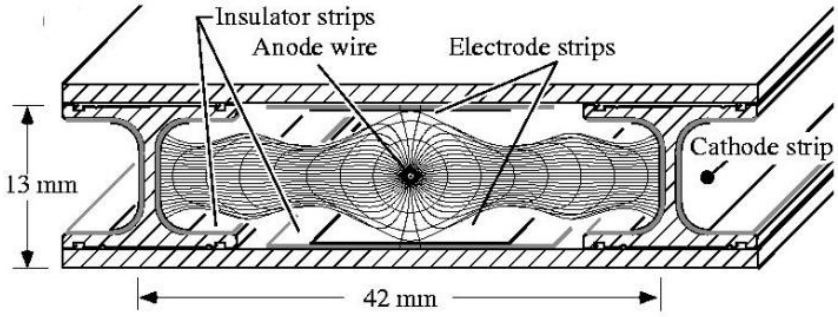


Figure 2.10: Sketch of a cell showing drift lines and isochrones. The plates at the top and bottom of the cell are at ground potential. The voltages applied to the electrodes are +3600 V for wires, +1800 V for strips, and -1200 V for cathodes. [56]

DT are mounted on barrel wheels, on 12 sectors, each one covering a ϕ angle of 30°. Each sector is further segmented in four station (named MB1, MB2, MB3, MB4) placed at increasing radial distance from the beam pipe. Each one of the first three station contains 3 superlayers, each containing 4 layers of tubes. The upper and lower superlayer provide $r - \phi$ coordinates, while the chambers in the middle give a measurement of the coordinate in z direction (having axis perpendicular to z axis). Chambers are arranged in such a way to avoid possible cracks which may give rise to dead regions.
The iron yoke is interleaved between two subsequent stations, with a variable thickness calculated to optimize muon p_T measurement: actually chamber sizes increase moving towards the outer part of barrel. With an efficiency close to 99%, the space resolution can reach about 200 μm, depending on the incidence angle and on the magnetic field.

Cathode strip chambers

Cathode Strip Chambers (CSCs) are employed in endcap region ($0.9 < |\eta| < 2.4$), where the high particle flux and the presence of magnetic field in the instrumented region make CSC more suitable than DTs. Actually the former has also shorter drift paths and higher rate capability. As DTs they are named

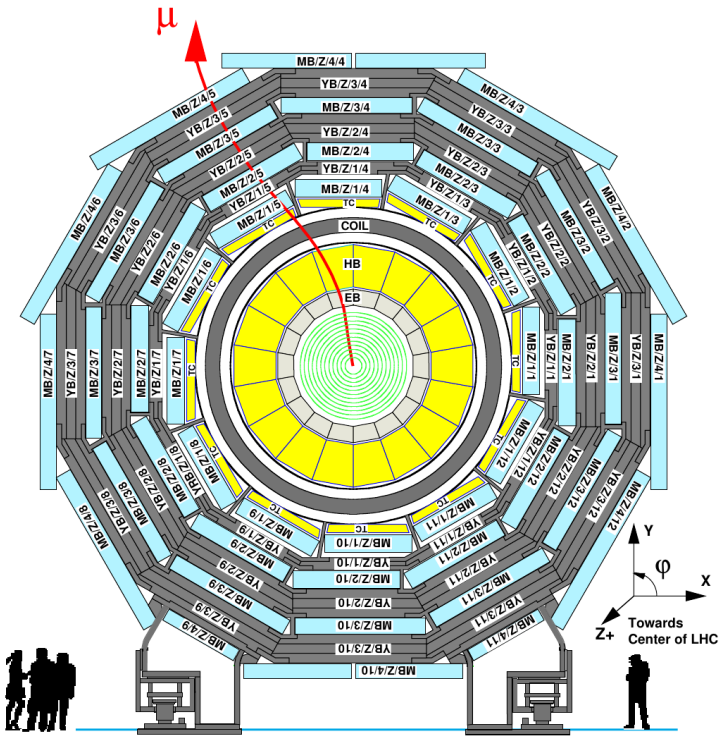


Figure 2.11: Layout of the CMS barrel muon DT chambers in one of the 5 wheels. [56]

as ME1, ME2, ME3, ME4 (Figure 2.12).
Each chambers is made up by six layers of anode wires, interleaved by 7 trapezoidal cathode strip panels, and are filled with a gas mixture composed by Ar(30%), CO_2 (50%) and CF_4 (20%). Wires are oriented along ϕ direction, defining η coordinate, strips run radially with $\Delta\phi$ distance. ϕ coordinate is obtained interpolating positions of strip pulses and weighting them by the induced charge. Each particle hit is encoded along r and ϕ coordinates, using anode wires and cathode strips information.
486 chambers are mounted on the endcap return iron yoke, which provide both bending field and structural support for CSCs. Single chamber size varies from station to station (the largest having size 3.3×1.5 m). Each disk is split in 18 sectors for ME1 and 36 for ME2, ME3, ME4, the first chamber starting from $\phi = -5$, the last one at $+15°$ (or $+5$).

Resistive plate chambers

Resistive Plate Chambers (RPCs) are employed both as a complement to CSC and DT systems and as a redundant muon detection system. Thanks to their good timing properties, they are mainly used for triggering purposes, providing an unambiguous bunch crossing identification (time resolution < 4 ns). In

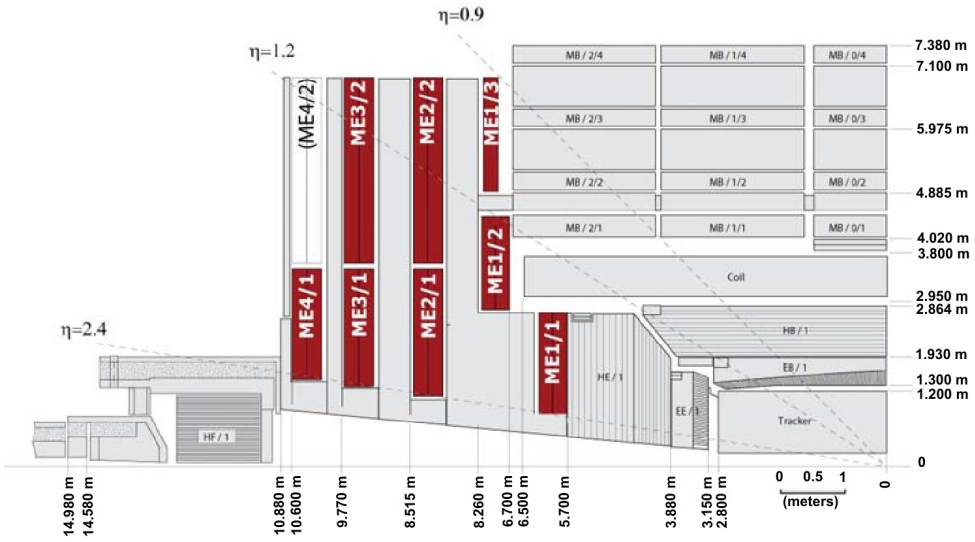


Figure 2.12: Quarter view of the CMS detector. Cathode strip chambers of the endcap Muon system are highlighted.[56]

Figure 2.13 a quarter view of CMS detector is shown, where RPC chambers have been highlighted. Each chamber is made up by at least two high resistivity HPL (High Pressure Laminate) panels ($10^{11}\Omega \cdot m$), forming a 2 mm-thick gap, filled with a quenching gas mixture (96% $C_2H_2F_4$, 3.5% C_4H_{10} and 0.5% SF_4). A thin graphite layers covers each Bakelite panel externally, serving as a high voltage (HV) electrode. The whole system is then enclosed between two mylar sheets, which separates high voltage electrodes from readout strips. Indeed RPC chambers are made by two adjacent gaps, to optimize detection efficiency. In such cases common readout strips are placed between the two gaps.
RPCs are mounted both on barrel and endcap yoke. In barrel region ($0<|\eta|<1.2$), they are arranged in 12 ϕ sectors, each of them being further split into four stations, respectively named RB1in, RB1out, RB2in, RB2out, RB3 and RB4. In endcap region ($0.9<|\eta|<1.6$) they are mounted on three endcap disks, divided in 36 partially overlapped sectors, which avoids cracks and dead regions. Each disk should be made of three concentric rings, assuring an η coverage until $|\eta| = 1.6$. In fact just two rings have been implemented, while the complete building of the high η region ring is foreseen for future upgrades.

2.5.4 Trigger system

With a design energy of 14 TeV in the center of mass and a luminosity of 10^{34} collision per second per cm^2, the LHC is able to deliver 40 millions of bunch collision per second, namely an average of 20 interaction per bunch crossing. A rate of 100 TB/s, coming from the 10^8 CMS readout channels, is expected. Given the magnitude of this numbers, a persistent storage of data is impossible

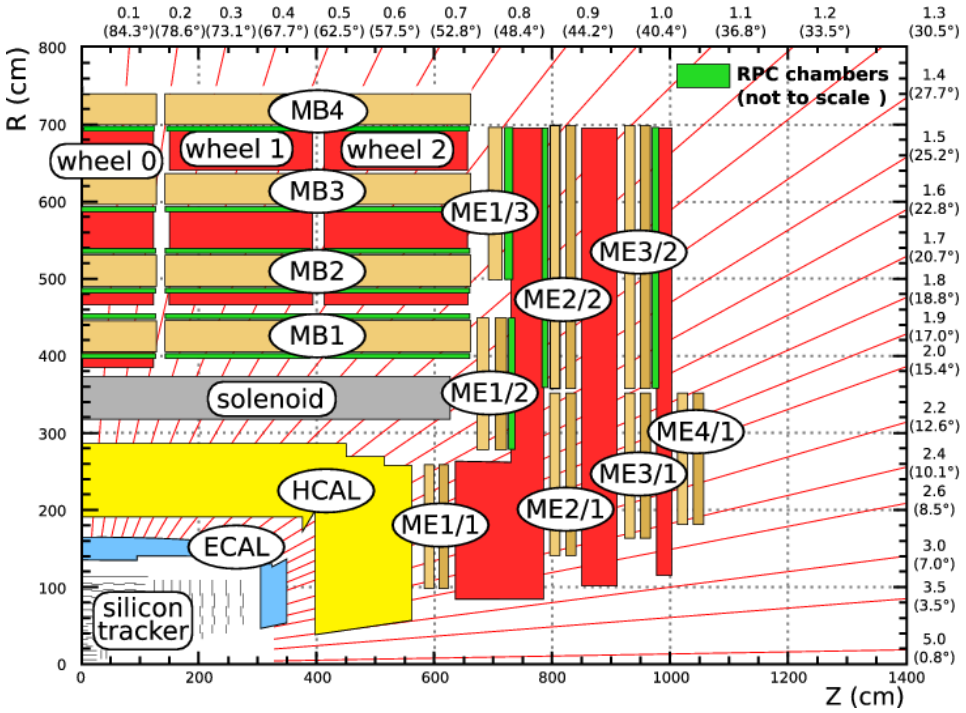


Figure 2.13: Quarter view of the CMS detector. RPC system has been highlighted. [56]

and a drastic reduction of online data flow is compulsory. A target event rate of $\mathcal{O}(100)$ Hz is achieved employing a Trigger System, which represents the first step in the physical event selection. Such a system is based on two steps, the first named Level 1 (L1) Trigger, namely a hardware-based trigger, providing a first coarse event selection, the second named High Level Trigger (HLT), where selection algorithms are performed by a dedicated software.

Level 1 trigger

The Level 1 trigger is the first stage of trigger processing, and provides a drop in the event rate to about 100 kHz (matching input rate capability of HLT system). Since the processing time is approximately 3.2 μs, bunch crossing information is pipelined into a FIFO buffer memory, able to host 128 events. Since short processing times are required, the L1 system takes into account just a fraction of the whole information coming from sub detector systems, ignoring calibration data. Hardware implementation makes use of Field Programmable Gate Array technology (FPGA), Application Specific Integrated Circuit (ASIC) technology and programmable memory Lookup Tables (LUT). L1 workflow is based on Local, Regional and Global components (Figure 2.14): the Local Muon Trigger search for signals compatible with those released in DT, CSC and RPC system by a particle originating from the IP; the Local Calorimeter Trigger look for energy deposit clusters in ECAL, HCAL or HF towers. A Regional Trigger takes over this information, determining if such

objects are electron or muon candidates through a pattern logic. A Regional Trigger combines all the available information in a spacial region in order to form a candidates, ranking them according to their energy or momentum. The four objects with the highest rank are the passed to the Global Muon Trigger and Global Calorimeter Trigger, which calculates visible and missing energy and makes a further selection on the four highest ranked objects (isolated/not-isolated electrons, muons and jets) over the whole experiment. Such candidates are then passed to the Global Trigger (GT), which decides if the event can be rejected or not. The GT result is communicated through the sub-detectors through the Timing, Trigger and Control (TTC) system. If the event is accepted all the detectors are read out through the detector DAQ (Data Acquisition System). Thus the complete information is passed to the event builder network, and then to the HLT system.

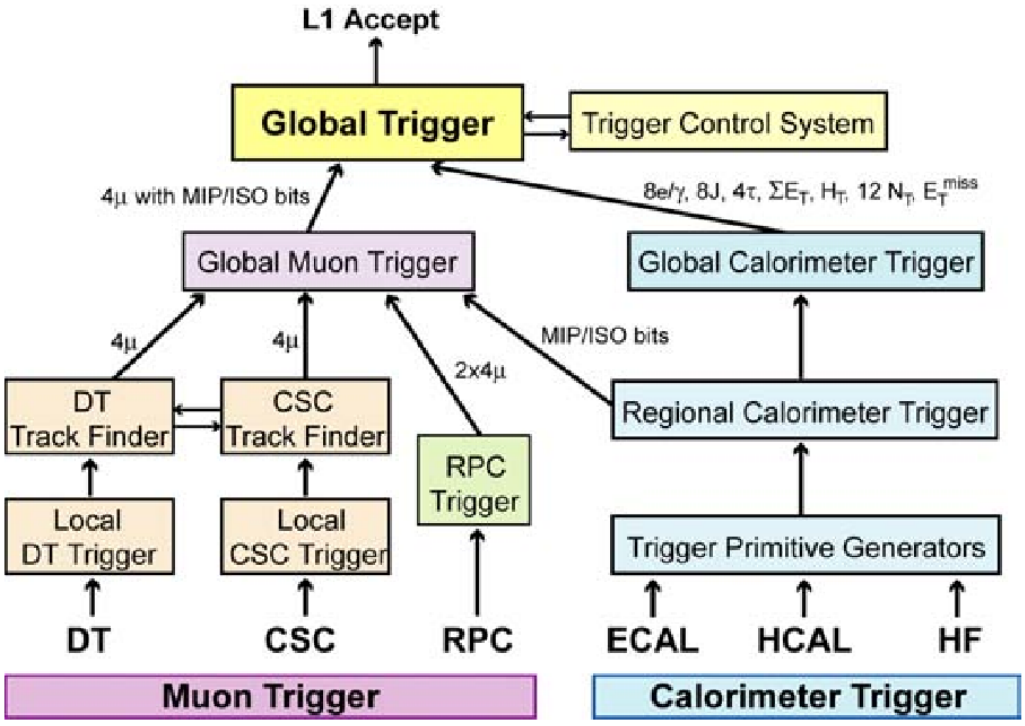


Figure 2.14: Architecture of Level-1 Trigger [56].

High Level Trigger

High Level Trigger (HLT) is a completely software based system, which reads high-resolution data from a pipelined memory. Actually, the time scale available for HLT processing is in the order of 1 second. A PC filter farm of 1000 computers determines the type and multiplicity of particles which belong to a given event and, at the end of the processing steps, the output rate results to be approximately 100 Hz, which is the maximum rate for mass storage.

The selection sequence is split in two subsequent logical steps: Level-2 and Level-3. The Level-2 elaborates the full information from both calorimeters and muon detectors, and drop the rate of one order of magnitude. Tracker informations are used in Level-3 selection instead, due to the high amount of time requested by reconstruction algorithms.
For each event trigger paths, namely a ordered sequences of filters based on different algorithms, check if all the event requirements are met and the single event can be written on tape. The information is then stored and made available to an offline use. HLT system is strongly customizable: therefore, because of the different running conditions of the LHC during the year, several *HLT menus*, namely sets of trigger paths grouped by typology, are modified each months to suit the collider settings.

Chapter 3

CMS event reconstruction

3.1 The CMS software

The CMS software, named CMSSW, is the official software framework used by the whole Collaboration for calibration, simulation, reconstruction and data analysis purposes. All these tasks must meet some technical requirements, in terms of computing resources (memory consumption, CPU processing time) and financial constraints. At the same time, they must provide a solution for three big challenges [56]:

- **Large scale**: high granularity, resulting from the detailed mapping of the whole detector, coupled to the very large statistics datasets needed to discover even rare signals, require a huge computing effort. A large scale system, supporting efficient approaches to data reduction and pattern recognition is mandatory. Moreover, the software must be developed by geographically distributed group of experimenters.

- **Flexibility**: any user should be able to access any data item recorded or calculated during the lifetime of the experiment. Moreover the software must evolve through the goals of the experiment, and used in different environments. New experimental conditions and analysis requirements cannot be defined in advance and once for all.

- **Manageability**: the software and computing resources as well should be maintained for, at least, 15 years; a generation turn over of hardware and software should have be foreseen.

CMSSW represents the modular, high-level system with loosely coupled components and well defined interfaces, which has been chosen to meet all the previous requirements. It is an object-oriented framework, written in C++ and Python, built as a collection of different modules, designed for specific functions.

3.1.1 Architecture design

The overall structure for the CMS software is composed by [111]:

- An application framework customizable for any computer architecture
- Physics software modules, each corresponding to a well defined data processing step and with a common interface to the driving software framework.
- Services and utility toolkits, that can be used by any of the physics modules

The framework defines the top level abstractions, their behavior and collaboration patterns. It is composed by two parts: a set of classes that capture CMS specific concepts like detector components and event features; a control policy that controls the instances of those classes taking care of the flow of control, module scheduling, input/output, etc.
The physics and utility modules can be plugged in the application framework at runtime, and a choice among several versions of modules can be made. The physics modules do not communicate with each other directly but only through the data access protocols that are part of the framework itself.
Service and utility toolkits can be divided into two categories: physics type services (histogrammers, fitters, mathematical algorithms, geometry and physics routines..) and computer services (data access, inter module communication, user interface). The application framework and the services also decouple physics modules from the underlying technology used for accessing to external services (e.g. calibration database). This approach grants a smooth transition to new technologies, localizing changes in the framework and in the specific components of the toolkit.

3.1.2 Framework and modules

The Event Data Model (EDM) represents the most important technical structure of CMSSW. As the word suggests, this structure is based on the concept of *Event*, which represents the physical information extracted from the detector data (e.g. list of reconstructed muons, electrons, jets...). A C++ object, which stores all the event information, is made available for any physics modules.
The Event Data Model, as well as the set of services dedicated to calibration, alignment, etc. are part of the Framework, namely the general infrastructure needed for modules development. The framework has been conceived to be easy to use and to encourage users to make their own data objects, in order to remove the need of specialized private data analysis formats.
At the Event level this modular approach for example translates into the possibility to develop and test distinct components of a sequence, which can correspond to trigger emulation, simulation, reconstruction and analysis. Modules

are used as functional units that can be employed as event data producers (used in L1 trigger, emulation, HLT, reconstruction, simulation etc.), filters (useful for trigger purposes and offline selection), analyzers designed to produce histograms and summaries, but not to modify event data) and other use cases dedicated to the production and use of calibration data.
From a technical point of view the reconstruction workflow starts as a simple collection of RAW data, stored in a single entity and corresponding to the digitized data produced by the CMS detectors. When raw event data are processed, the resulting higher-level objects (e.g. Calorimeter clusters of cells, reconstructed tracks) products represent the so-called RECO tier of the event information (RECO). The C++ objects representing the various collections of reconstructed quantities have a format recognized by the ROOT [81] software analysis framework and can therefore be used in this framework for final When data are put into the Event, the provenance information (namely the module that created data) allows to keep track of how a particular reconstruction result has been obtained.

3.2 Simulation

In CMSSW, the description of CMS detector and the simulation of the propagation of the primary particles through the detector are based on the GEANT4 [76] simulation toolkit which is integrated to the CMS software framework and Event Data Model. All CMS detectors are simulated both in the central region (namely Tracker, Calorimeters and Muon Systems) and in the forward regions (such as TOTEM [57]). A detailed map of the magnetic field produced by the solenoid is included. The full simulation program models in detail also the detector response (called *digitization*).
The simulation component of CMSSW has been deployed since the 2004 CMS Data Challenge and has been extensively validated by detailed comparison with test-beam data as well as GEANT3 available simulations.
The detailed simulation workflow is as follows:

- A physics group configures an appropriate Monte Carlo event generator (several generators are used, such as MadGraph [73], POWHEG [71], Pythia [72], ALPGEN [75]...) to simulate the primary interactions between colliding protons: in this work samples generated with MadGraph 5, Pythia 6 and 8, POWHEG and ALPGEN have been used;

- The production team/system runs the generator software to produce generator event data files in HepMC format;

- The physics group validates the generator data samples and selects a configuration for the GEANT4 simulation (detector configuration, physics cuts, etc.);

- An official CMS production team/system runs the GEANT4-based simulation of CMS, with generator events as input, to produce (using the standard CMS framework) persistent hits in the sensitive detectors;

- An official CMS physics group validates these hit data which are then used as input to the subsequent digitization step, allowing for the addition of pile-up. This step converts hits into digital detector signals (also known as "digis") which correspond to the output of the CMS detector electronics;

- Finally event reconstruction can be performed, possibly including L1 trigger emulation and HLT software modules. The reconstruction process starts from locally-reconstructed quantities (such as Reconstructed Hits and Segments, in the muons case) and finally provides a list of physics objects.

3.2.1 Event production

Event generation

The event generation always starts with the choice of an event generator. A quite common strategy is to employ a Matrix Element generator, such as MadGraph5, which is able to handle a partonic process given a pre-defined input model (some limitations on input models might be imposed by the Matrix Element generator).
MadGraph is able to calculate the matrix element of a physics process at leading order, starting from the factorization theorem, which extend parton model idea (developed for deep inelastic scattering) to hard scattering processes. This approach absorbs the contribution of large logarithms from gluons emitted collinear with incoming quarks (whose contributions prevent the convergence of perturbative expansion) and can be factorized into renormalized parton distributions.
The hadronic cross section σ_{AB} of a generic process $\sigma(AB \to X)$ can be written:

$$\sigma_{AB} = \int dx_a dx_b f_{a/A}(x_a, Q^2) f_{b/B}(x_b, Q^2) \hat{\sigma}_{ab\to X} \tag{3.1}$$

being a, b respectively incoming quark belonging to incoming hadrons (A and B), while Q^2 represents the large momentum scale that characterizes the hard scattering process.
To take into account higher order contributions one should further expand the process perturbatively, as a sum of higher order terms $\hat{\sigma_1}$, $\hat{\sigma_2}$:

$$\begin{aligned}\sigma_{AB} = \int & dx_a dx_b f_{a/A}(x_a, \mu_F^2) f_{b/B}(x_b, \mu_F^2) \\ & \times \left(\hat{\sigma_0} + \alpha_S(\mu_R^2)\hat{\sigma_1} + \alpha_S^2(\mu_R^2)\hat{\sigma_2} + ...\right)_{ab\to X}\end{aligned} \tag{3.2}$$

where μ_F is the factorization scale (which can be conceived as the scale that separates long and short distance physics) and μ_R is the renormalization scale for QCD running coupling. In perturbation theory the cross section calculated to all orders is invariant under the changes of these parameters, since the μ_F^2 and μ_R^2 dependence of coefficients compensate exactly the scale dependence of the parton distribution and the coupling constants.
In absence of a complete higher-order expansion, one usually chooses μ_F and μ_R in order to avoid the appearance in the expansion series of large logarithms. Often $\mu_F = \mu_R$ is assumed (suitable choices for scales could be M_Z for Drell-Yan process, top mass for $t\bar{t}$ production etc.). Finally a numerical integration over the variables x_a, x_b and any phase-space variables associated to the final state is performed.
It's worth to note that leading order calculation does not provide accurate estimate of LHC processes cross-sections. However a reduction of uncertainties can be obtained performing (at least) next-to-leading order (NLO) calculations. First, one should take into account all the diagrams that can contribute with an additional strong coupling α_S, appearing as virtual (loop) contribution or real radiation; in principle, one can also reduce the scale dependence on unphysical μ_F and μ_R, developing α_S expansion series; also the k-factor, namely the ratio between NLO and LO cross section, can be used as a solution to encapsulate NLO correction to lowest order calculation: however, this ratio varies according to phase-space kinematic cuts, on the PDF and more in general on the phase space region.

Parton showering

Once the hard scattering particles have been made persistent, parton shower can be performed to simulate parton evolutions from high energy scale to an energy scale close to Λ_{QCD}.
At this lower energy scale, one can exploit a non-perturbative model that can be used to move from partons to hadrons description. Using DGLAP formalism [105], the evolution of parton fragmentation function can be calculated writing the solution of DGLAP equations with the help of Sudakhov form factors:

$$\Delta(t) \equiv \exp\left(-\int_{t_0}^{t} \frac{dt'}{t'} \int \frac{dz}{z} \frac{\alpha_S}{2\pi} P(z) \frac{f(x/z,t)}{f(x,t)}\right) \tag{3.3}$$

Such form factors represents the probability of evolving from the higher (hard) scale t to the lower (cutoff) scale t_0 without the emission of gluons with an energy higher than a given value. $P(z)$ represents the splitting function for the branching under consideration. Parton showers generate the values of an evolution variable t, a momentum fraction z and an azimuthal angle ϕ, as well as the flavor emitted during showering.

Jet-parton matching

Parton shower (PS) describe processes in regions which are dominated by soft and collinear gluon emission, based on Sudakhov form factors. Matrix Elements (ME) provide a description of high-energetic and well separated partons, as well as the effects of interference between amplitudes with the same external partons. However it diverges as partons become soft or collinear [113].
Parton-jet matching procedure avoids overlap between ME and PS phase-space description in multi-jets simulations, and it can be performed according different schemes. In MadGraph three different matching schemes (CKKM, MLM cone-jets, MLM k_T-jets) are implemented. These algorithms mainly differ for the jet definition used, for the method employed to accept/reject events and for details about jet veto starting conditions in parton shower.
For sake of shortness, just k_T-jet MLM principle will be illustrated, even though the extension to cone-jet algorithm is quite straightforward. Final state partons are clustered according to the k_T algorithm. The smallest k_T value is constrained above a cutoff scale xq_{cut}, may be seen as the minimal distance in the phase space between extra partons. k_T value is also used as renormalization scale for α_S in the vertices corresponding to QCD emission. Factorization scale is assumed to match the transverse mass $m_T^2 = p_T^2 + m^2$ of the particle produced in the central process. The event is then passed to parton shower program (Pythia). After showering, but before hadronization and decays, the final state parton are clustered into jets, setting a cutoff scale $q_T > xq_{cut}$ (for $t\bar{t}$ production, one can set $xq_{cut} = 20$ GeV/c, and q_{cut} between 30 and 40 GeV/c). These jets are then compared to the original partons from the matrix element. If the jet measure $k_T(\text{ parton, jet}) < q_{cut}$, it matches to the closest parton. The event is accepted if all the jets are matched (except for highest multiplicity samples, where extra jets are allowed to exist below the k_T scale of the softest parton of matrix element in the event); otherwise is removed from the simulated sample.
As a rule of thumb a proper q_{cut} value can be chosen assuring that the differential jet ratio distribution (Figure 3.1) has a global shape which is invariant under the choice of cutoff threshold.
The q_{cut} value can be chosen as the value that assures a smooth transition from one region of the phase-space to the other.

3.3 Reconstruction

Once particles have been generated, Geant4 [76] simulation tool take them over, propagating them through the whole detector. The energy released in the interactions between radiation and matter is then converted into digital pulses. On the basis of the collected signals, the object reconstruction is initiated by a set of algorithms which transform raw data into *Candidates*, namely the prototypes of the physical objects (electrons, muons, jets, photons..) that will

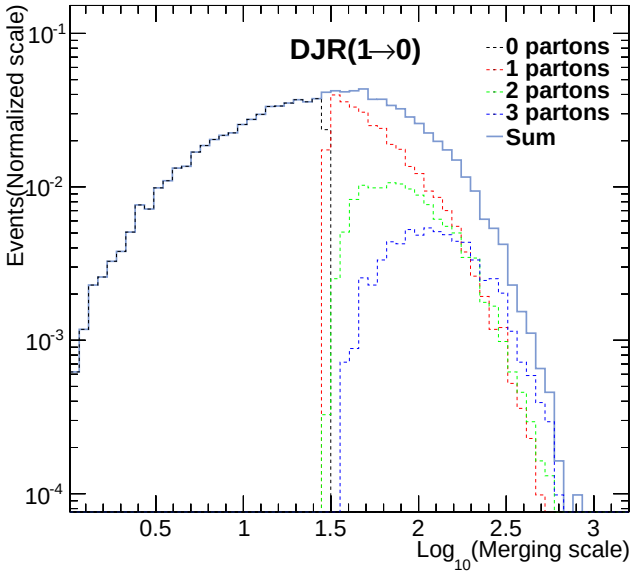


Figure 3.1: Differential Jet Ratio for a $t\bar{t}+1,2,3,4$ jets sample at 8 TeV. This plot represents the distribution of the transition value Q, namely the value of the resolution for which an n jet event becomes an $n-1$ jet event (in this case, the transition from 1 to 0 jets is shown). The plot has been produced using Pythia6 on top of a MG5 sample, setting q_{cut} to 30 GeV. The contribution of one, two three and four extra parton events are shown, as well as the total number of events (in light blue).

be used in the analysis.
In the following paragraph, a general but not exhaustive review on fundamental reconstruction methods of muons, electron, jets and transverse missing energy (E_T^{miss}) will be given, focusing on the most relevant issues that are implied by the reconstruction process.

3.3.1 Muon reconstruction

Muon reconstruction starts with the so-called "trajectory seeding". In CMS trajectory seeds are hit-based seeds (or state-based seeds, using momentum information). Hit-pairs (or hit-triplets) are required to be compatible with beam spot (further criteria can be added, for instance imposing the hit position is placed in a given region). The Seed Generator is based on DT and CSC segments: the former provides track segments in the ϕ projection (being $\Delta\phi$ the bending angle with respect to the vertex direction) and hit patterns in η projection; the latter delivers three dimensional track segments.
Trajectory building then starts in the direction specified by seed, towards subsequent layers: in the standard configuration parameters are propagated from outer detector layers toward the innermost compatible ones. Compatible hits

are searched and the track finding and fitting is accomplished by an iterative Kalman filter technique [101]. Material effects, mainly due to random Coulomb scattering, are included in the iterative steps, since they introduce a gaussian-distributed uncertainty on scattering angle. Particle propagation is a very time consuming phase. Step by step, along trajectory propagation new hits informations are included in trajectory description using an outside-in reconstruction, as well as the knowledge of $\vec{B}$ field and detector material. The process is stopped when the innermost compatible layer of muon detectors is reached.
Since this procedure may give rise to a number of trajectory that may share the same hits, a "cleaning" process resolves all the ambiguities, keeping a number of track candidates. Finally the "trajectory smoothing" takes over any remaining trajectories and perform a backward fitting using the the covariance matrices that have been used in the previous intermediate measurements. The trajectory is extrapolated to the point of closest approach to the beam line in the transverse plane and in order to improve the momentum resolution, a constraint on the nominal IP is imposed. A Stand-Alone Muon is thus reconstructed.
Since CMS has been conceived to detect muons over a wide range of energies and a large acceptance, the use of all the information coming from several types of sub-detectors is desirable in order to obtain an accurate description of it. Global Muon is the reconstructed entity which merges muon-system and inner tracker information, this latter used also to build Tracker Muon candidates and being more accurate just for low momentum particles ($p_T < 5$ GeV/c).
Track matching usually seek for a suitable region of interest defined on $\eta - \phi$ plane, which contains a set of tracker tracks that could match stand-alone muons ones. These tracks are built inside-out from hit pairs coming from different pixel or strip layers. Even in this case a Kalman-Filter strategy is applied to reconstruct tracks using all the available information.
Then, more and more stringent constraints are applied to choose the best tracker track where stand-alone muon should have come from.
A final fit of tracker and stand-alone muon tracks definitely define the global muon, looping over all the possible track pairs. The global muon track with the best χ^2 is retained, such that for each stand-alone muon just one global muon is retained.
The majority of muons are constructed either as Global Muon or Tracker Muons. Just 1% of muons are reconstructed as Standalone-muon tracks only.

Muon Identification Variables

A number of variables can be used for muon identification and quality selection in muon analysis, which have been described and reported in detail in [104]. Some of them are

- The number of track segments built from hits in muon chambers with the

inner track extrapolation; such a quantity can be useful to reject muon from light flavors decays;

- The transverse impact parameter d_0 in $x-y$ plane, defined as the distance between the point of closest approach to the beamline and the beamline itself[1]. The transverse impact parameter distribution tails are dominated by pion and kaon decays in flight. (A longitudinal impact parameter can be also defined as the z-coordinate of the point of closest approach along the trajectory).

- The number of valid hits of muon hits both in tracker and muons system (according to the tightness of the selection performed);

- The χ^2 of the track fit both for the silicon tracker tracks or for global track;

- The *combined isolation variable*, which is able to distinguish prompt muons from non-prompt muons, for example coming from jets. Such a variable is calculated building a cone around muon trajectory, with radius $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$. A standard value for ΔR is 0.3. The scalar sum of tracks p_T inside the cone as well as the the energy loss inside ECAL and HCAL is calculated, excluding contribution from the candidate itself. The relative isolation is defined as the ratio of the total energy inside the cone to the transverse momentum of the candidate.
  In fact a more refined isolation variable based on reconstructed particles, named *Particle Flow Isolation variable*, can be used to identify good muon candidates. Such a variable will be defined more extensively in Section 4.4.1.

3.3.2 Electron reconstruction

The "ECAL driven" electron reconstruction starts by the reconstruction of ECAL "superclusters" of transverse energy with $E_T > 4$ GeV. "Superclusters" are clusters of adjacent ECAL cells, whose energy deposits are compatible with those expected from an electron shower and are matched to track seeds (pairs or triplets of hits) in the inner tracker layers. From these track seeds, electron tracks are built.
When a shower develops into the electromagnetic calorimeter, it can involve several neighboring crystals. Also bremsstrahlung can be found on ECAL cells at small phi angles with respect to the electron's one. Crystals are therefore clustered together, in a different way according to the geometric properties of the detector region.

[1]The transverse impact parameter can be defined with respect to the beam spot (namely the the luminous region produced by the collisions of proton beams) or the primary vertex of interaction.

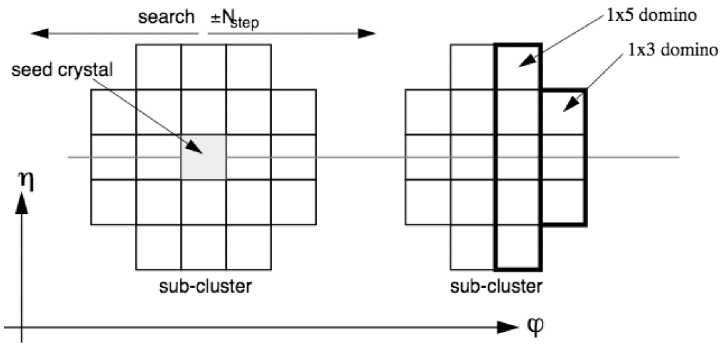


Figure 3.2: An illustration of the Hybrid clustering algorithm used in the ECAL barrel region is given. The algorithm starts exploring a 3×1 "domino" of crystals in $\eta - \phi$; then moves towards the seed, exploring crystals symmetrically around it. All the remaining dominoes are grouped around local energy maxima connected in ϕ direction. If dominoes belonging to any local maxima with highest energy below E_{seed} are dropped. More details can be found in [102].

In the barrel region, the so-called "Hybrid-algorithm" is performed, which exploit the $\eta-\phi$ geometry of crystals and collect energy from bremsstrahlung-free and bremsstrahlung-accompanied electron-like showers. The region around a "seed crystal" is explored symmetrically along the ϕ coordinate. An hybrid supercluster made up of a series of sliced showers at constant η but spread in ϕ-direction is produced (Figure 3.2). The entire energy deposit is well contained in clusters of 5×5 adjacent crystals. More details about this method are given in [102].
In the endcap region a similar idea is implemented with some differences, since crystals are not arranged in a $\eta - \phi$ geometry. The Multi5×5 algorithm is performed. Energy released in the Preshower detector is just added to each endcap supercluster before applying any other correction. Endcap uses just endcap hits, barrel cluster just barrel hits, except for clusters that are found within a maximum distance from each other. Such process is called "superclustering".
Further corrections must be applied to control energy scales in ECAL. Such corrections keep track of energy losses or electron interaction inside the surrounding material (like the tracker material) giving rise to bremsstrahlung or photon conversion.

In the first filtering step of the so-called ECAL-driven seeding, supercluster are matched to track seeds, which are composed by pairs or triplets of seeds in the inner tracker layers.
Actually electron reconstruction takes advantage from the fact that supercluster energy-weighted position lies on the helix that would be followed by the electron in the absence of bremsstrahlung. Therefore trajectories can be backpropagated through the magnetic field and once a hit is matched on the first

tracker layer, helix parameters are recalculated using this information; then a further hit is searched in the second layer using a smaller window. Trajectories are constructed using a dedicate modeling of the electron energy loss and are fitted with a Gaussian Sum Filter (GSF) [107], which is used for the estimation of track parameters and to find the best track.
The filtering is complemented by a loose preselection, based on the matching between GSF tracks and the supercluster in η and ϕ, in order to reduce the probability to reconstruct jets as fake electrons and at the same time not to affect analysis level selection criteria [107].

Electron identification variables

Electron selection variables can be grouped in three categories: identification, selection and conversion rejection variables. One may decide to cut on the upper value of such variables, being each threshold previously tuned to provide standard Electron IDs with different robustness.
Identification variables are usually intended to discriminate between real and fake electrons. In 2012 selections were

- H/E where H is the energy deposited in the HCAL towers in a cone of radius Δ R = 0.15 centered on the electromagnetic supercluster position, while E represents the energy of electromagnetic supercluster
- $\Delta\eta_{in} = \eta_{sc} - \eta_{in}^{extrap}$ is the energy-weighted centroid position in η of the supercluster, while η_{in}^{extrap} is the associated track pseudorapidity at ECAL surface as extrapolated from the innermost track layer
- $\Delta\phi_{in} = \phi_{sc} - \phi_{in}^{extrap}$ is the energy-weighted centroid position in ϕ of the supercluster, while ϕ_{in}^{extrap} is the associated track pseudorapidity at ECAL surface as extrapolated from the innermost track layer
- $\sigma_{i\eta i\eta} = \sqrt{\sum_i^{5\times5} \omega_i(\eta_i - \bar{\eta}_{5\times5})^2 / \sum_i^{5\times5} \omega_i}$ where the index i runs over all the crystals in a 5×5 block of crystals centered on the seed crystal, η_i is the η position of i^{th} crystal, $\bar{\eta}_{5\times5}$ is the energy weighted mean η of the 5×5 block of crystals and ω_i is the weight of the i^{th} crystal and is defined as $\omega_i = 4.7 + \ln(E_i/E_{5\times5})$, where E_i and $E_{5\times5}$ are the energy of the i^{th} and 5×5 block of crystals respectively.
- The transverse and longitudinal impact parameter (d_{xy} and d_z)
- The $|1/E - 1/p|$ being E and p respectively supercluster energy and the track momentum.

Isolation variables are usually defined as a sum of transverse energy or momentum measured by detectors, in a cone on the $\eta - \phi$ plane. The definition and the procedure to compute the isolation variables have been already explained in the previous Section (3.3.1).

Conversion variables are used to identify electrons coming from photon conversion. Due to the amount of material in the CMS tracking system, a sizeable fraction of γ photons converts in electron pairs.
This undesirable background can be reduced by applying a selection based on characteristic topology of converted photons. These are the variables used:

- Missing hits, namely the number of missing electron hits in the innermost tracker layers. Actually electron conversion takes place inside tracker layers and is characterized by a vertex displaced with respect to the primary one. Extrapolating backward electron track, one can check if any hits are missing on active detector layers and, in this case, reject the electron candidate.

- $\Delta \cot\theta = \cot(\theta_{partner} - \theta_{elec})$, being θ_{elec} the direction of electron candidate track and $\theta_{partner}$ the direction of the closest partner track, chosen among all the tracks with opposite charge in a cone with $\Delta R < 0.3$ around the electron candidate. A cutoff threshold may be set considering that electrons coming from photon conversion remain almost parallel each other at the conversion point, and remain parallel in the $r-z$ plane.

- $D(\text{elec}, \text{partner})$ is the distance between the electron track and its closest partner track, namely the distance between the trajectory points where the electron and the partner track directions are parallel each other. Electron produced from photon conversion are characterized by a small D value. Hence a cutoff threshold can be set.

3.3.3 Particle flow

Particle flow algorithm aims to reconstruct all the stable particles in the event, namely electron, muons, photons, charged and neutral hadrons, combining the information from all the CMS subdetectors. It's worth to note that most of stable constituents have usually low p_T values (as the final products of exotic particle decay chains): thus an accurate, efficient and low-fake rate reconstruction must be performed. The list of particles is then used to build jets, to determine missing transverse energy E_T^{miss}, to reconstruct and identify taus from their decay products, to give an estimate of lepton isolation, to b-tag jets etc.
The information of the basic reconstruction objects are combined and linked through a linking algorithm to form physical objects. Particle Flow algorithm is roughly composed by three steps

- **Iterative tracking:** it is based on the information coming from the tracker detector, which is able to provide an accurate measurement of charged particle direction at the production vertex. Track are first reconstructed with very tight requirements, with a moderate efficiency but also a very low fake rate. Once a tight trajectory has been built, assigned

hits are removed, seeding criteria are loosened. Combinatorics is thus reduced and fake rate is kept low. In the first three iterations, 99.5% of isolated muons and 90% of charged hadrons are identified. In the subsequent iterations, a relaxed constraint on the origin vertex is chosen to allow the reconstruction of secondary charged particles, take origin from photon conversion or nuclear interactions in the tracker material.

- **Calorimeter clustering:** this step aims to four objectives: to detect and measure neutral particle; to separate them from charged hadrons; to reconstruct and identify electrons and bremsstrahlung radiation; help energy measurements of charged hadrons, and low quality or high-p_T tracks. A specific clustering algorithm is performed separately on different components of ECAL, HCAL and PS. Clustering starts from an energy maxima in calorimeter cell (seed); cells with common side and with a signal two standard deviations above the electronic noise are then aggregated in "topological clusters", which give rise to as many "particle-flow clusters" as seeds. An iterative procedure determines position and energy of clusters.

- **Link algorithm**: it connects each elements to fully reconstruct objects and single particles avoiding double counting, providing, for each pair of elements, a "distance" which quantify the quality of the link. Some "blocks" containing two or three elements are produced, being the base of particle reconstruction and identification. Links can connect charge-particle track and calorimeter clusters, or two calorimeter clusters, charged particle tracks in the tracker and muon track in muon system. More details about link algorithm can be found in [106].

Once blocks have been built, particle flow algorithm perform the reconstruction and identification step of the muon, electrons and all the remaining tracks, making available a full event description for the analysis.

Particle Flow Performance

Particle flow algorithm is able to provide a substantial improvement both in jet reconstruction and resolution.
In Figure 3.3 (top) a comparison between matching efficiency for Calo-Jets and PF jets is shown. For Calo-Jets just calorimeter information has been used. In both cases, simulated samples have been used to match reconstructed jets (using iterative-cone algorithm) with stable particle jets at generator level. Matching process requires a minimum distance ΔR of 0.1 in the (η, ϕ) plane. For PF jets greater than 20 GeV, the plateau is reached at 40 GeV. In fact an improvement in the agreement between Calo and PF jets can be obtained moving from $\Delta R = 0.1$ to $\Delta R = 0.2$. However, PF algorithm is able to identify jets as small as 5 GeV/c, whereas at low p_T low calorimeter energy threshold

affects jet reconstruction.
Distributions of the relative difference between the transverse momenta of reconstructed jets (p_T^{reco}) and generated jets (p_T^{gen}) are also shown in Figure 3.3 (bottom). Jet energy scale in Particle Flow has been shown to be closer to the unit, requiring smaller residual corrections which makes reconstructed energy matching the generated one.

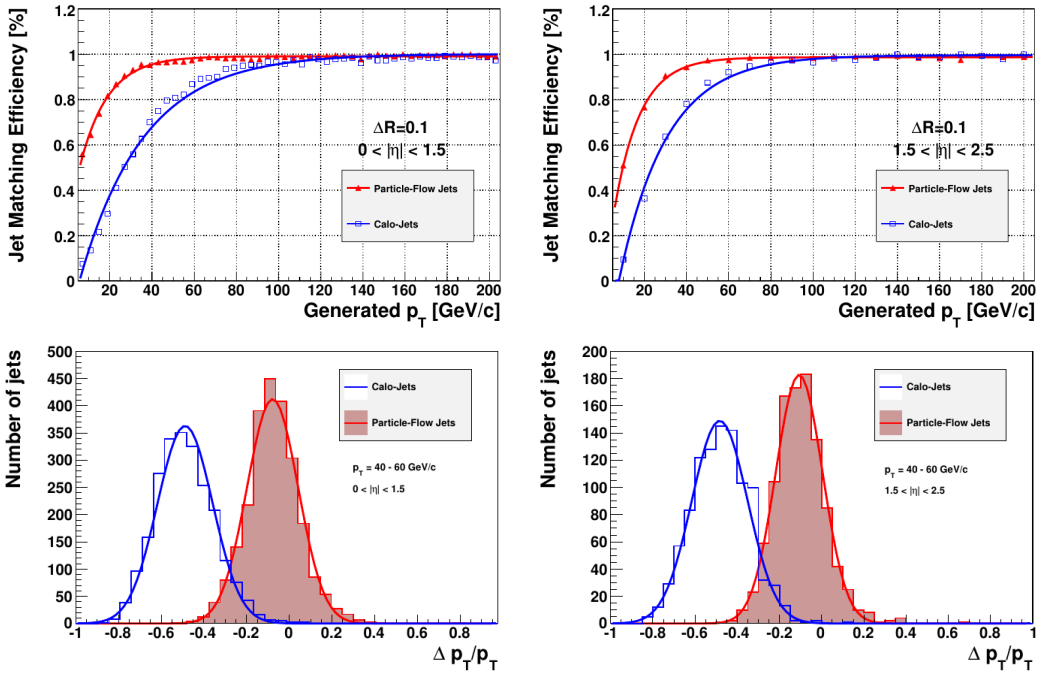


Figure 3.3: Performance of Particle Flow algorithm. Top: jet matching efficiency as obtained for Calo-jets (open squares) and particle-flow jets (triangle) in the barrel, with a matching distance of 0.2; efficiencies and fake rates are fitted to exponential functions of p_T. Bottom: Distribution of $((p_T^{rec} - p_T^{gen})/p_T^{gen}$ for p_T^{gen} between 40 and 60 GeV/c. A gaussian is fit to all distributions, to determine the response and the resolution. [106]

3.3.4 Jet reconstruction

Anti-k_t algorithm

Jets represents the signature of quarks and gluons emissions, which hadronize and give rise to a number of hadrons as a consequence of quark confinement predicted by QCD. Hadrons fly in the same direction of the parton object which they are generated from and release their energy mainly in ECAL and HCAL cells. A calotower, namely the combination of consecutive ECAL and HCAL cells, define jet energy in the $\eta - \phi$ plane.
In this analysis the clustering algorithm deputed to jet reconstruction is the

anti-k_T with cone size 0.5, which in few years has become the most used algorithm in CMS analysis. Energy reconstruction and calibration has been performed combining subdetectors information through a PF algorithm.
Anti-k_T algorithm is based on the generalization of Cambridge/Aachen algorithms. A distance measure d_{ij} is defined:

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2} \tag{3.4}$$

$$d_{iB} = k_{ti}^{2p} \tag{3.5}$$

where $\Delta^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and k_{ti}, y_i and ϕ_j are respectively the transverse momentum, rapidity and azimuthal angle of particle i. R is a radius parameter (set to 0.5 in AK5 jets), while p govern the relative power of the energy versus geometrical scale Δ_{ij}. Anti-k_T algorithm holds for $p = -1$.
The algorithm just loops over all the involved towers, looking for the entities (particles or pseudojets) which minimize d_{ij}.
If there are no other hard particles within a distance of 2R, it just accumulates soft particles around itself; if there are hard particles within a distance $R < \Delta_{12} < 2R$ two jets can be produced, even though at least one of the two will not be perfectly conical; if $\Delta_{12} < R$ the two hard particles cluster in a single jet. After each iteration are progressively removed from the main entity collection, until all entities are clustered.
It can be shown that soft particles tend to cluster with hard ones before clustering among themselves, due to the lower value that d_{ij} assumes in presence of hard radiation. Therefore soft particles do not change significantly jet shape, while hard ones do, making this algorithm quite stable also in pileup environments. More details about anti-k_T algorithm can be found in [80].

b-tagging

b-jets identification is a critical feature for many high energy processes. In CMS it is based on an algorithm which provide a discriminating variable , namely a number which can be used to select different regions, providing a compromise between efficiency and purity.
Such an algorithm is called "tagger" and can take into account quantities like IP significance d_0/σ (being σ the uncertainty on d_0), or output of likelihood ratio or multivariate analysis.

For instance the "simple secondary vertex algorithm" is an high-purity algorithm based on the reconstruction of at least one secondary vertex. It starts from the assumption that $b-$quarks, may fly for a significant distance before hadronizing, significantly displacing jet vertex from IP. For B hadrons with finite lifetime, the typical scale for IP is $c\tau \simeq 480\mu m$. Thanks to the high resolution provided by the pixel system, a 3D vertex reconstruction can be performed. However, such an approach limits the b-jet efficiency to the probability of finding a vertex when a B hadron decays (about 60-70%). For this

reason, several observables may be used in addition (like IP significance, track-by-track probability etc.). More information can be found in [86].
In this analysis a "combined secondary vertex" algorithm has been used instead. It is a sophisticated tag which exploits all the known variables, in particular IP significance, secondary vertex information and jet kinematics. Thus even if no secondary vertices are found, a discrimination value is provided. These variables are then used as input for a likelihood ratio and used twice to discriminate between b-jet and c-jet; a weighted sum is performed and then combined in a unique number. More details can be found in [86].

Jet energy scales

The precise measurement of absolute jet energy, however, is not an easy task: a number of sources may contribute to modify jet energy with respect to the original value. Geometry and calorimeter inefficiencies or miscalibrations, radiation losses by hadronic calorimeters may lower jet energy measurements, while radiative activity due to the presence of pileup environments for example may contribute to overestimate it. Jet energy scale (JES) in simulation has to be corrected to fix these unwanted effects, and some jet energy corrections must be applied. Such corrections are applied sequentially and with a fixed order, each one taking care of a different effect. Three corrections are mandatory to perform a jet-based analysis: L1 corrections, aiming to remove energy deposited by jets deriving from pileup activity; L2 relative corrections, which make the jet response flat as a function of pseudorapidity; L3 absolute corrections, which remove any dependence on p_T in jet response. A further correction is applied to remove a small difference (about 10%) depending on pseudorapidity between data and simulation. A detailed description of energy corrections can be found in [99], [100].

3.3.5 Missing transverse energy reconstruction

Missing Transverse Energy reconstruction (E_T^{miss}) is defined as

$$\vec{E_T^{miss}} = -\sum_n (E_n \sin\theta_n \cos\phi_n \hat{\mathbf{i}} + E_n \sin\theta_n \sin\phi_n \hat{\mathbf{j}}) = \\ = \not{E}_T^{miss,X} \hat{\mathbf{i}} + \not{E}_T^{miss,Y} \hat{\mathbf{j}} \tag{3.6}$$

where the index n runs over all the input object (energy deposits in towers, hits, generator-level particles, unclustered objects...), and $\hat{i}$, $\hat{j}$ are the unit vectors in the direction of x and y axis [109], where z points towards the direction of the beam. If no physical sources of missing energy are present, both $\not{E}_T^{miss,X}$ and $\not{E}_T^{miss,Y}$ components are expected to be distributed according to a gaussian centered at zero, with a standard deviations σ, function of detector features. In fact, resolution on E_T^{miss} can be written as a function of different contributions:

$$\sigma(E_T^{miss}) = A \bigoplus B\sqrt{E_T - D} \bigoplus C \cdot (E_T - D) \tag{3.7}$$

where A stands for noise contributions (pile-up effects, underlying events, electronic noise), B represents a stochastic term due to the way particle energy in calorimeters is sampled, C accounts for constant contributions due to non-linearities, cracks and dead material and D is a generic offset.
In case of a QCD dijet event, where no physical E_T^{miss} (neutrinos) is expected, an energy unbalance can be just given to detector finite resolution or acceptance.
CMS uses different kinds of MET reconstruction: the most fundamental one, provides the so-called CaloMET, namely E_T^{miss} calculated from all energy deposits in calorimeter towers in EB, EE, HB, HE, HF.
Reconstructed particle using Particle Flow algorithm can be used as well to determine E_T^{miss} (pfMET), including track information. In some sense, calculating E_T^{miss} from Particle Flow is something analogous to use generator-level particles, except for the fact that reconstructed particles are effectively used.
In general missing transverse energy should be corrected by various effects: the most basic corrections, named *Type I*, embed L1,L2,L3 jet corrections in E_T^{miss} computation, aiming to remove dependence on p_T, η and non-linearity in jet energy scales. Therefore E_T^{miss} can be written as:

$$\vec{\not{E}}_T^{miss,TypeI} = \vec{E_T^{miss}}^{raw} - \sum_{Njets} [\vec{p_T i}^{L123} - \vec{p_T i}^{raw}] \tag{3.8}$$

being $\vec{p_T}^{L123}$ the full-corrected jet transverse momenta.
More specifically uncorrected Particle Flow E_T^{miss} can be written as the negative vectorial sum of the transverse momentum of jets (classified in two classes, with a threshold at 10 GeV/c) and all the other reconstructed objects:

$$\vec{E}_{\mathrm{T}}^{\mathrm{uncorr.}} = - \sum_{\mathrm{jet}\vec{p}_{\mathrm{T,\ jet}}>10\mathrm{GeV}} \vec{p}_{\mathrm{T,\ jet}}^{\mathrm{uncorr.}} - \sum_{\mathrm{jet}\vec{p}_{\mathrm{T,\ jet}}\leq 10\mathrm{GeV}} \vec{p}_{\mathrm{T,\ jet}}^{\mathrm{uncorr.}} - \sum_{i\notin\mathrm{jets}} \vec{p}_{\mathrm{T,\ i}}. \tag{3.9}$$

When Type-I corrections are used, which are applied to jets having $p_T > 10$ GeV/c, Equation 3.10 reads:

$$\vec{E}_{\mathrm{T}}^{\mathrm{uncorr.}} = - \sum_{\mathrm{jet}\vec{p}_{\mathrm{T,\ jet}}>10\mathrm{GeV}} \vec{p}_{\mathrm{T,\ jet}}^{\mathrm{L123}} - \sum_{\mathrm{jet}\vec{p}_{\mathrm{T,\ jet}}\leq 10\mathrm{GeV}} \vec{p}_{\mathrm{T,\ jet}}^{\mathrm{uncorr.}} - \sum_{i\notin\mathrm{jets}} \vec{p}_{\mathrm{T,\ i}}. \tag{3.10}$$

A further correction, called *Type II*, can be also applied to account for unclustered energy and out-of-cones energy deposition, adding a further scale term. However *Type II* corrections have been conceived to correct CaloMET and are not recommended to be used with Particle Flow E_T^{miss}. More information about *Type II* corrections can be found in [108].
Indeed it has been proven that Particle Flow approach brings a significant improvement with respect to the E_T^{miss} reconstructed with only calorimeter information. One of the first comparisons between Particle Flow MET and CaloMET is shown in Figure 3.4 (top), where the transverse energy relative

resolution, that can be read as

$$\frac{\sum E_T^{reco} - \sum E_T^{true}}{\sum E_T^{true}} as \tag{3.11}$$

is reported as function of the "true" total visible missing energy (being E_T^{reco} the reconstructed energy in the event, while E_T^{true} the " "real" one, obtained by the simulation); in Figure 3.4 (bottom) E_T^{miss} resolution, namely $\not{E}_T^{miss,reco} - \not{E}_T^{miss,true}$, accounting for muon and detector acceptance, has been reported for a $t\bar{t}$ selection. More details can be found in [108]

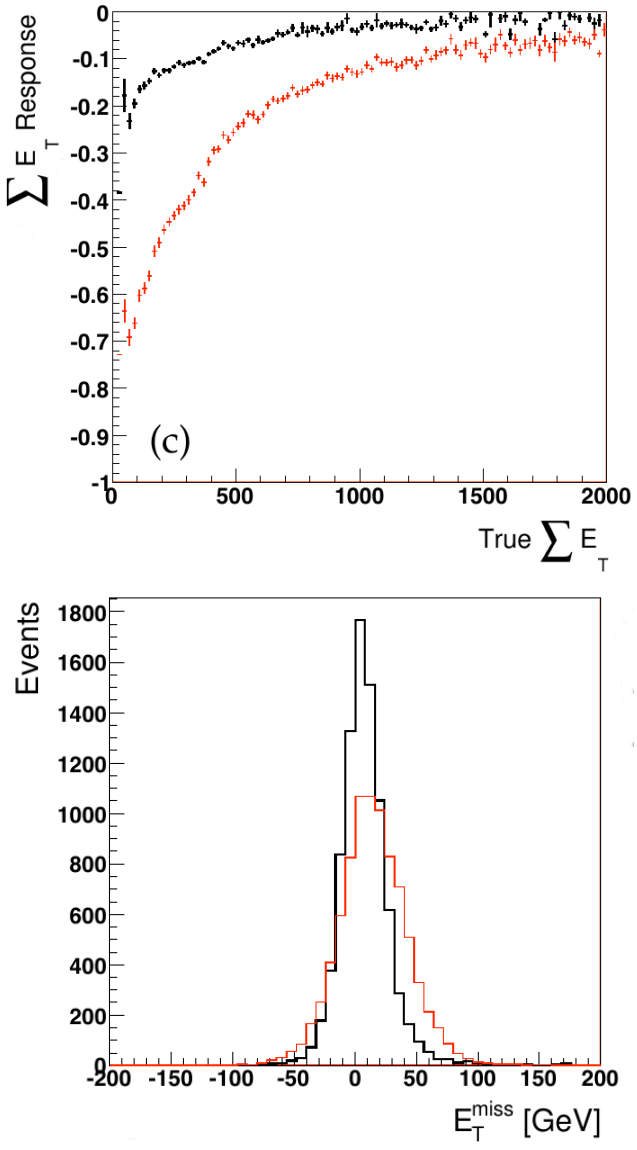


Figure 3.4: Fully simulated QCD dijet events. Top: the transverse energy relative resolution, as a function of the "true" total visible transverse energy of the event: bold black points represent corrected particle flow reconstruction, the red points represent calorimeter based reconstruction (corrected for the calorimeter jet energy scale). Bottom: E_T^{miss} resolution: the black histogram represents corrected particle flow reconstruction, the red histogram represents calorimeter based reconstruction (corrected for the calorimeter jet energy scale and the possible presence of muons.) [108]

Chapter 4

Analysis strategy

4.1 Introduction

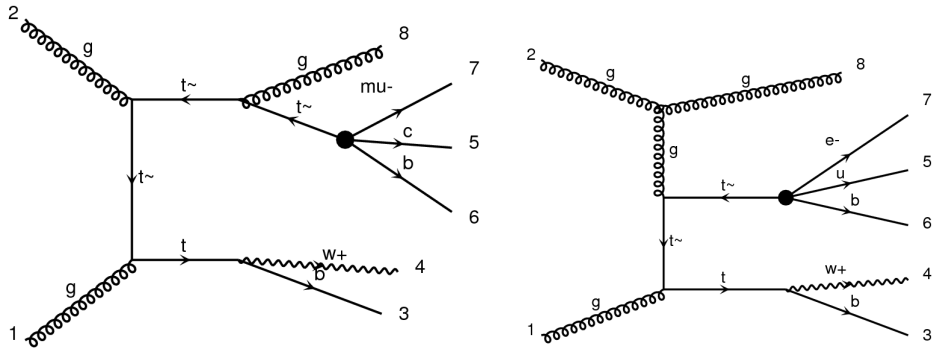
This analysis aims to find evidences of Baryon Number Violating (BNV) top quark decays in one lepton plus five jets topologies. It has been performed using data collected by CMS Experiment during 2011 ($\sqrt{s} = 7$ TeV) [68] and 2012 ($\sqrt{s} = 8$ TeV) [69], corresponding to an integrated luminosity respectively of 5 fb^{-1} and 19.5 fb $^{-1}$. An upper limit on top quark BNV decay Branching Ratio (BR) has been set both in 2011 and 2012, this latter much improving previous results. In fact few differences exist between the 2012 and 2011 analyses: apart the larger integrated luminosity and the higher energy in the center of mass, which give larger yields, in 2012 new triggers have been chosen for the muon analysis; a different method for QCD estimation has been adopted; finally a new approach to deal with systematic uncertainties has been implemented.
Because of the low expected value of BR (much less than 1%), the most suitable final states to look for such decays are those coming from $t\bar{t}$ decays, where one of the two tops decays hadronically, while the other decays in one charged lepton and two jets (in Figure 4.1 the tree level process is shown).
Two analyses have been performed independently, respectively for both muon and electron flavor. Actually muon and electron clear signatures make them suitable for the search of rare processes; on the contrary τ channel has not been taken into account due lower sensitivity and the more complex final state topology.
The most striking feature of such a baryon number violating process is the absence of neutrinos in final state: the lack of physical missing energy E_T^{miss} represents the main difference with a SM $t\bar{t}$ semi-leptonic decay with one extra jet due to initial or final state radiation. A further remarkable difference is given by the charge signature of BNV top decay: one of the two jet in the final state is started by a b-quark with the same charge of the parent top quark. Therefore when one considers the full $t\bar{t}$ final state, two same sign b-jet are actually found, while no constraints are set on the lepton and light jet flavor.

This analysis relies on the effective model described in Section 1.2.3, which substantially makes BNV top decay equivalent to a four-point interaction. Fermion-flavor-dependent parameters a, a'... that appear in Equation have been set to 1, with a negligible impact on event kinematics, while the mass scale Λ has been set to 1 TeV/c^2.
A counting experiment has been performed to simplify the analysis and to make it as model-independent as possible. The simulated events yield in a region (*tight selection*) where the signal presence is enhanced has been compared to data yield; in order to rely as little as possible on simulated signal, before applying a tight selection, MC and data have been normalized in a region where non top-background is suppressed.
The analyses on electron and muon channels have been performed separately, and then combined to extract final results.

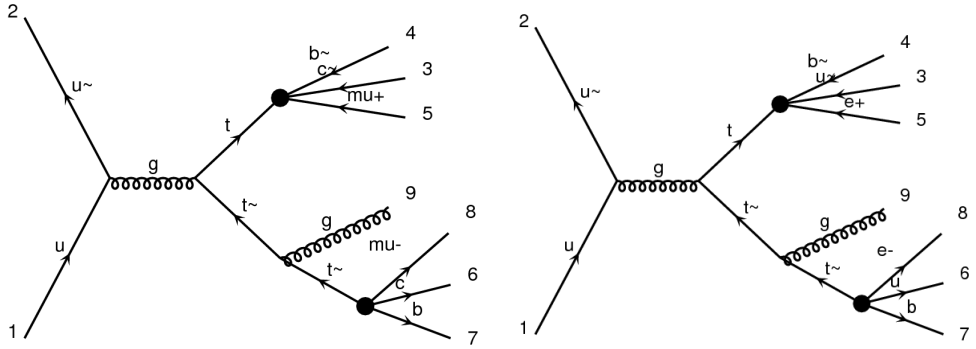
4.2 Datasets and triggers

4.2.1 Datasets and backgrounds

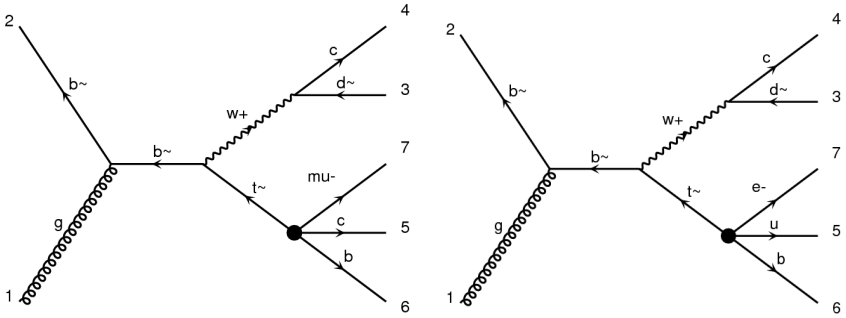
As already mentioned, in this work data gathered by CMS Experiment at LHC during 2011 and 2012 are used, corresponding to an integrated luminosity of 5 ± 0.19 fb^{-1} in 2011 ($\sqrt{s} = 7$ TeV) and 19.52 ± 0.49 fb^{-1} in 2012 ($\sqrt{s} = 8$ TeV).
The main background to the BNV decay process is given by the SM $t\bar{t}$ production with extra jets coming from initial and final state radiation and then merged using "MLM" matching procedure [70] (see Section 3.2.1 for more details). Official CMS simulated samples have been used, imposing $\mathcal{B}(t \to bW)$ equal to 1, and top mass matching 172.5 GeV/c^2.
Single-top production in association with W boson, represents the second most significant background which may contribute in principle to BNV signal detection. Even single-top production in s-channel and t-channel have been considered as well, though their impact is largely subdominant.
The dominant non-top background is given by production of W and Z boson in association with extra-jets (respectively W+jets and Z+jets); a further contribution comes from double-bosons processes like WW, ZZ, WZ and from $t\bar{t}$ in association with vector bosons ($t\bar{t}$W, $t\bar{t}Z$). Finally a non-negligible QCD multi-jets contribution is expected.
Single-top samples have been generated with POWHEG [71] and double bosons events have been produced using Pythia6 generator [72]. W+jets, Z+jets, $t\bar{t}Z$ and $t\bar{t}W$ $t\bar{t}$ have been generated with MadGraph v.5.1.3.30 [73] interfaced with Pythia6 [72] to provide hadronization and parton fragmentation.
Three datasets involving BNV processes have been produced: parton level events have been generated by Madgraph 5.143 and then passed to Pythia 8.165 [74]. The main sample ($t\bar{t}_{BNV\_SM}$) is composed by $t\bar{t}$ events, where one top decays hadronically, while the other undergoes a BNV decay; a dataset



(a) One of the possible baryon number violating diagrams for $t\bar{t}$ pairs, both in muon and electron channel. The first top undergoes a BNV decay, while the second top decays hadronically. Extra jets in initial and final state have been added, to account for ISR and FSR.



(b) One of the possible baryon number violating diagrams for $t\bar{t}$ pairs in muon and electron channel, where both the top quarks undergo BNV decay.



(c) One of the possible baryon number violating single top diagram, produced in association with a W boson.

Figure 4.1: Simulated BNV process in 2011 and 2012 analysis; in 2012 analysis one extra jet has been added to account for initial and final state radiation in $t\bar{t}$ pairs. Due to the low cross section, no extra radiation has been added to Single top production process. Black blobs stands for new physics vertices.

where both top quarks decay in BNV mode ($t\bar{t}_{BNV\_BNV}$) has been generated as well; finally, a single-top dataset has been produced, where a top which decays violating baryon number symmetry is produced in association with a W boson (tW_{BNV}). For the 2012 simulated signal samples, samples with one extra jet have been added to better simulate initial and final state radiation in the $t\bar{t}_{BNV\_SM}$ and $t\bar{t}_{BNV\_BNV}$ processes. Due to the low cross section, related to higher order initial and final state radiation processes and because of the heavy computational effort, no further extra jets have been added.
As suggested in [7], the BNV decay of the (anti-) top quarks is exclusively of the type $t \to \bar{b}\bar{c}\mu^+$ for muon channel, and $t \to \bar{b}\bar{u}e^+$ for electron channel. In Table 4.1 all the simulated samples have been summarized, and process generators have been reported.
After generation, hadronization and parton fragmentation step, the propagation of the particles produced at the interaction vertex has been simulated by Geant4 v9.2 toolkit [76]. Event reconstruction has been performed, exactly as for real data, with the CMSSW software suite, overlaying simulated minimum-bias events to reproduce the pileup measured in data.

Process	**Generator**
$t\bar{t} + Jets$, $W + Jets$, $Z + Jets$ ttW, ttZ	Madgraph5 + Pythia 6
tW, $t-$channel, $s-$channel	Powheg
WW, WZ, ZZ	Pythia6
BNV Signals	Madgraph5 + Pythia 8

Table 4.1: Background and Signal processes used in this analysis; for each MC sample, the generator program used for events production has been reported; when Madgraph5 is used for parton level generation, Pythia provides hadronization and parton-fragmentation.

4.2.2 Triggers

The choice of the triggers is driven by the final state topology. In 2011 isolated single-muon triggers have been chosen, with a p_T threshold ranging between 17 and 30 GeV, depending on the data taking period. In 2012 a further requirement of three jets emitted in the central region ($0 \leq \eta \leq 2.1$) having transverse momentum between 30, 40 and 50 GeV at trigger level, allowed to keep to 17 GeV the muon p_T threshold at trigger level. For the electron channel one isolated electron plus three central jets with a p_T threshold of 30 GeV in 2011 and 30, 40, 50 GeV in 2012 have been required.
A common software emulated trigger threshold has been applied offline to make the trigger selection homogeneous across all periods. The chosen thresh-

Year	Muon Channel Trigger	Electron Channel Trigger
2011	1 Iso. Mu ($p_T > 30$ GeV; $\|\eta\| < 2.1$)	1 Iso. Ele. ($p_T > 25$ GeV) and 3 Jets ($p_T > 30$ GeV; $\eta < 2.4$)
2012	1 Iso. Mu ($p_T > 20$ GeV; $\|\eta\| < 2.1$) and 3 Jets ($p_T > 30$,40,50 GeV; $\|\eta\| < 2.4$)	1 Iso. Ele. ($p_T > 25$ GeV; $\|\eta\| < 2.5$) and 3 Jets ($p_T > 30$,40,50 GeV; $\|\eta\| < 2.4$)

Table 4.2: Tightest trigger thresholds applied for electron and muon analysis; these cuts match the tighter online trigger requirements respectively imposed in electron and muon analysis during 2011 and 2012 data taking period.

old value is equal to the highest trigger threshold among those used in either the 2011 or 2012 runs. In Table 4.2 are summarized the trigger thresholds adopted for 2011 and 2012 analysis.

4.3 Analysis strategy

A counting experiment has been performed to compare SM yield in a phase space region where the presence of BNV signal is enhanced (tight selection). A significant excess of data over MC would be interpreted as an hint of new physics, possibly compatible with existence of top quark BNV decay.
In principle the BNV decay could manifest itself also in $t\bar{t}V$ and single-top decay. However, due to the tiny branching ratio expected for the top quark BNV decay, just BNV decay contributions from tW and $t\bar{t}$ events have been taken into account to compute the expected signal yield. Actually a different strategy would have made the analysis much more complex. The BNV contribution from single top production $t-$channel and $s-$channel, as well as for $t\bar{t}V$ events has been neglected.
The analysis selection starts targeting a control region, where the presence of non-top background is relatively suppressed. In this region the event yield was dominated by $t\bar{t}$ and W+jets process, the total uncertainty coming primarily from cross section and luminosity systematic contributions, in addition to the selection efficiency in basic. On top of this "basic selection" a further "tight" selection is applied, defining a new region, where the presence of the signal is strongly enhanced.
The total expected yield in "tight" according to a classical approach could be written:

$$\begin{aligned} N^T_{exp} &= N^T_{top} + N^T_{bck} = (N^T_{t\bar{t}} + N^T_{tW}) + N^T_{bck} = \\ &= L\left[\sigma_{t\bar{t}}\epsilon^{(T)}_{t\bar{t}}(BR) + \sigma_{tW}\epsilon^{(T)}_{tW}(BR)\right] + N^T_{bck} \end{aligned} \tag{4.1}$$

being N^T_{top} the expected top yield, resulting by the sum of $t\bar{t}$ and tW contributions; L is the integrated luminosity; $\sigma_{t\bar{t}}$ (σ_{tW}) the production cross section

of $t\bar{t}$ (tW) events; N^T_{top} the number of top-like events and N^T_{bck} the number of background events; $\epsilon^T_{t\bar{t}}$ (ϵ^T_{tW}) are respectively efficiencies to pass tight selection for $t\bar{t}$ (tW) events.
For a generic event ϵ^T depends on the processes which effectively takes place. For a $t\bar{t}$ event such an efficiency can be written as

$$\epsilon^T_{t\bar{t}} = 2BR(1-BR)\epsilon^T_{BNV,SM} + (1-BR)^2\epsilon^T_{SM,SM} + BR^2\epsilon^T_{BNV,BNV} \tag{4.2}$$

while for tW channel one can write:

$$\epsilon^T_{tW} = (1-BR)\epsilon^T_{SM} + BR\epsilon^T_{BNV} \tag{4.3}$$

Equation 4.2 (Equation 4.3) accounts for all the possible $t\bar{t}$ (tW) decay modes, each efficiency being weighted by its branching ratio; T stands for "tight selection".
Following the classical approach, this estimate would be affected by the uncertainties due to the SM $t\bar{t}$ and tW cross sections (which represent the main background for this signal, being non top background subdominant) and by the full uncertainty on integrated luminosity.
In fact the impact of systematic uncertainties and the uncertainty in the selection efficiency can be reduced if the top yield in Tight N^T_{top} is expressed as a function of the yield in Basic. First of all the top yield in basic N^B_{top} has been written as the difference between the total number of observed events in Basic as N^B_{obs} and the number of expected non-top background events as N^B_{bck}:

$$N^B_{top} = N^B_{obs} - N^B_{bck} \tag{4.4}$$

Actually Equation 4.4 assumes that simulation is normalized to data.
If one writes the probability for a $t\bar{t}$ (tW) event to pass tight selection once basic selection has been passed as $\epsilon^{(T|B)}_{t\bar{t}}$ ($\epsilon^{(T|B)}_{tW}$), the total number of expected events N^T_{exp} in tight reads:

$$\begin{aligned} N^T_{exp} = N^T_{top} + N^T_{bck} = \left(N^B_{obs} - N^B_{bck}\right) \times \\ \times \left[\frac{N^B_{t\bar{t}}}{N^B_{t\bar{t}} + N^B_{tW}} \times \epsilon^{(T|B)}_{t\bar{t}} + \frac{N^B_{tW}}{N^B_{tW} + N^B_{t\bar{t}}} \times \epsilon^{(T|B)}_{tW}\right] + N^T_{bck} \end{aligned} \tag{4.5}$$

$N^B_{t\bar{t}}$ (N^B_{tW}) standing for the number of $t\bar{t}$ (tW) events in Basic selection.
All the quantities that appear in Equation 4.5 inside square brackets in principle depends on BR, on the total production cross sections for $t\bar{t}$ and tW (respectively $\sigma_{t\bar{t}}$ and σ_{tW}) and on the Luminosity L. It's easy to show that Luminosity cancels out and Equation 4.5 turns out to be

$$\begin{aligned} N^T_{exp} = \left(N^B_{obs} - N^B_{bck}\right) \times \\ \times \left[\frac{1}{1 + \frac{\sigma_{tW}\epsilon^B_{tW}(BR)}{\sigma_{t\bar{t}}\epsilon^B_{t\bar{t}}(BR)}} \times \frac{\epsilon^T_{t\bar{t}}(BR)}{\epsilon^B_{t\bar{t}}(BR)} + \frac{1}{1 + \frac{\sigma_{t\bar{t}}\epsilon^B_{t\bar{t}}(BR)}{\sigma_{tW}\epsilon^B_{tW}(BR)}} \times \frac{\epsilon^T_{tW}(BR)}{\epsilon^B_{tW}(BR)}\right] + N^T_{bck} \end{aligned} \tag{4.6}$$

being $\epsilon^B_{t\bar{t}}$ (ϵ^B_{tW}) the efficiencies of the basic selection for $t\bar{t}$ (tW) events. These quantities are functions of more fundamental efficiency values, each corresponding to a decay mode:

$$\epsilon^B_{t\bar{t}} = 2BR(1-BR)\epsilon^B_{BNV,SM} + (1-BR)^2\epsilon^B_{SM,SM} + BR^2\epsilon^B_{BNV,BNV} \tag{4.7}$$

$$\epsilon^B_{tW} = (1-BR)\epsilon^B_{SM} + BR\epsilon^B_{BNV} \tag{4.8}$$

where B stands for basic selection.
Some observations about Equation 4.6 can be done:

- In the first term in square bracket the dependence on the $t\bar{t}$ cross section is confined to a term $\ll 1$, added to 1.
- The second term in square bracket is indeed affected by the $t\bar{t}$ cross section uncertainty; however the total contribution of the second term, which represents the tW contribution, represents less than 5% of the contribution of $t\bar{t}$ events in tight.
- Instead of plain efficiency terms for passing the tight selection in Equation 4.6 just simple efficiency ratios appear, such that most of correlated uncertainties cancel out.
- The impact of integrated luminosity is confined to non-top background, whose contribution is subdominant.
- Non-top background cross section (e.g. W+jets) in a classical approach (Equation 4.1) would have been the dominant source of systematic uncertainties. In this case their impact is largely reduced, because - as can be seen in Equation 4.6 - N^T_{bck} and N^B_{bck}, which have uncertainties that are highly and positively correlated, appear with opposite sign. Thus an increment of N^T_{bck} causes a reduction of N^B_{top}: therefore the overall expected yield in Equation 4.6 remains relatively stable
- From Equation 4.6, 4.8 and 4.7, due to the small BR value, one can easily note that analysis is mostly sensitive to uncertainties in the ratio of $\epsilon^T_{SM,SM}$ to $\epsilon^B_{SM,SM}$, N^B_{bck}, and N^T_{bck}.

4.4 Event selection

As already mentioned the main background for this analysis is given by SM $t\bar{t}$ events. As shown in Equation 4.6, the yield of top background in basic is given by the subtraction of non-top background from the total observed yield, while in tight the number of events is computed taking into account the efficiency of passing the tight selection, once basic has been passed. The efficiency values

are calculated from the simulation, while the next-to-leading order (NLO) $t\bar{t}$ and tW cross sections are used to compute the ratio $\sigma_{t\bar{t}}/\sigma_{tW}$ in Equation 4.6. The second background is given by W and Z production in association with jets, whose cross sections are taken for theoretical predictions for W+jets$\rightarrow l\nu$+jets and $Z \rightarrow l\bar{l}$+jets processes, computed by FEWZ at next-to-next-to-leading order (NNLO) [77]. The yields in basic and in tight selection are computed according to the simulation and using the measured value of the integrated luminosity.
All the contributions in tight and basic selection coming from single top production via $s-$channel and $t-$channel, WW, WZ, ZZ, $t\bar{t}W$ and $t\bar{t}Z$ are evaluated from simulations as well. The cross section value for single top production via s-channel has been obtained from next-to-next-to-leading-logarithm resummation of soft and collinear gluon corrections Reference. In all the other cases NLO theoretical predictions, as obtained from MCFM, have been adopted. Event object reconstruction has been performed using the particle flow (PF) algorithm [106]. For each events, the following quality requirements have been applied:

- For events with at least 10 tracks, at least 25% of high purity tracks[1] has been required.

- Events with high calorimeter noise in the HCAL barrel or endcap have been discarded.

- Events have been required to have at least one good reconstructed primary vertex; among all the reconstructed vertices, the one with the highest associated $\sum p_T^2$, is assumed to be related to the hard scattering vertex, and all the reconstructed object of the analysis must be consistent with it.

Due to the high luminosity of 2011 and 2012 running conditions, more than one proton-proton interactions per bunch are expected, giving rise to the so-called pileup interactions. Such interactions are usually simulated as multiple minimum bias events, overlaid to the hard scattering interaction.
The pileup activity cause an increase in the number of primary vertices in the events: the expected distribution can be calculated using the information provided by the CMS bunch-by-bunch luminosity measurements with the total pp inelastic cross-section.
The average number of reconstructed vertices during the integrated 2011 data taking period was approximately 9, while during 2012 was 15.

[1]Due to the dense track environment in CMS events, a preliminary track "cleaning" is needed to reduce fake reconstructed tracks: several quality cuts are imposed on tracks, the most important ones being the track fit-χ^2, the longitudinal and transverse impact parameters (and their significance), the number of crossed layers with measurements, track η and p_T. The full *high purity* is described in detail in [78].

4.4.1 Basic selection

As already mentioned, the analysis has been performed independently in muon and electron channels.

Muon selection

Muons must pass a set of offline quality cuts, reported in Appendix A. In 2011, muons are required to have a minimum p_T of 35 GeV/c, with a pseudo-rapidity $|\eta| < 2.1$: such a threshold was set to keep analysis working point in trigger efficiency plateau so that the efficiency of trigger muon selection can be considered constant and independent from p_T; in 2012, being the trigger muon threshold lowered to 20 GeV, also offline p_T requirements decreases to 25 GeV/c. To reject non-prompt muon (cosmics, muons form b hadron decays, etc.) the transverse impact parameter relative to the beam axis is required to be smaller than 0.2 (0.02 in 2011), while the longitudinal position of the muon track at its closest approach to the beam line is required to lie within 1 cm (0.5 in 2011) from the position of the hard scattering vertex.
The isolation of muon candidates is defined according to the particle flow isolation algorithm. In such an approach, a cone with radius ΔR, calculated as

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} \tag{4.9}$$

is defined around the muon track. Inside the cone the energy deposited by charged hadrons (E^{ℓ}_{CH}), neutral hadrons (E^{ℓ}_{NH}) and by gamma photons (E^{ℓ}_{γ}) are computed and a new variable I^{ℓ}_{rel} (relative lepton isolation) is defined:

$$I^{\ell}_{\mathrm{Rel}} = \frac{E^{\ell}_{\mathrm{CH}} + E^{\ell}_{\mathrm{NH}} + E^{\ell}_{\gamma}}{p^{\ell}_{\mathrm{T}} \cdot c} \ , \tag{4.10}$$

Due to the not negligible pileup contribution a correction must be applied to the isolation variable to account for effects of additional interactions. In muon channel the contribution of energy deposited in the isolation cone by charged particles not associated to the primary vertex is first calculated, to correct jet energy from the pileup contribution. This amount is multiplied by the naive ratio of neutral to charged hadron production in the hadronization process of pileup interactions (0.5) and subtracted by the isolation variable. [79].
Corrected isolation is required to be less than 0.12 for muons (Figure 4.5), being $\Delta R = 0.3$.

Electron selection

Electrons must pass a set of tight quality cuts, reported in Appendix A. Coherent sets of cuts have been optimized for 2011 and 2012 data taking periods, which gave slightly different choice for cuts thresholds. Actually in 2012 (2011)

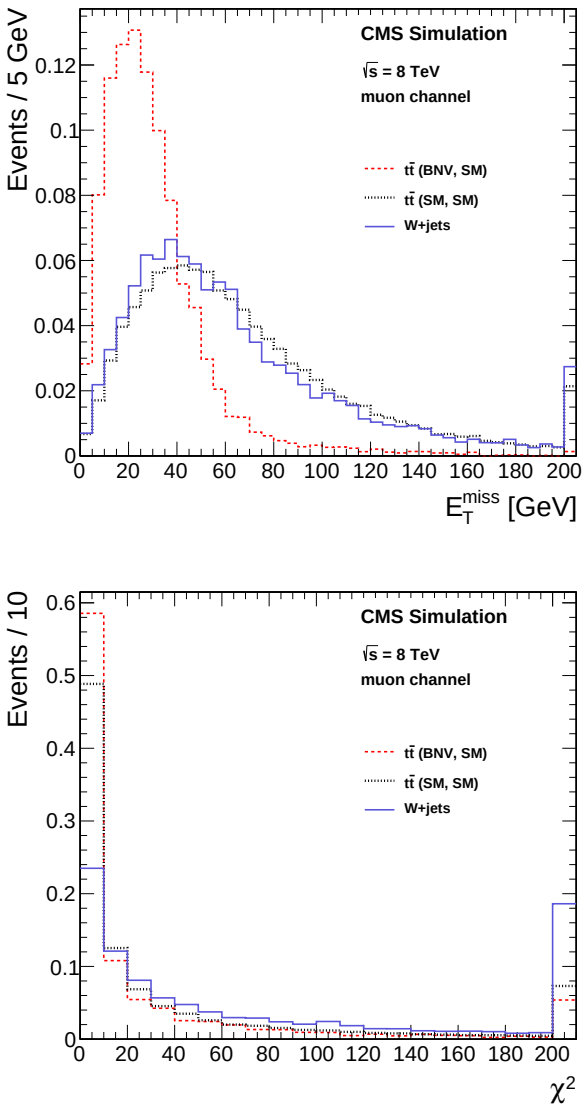


Figure 4.2: Distribution in MC of E_T^{miss} (top) and χ^2 (bottom) variables, as defined in Section 4.4.2, for W+jets, SM-SM $t\bar{t}$ and SM-BNV $t\bar{t}$ events after the basic selection in the muon analysis. All allowed W decays are considered for the top quark SM decays.

electrons have been required to have a transverse momentum greater than 30 (35) GeV/c, $|\eta| < 2.5$ (2.5), with the exclusion region ($1.442 < |\eta| < 1.566$). To reject non-prompt electrons the transverse impact parameter relative to the beam axis is required to be smaller than 0.2 (0.02 in 2011), while the longitudinal position of the electron track at its closest approach to the beam line is required to lie within 1 cm (0.5 in 2011) of longitudinal position of the hard scattering vertex. In addition, electron candidates must be isolated according to particle flow isolation algorithm (defined in Equation 4.10), with $I^{\ell}_{\mathrm{Rel}} < 0.1$

(Figure 4.5 bottom), being $\Delta R = 0.4$. In electron channel isolation variable is corrected, too, to account for pileup contribution: a median energy density (ρ) is determined event by event. The pileup contribution to the jet energy is estimated as the product of ρ and the area of the jet. Such contribution is finally subtracted from the numerator of Equation 4.10 [79].
A photon conversion rejection requirement is also applied (as explained in Section 3.3.2).

Lepton veto

Events with at least one further muon (electron) having $p_T > 10$ (20) GeV/c in $|\eta| < 2.4$ (2.4) and with $I^{\ell}_{\text{Rel}} < 0.1$ are vetoed.

Jet selection

Jet clustering is performed using the anti-k$_T$ algorithm [80] with a size parameter of 0.5. Charged hadrons identified by PF algorithm and associated to pileup activity, isolated muons and electrons are not passed as input of clustering algorithm. Jet energy is corrected using a factorized approach, where each level of correction takes care of different effects; each correction implies the application of a momentum scale factor, which depends on various related jet quantities (p_T, η...), aiming to make energy response function flat as function of jet features and to correct the disagreement in energy resolution observed in data and MC.
Finally, jets energy is also corrected taking into account neutral particle contribution from pile-up collisions (as already mentioned in Section 3.3.4), for which no particle subtraction is performed, due to large uncertainties on the originating primary vertex.
Five jets passing minimal quality cuts are requested for each event, with a p_T greater than 30 GeV/c and $|\eta| < 2.4$. In fact in 2012, a further requirement on p_T threshold has been added, asking for the three leading jet a p_T greater than 70, 55, 40 GeV/c. Actually this choice is motivated by the higher jet p_T threshold at trigger level, setting the working point on the plateau of the jet-trigger efficiency turn-on curve.
Moreover one of the five selected jets must be tagged as "b-quark" initiated jet by the "Combined Secondary Vertex" (see Appendix) algorithm, configured at a the "loose" working point (corresponding to a discriminating threshold of 0.244).

4.4.2 Tight selection

All the cuts listed in Section 4.4.1 defined the so called *basic region* (as explained in Section 4.3). *Tight region* has been conceived to enhance the presence of signal with respect to the expected background (Section 4.3). In fact, the absence of physical missing energy is a fundamental feature of the signal.

In particle flow reconstruction missing energy (E_T^{miss}) is calculated as

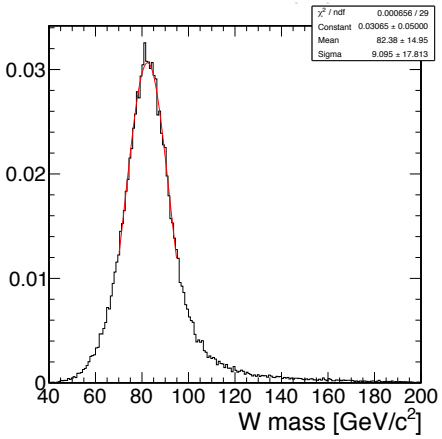
$$\vec{E_T^{miss}} = -\sum_{i=1}^{N} \vec{p_T^i} \tag{4.11}$$

being p_T^i the transverse momentum of i-th reconstructed event object (jets, muons etc.), including unclustered deposits in the calorimeters. Of course, jet energy corrections must be propagated for E_T^{miss} computation. In order to pass tight selection, events are required to have $E_T^{miss} < 20$ GeV (see Figure 4.2 (top) for a comparison among E_T^{miss} distributions for W+jets, SM $t\bar{t}$ and $t\bar{t}_{SM\_BNV}$ decay channels).
An additional requirement must be met to pass the tight selection, which accounts for the compatibility with the expected kinematics of signal events (where one top undergo an hadronic decay and one a BNV decay). A χ^2 variable is defined:

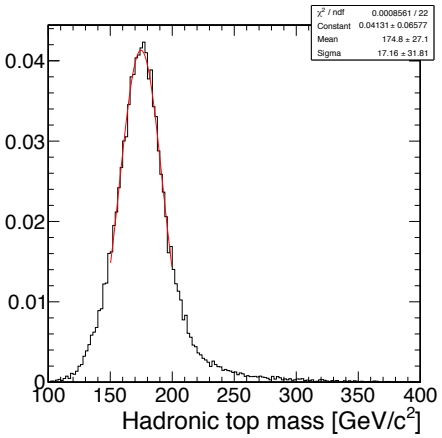
$$\chi^2 = \sum_i \frac{(x_i - \bar{x}_i)^2}{\sigma_i^2} \tag{4.12}$$

where the x_i are the reconstructed invariant mass respectively of the W boson, of hadronically decaying quark and of BNV decaying quark. $\bar{x}_i$ and σ_i are respectively the mean value and sigma of a gaussian fit of x_i distribution, obtained from $t\bar{t}_{SM\_BNV}$ events using MC truth matching to get the correct jet-parton association. In Figure 4.3a-4.3c invariant W boson, hadronic and top quark BNV mass distributions for muon analysis are shown; indeed for electron analysis no differences have been observed. In Table 4.3 also fit values are reported.
The χ^2 variable is then calculated, for every possible combination of lepton plus five jets (Figure 4.2, bottom). The combinations where one of the two jets associated with the W decay is tagged as a b-jet, using a "tight" working point of the CSV algorithm, are discarded. Among all the possible combinations the one providing the lowest chi-square value is retained. Using the truth information, the correct jet-to-parton association is chosen in 60% of the events that pass the basic selection. Finally in tight selection all the events with a χ^2 value greater than 20 are rejected.
As shown in Figure 4.2 (bottom) even though χ^2 variable does not provide a strong discriminating power between $t\bar{t}_{SM\_SM}$ and $t\bar{t}_{BNV\_SM}$, it helps to further suppress non top background and W+jets in particular.

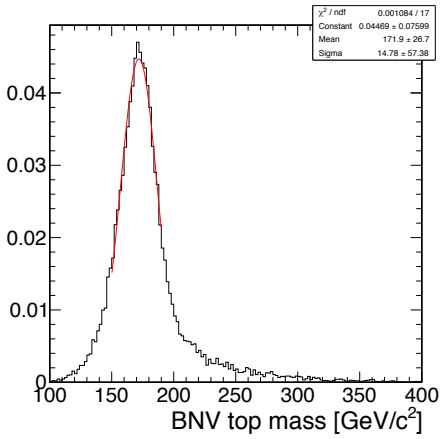
All the previous thresholds on leptons and jets (p_T, χ^2, E_T^{miss}, $b-$tagging discriminating variable) have been previously optimized to give the minimum expected limit on BR: for each point, corresponding to a set of values for these variables, the expected limit has been calculated. The set of values giving the lowest limit has been retained. In fact, since only slight discrepancies have been found among muon and electron channel (except for lepton p_T), the same threshold have been chosen for both the channels.



(a) Mass distribution of the hadronically decaying W boson



(b) Mass distribution of the hadronically decaying top quark



(c) Mass distribution of the BNV decaying top quark

Figure 4.3: Reconstructed distribution of the relevant quantities to build χ^2 variable. The jet association is performed using MC truth matching. All the samples are relative to muon analysis.

M_W	σ_W	M_t	σ_t	M_{tBNV}	σ_{tBNV}
82.4	9.1	171.9	14.8	174.8	17.2

Table 4.3: Fitted central value and σ of the Gaussian fit to the invariant mass of the W boson from the hadronically decaying top (M_W and σ_W), the reconstructed invariant mass of the hadronically decaying top (M_t and σ_t), and the reconstructed invariant mass of the BNV decaying top (M_{tBNV} and σ_{tBNV}). Numbers are in GeV/c^2.

4.5 Background evaluation

4.5.1 Multijet QCD background

In this analysis the QCD contribution is not negligible. Unfortunately the absolute contribution of multijet QCD is not fully reliable when estimated directly from simulation. Therefore multijet QCD background in tight has been computed using two methods, the former relying on a semi-data driven approach (used just as a cross check in 2012), the latter specifically developed for 2012 analysis, fully data-driven.
First, a control region, where QCD contribution was enhanced, has been created. Selection has been setup to enclose an "anti-isolation region", inverting the requirements on lepton isolation (namely $0.1 < I^{\ell}_{\text{Rel}} < 0.2$ for electrons and $0.12 <^{\ell}_{\text{Rel}}< 0.2$ for muons), while all the other cuts remained unchanged. In such a region multi-jet contribution is expected to be dominant and it could be estimated from the difference between the observed yield $N^{anti-iso}_{data}$ and the (not negligible) non-QCD yield $N^{anti-iso}_{non-qcd}$. This latter could be obtained from simulation, adding up the non-QCD backgrounds yields using the samples reported in Table 4.1.

The total QCD yield N_{QCD} can then be estimated as:

$$N_{QCD} = R(N^{anti-iso}_{data} - N^{anti-iso}_{non-qcd}) \tag{4.13}$$

In the semi data-driven approach R is given by the ratio between the number of events in the signal region (N^{iso}) and control region ($N^{anti-iso}$) in QCD simulated samples:

$$R = \frac{N^{iso}_{MC-QCD}}{N^{anti-iso}_{MC-QCD}} \tag{4.14}$$

In fact, due to the limited number of events in such samples, the requirement on the minimum number of jets for QCD has been lowered to 3 (instead of 5) and no χ^2 has been applied for tight selection. However, due to the more intense activity in events with a higher number of jets, one might expect to find a more populated anti-isolation region, yielding lower values of R: therefore the approximation used in this method results in (at most) an overestimation

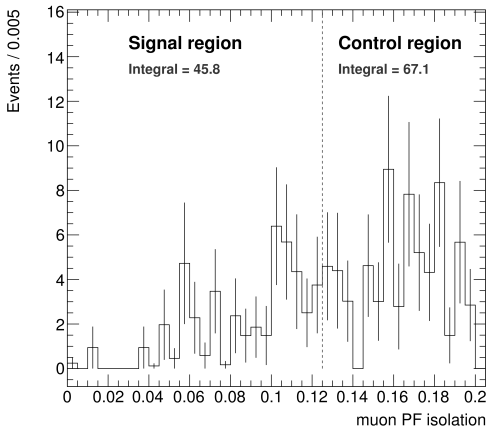


Figure 4.4: 2012 Analysis: Distribution of I^{ℓ}_{Rel} in data and MC for a selection of 1 muon + 3 jet on a simulated multi-jet QCD sample. A signal ($I^{\ell}_{\mathrm{Rel}} < 0.1$) and a control region ($I^{\ell}_{\mathrm{Rel}} < 0.2$) are identified. R is calculated as the ratio between the yield integral in control and signal region.

of QCD yield. In 2011 such strategy has been used to provide and estimate of QCD in basic and tight region for electron channel and just for tight region in muon channel (due to the large uncertainties, the estimate of multi-jet contribution in basic in muon channel requires a different approach, as will be explained in Section 4.5.1).

Selection	$N^{anti-iso}_{data}$	$N^{anti-iso}_{non-QCD}$	R	N^{T}_{qcd}
Basic	1276	631 ± 16	0.51 ± 0.10	329 ± 68
Tight	261	52.2 ± 3.5	0.51 ± 0.10	106 ± 22

Table 4.4: 2011 electron analysis: Relevant numbers for the estimate of the QCD multi-jet yield in basic and tight selection for the 2011 electron analysis. The same method as for the tight selection in the muon analysis has been used. Uncertainties are only statistical.

In 2012 a fully data-driven approach has been used, estimating the value of R from the fake rate, namely the probability that a genuine jet passes the lepton identification. If the jet passes all the lepton identification criteria and a loose isolation threshold ($I^{\ell}_{\mathrm{Rel}} < 0.2$) it is called "loose lepton"; if the jet passes also signal isolation threshold (0.12 for muons, 0.1 for electrons) it is called "tight lepton". An estimate of f as function of jet p_T can be obtained using data, as described in the next paragraph.

In this method R can be expressed as

$$R = \frac{f}{1-f} \tag{4.15}$$

Of course the relationship stated above is exact only for QCD events where just one jet is misidentified. The accuracy of such an approximation is at percent level for this analysis.
The overall uncertainty of this method is assumed to be 50% in tight selection for muons and electrons, and in basic selection for electrons. This uncertainty takes into account both the systematic uncertainty due to measurement of fake rate in a sample with a topology different with respect to the signal and the systematic uncertainty on $N_{non-qcd}^{anti-iso}$. Statistical uncertainties are in the range of 5-35%. The estimated QCD yields in electron channel (in basic and tight selection) are considered 100% correlated.

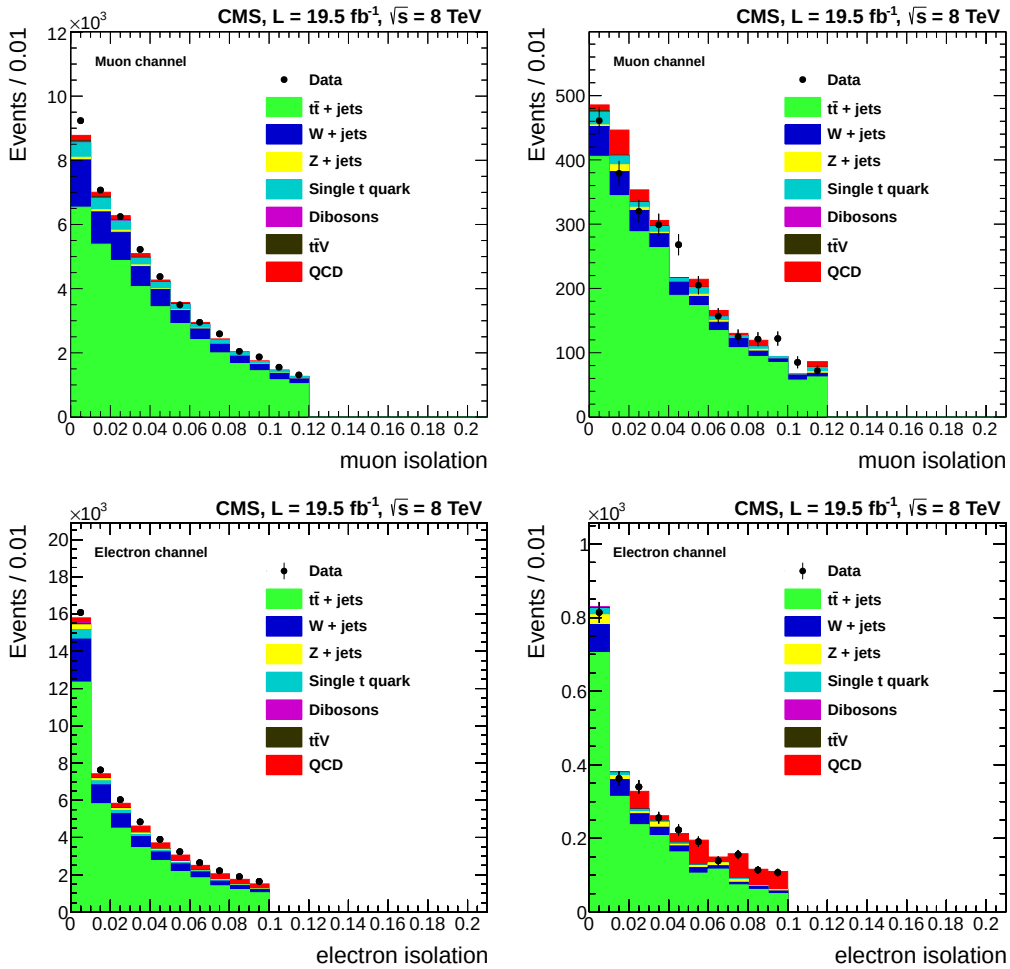


Figure 4.5: 2012 analysis: Distribution of I_{Rel}^{ℓ} in data and MC for basic (left) and tight selection (right), for muon and electron channel.

Estimation of f

An estimate of the fake rate can be obtained setting up a Z+jets selection, where one can look for a loose lepton leptons besides two well identified leptons (muons) coming from the Z decay.
Such a selection has been applied to a data sample of about 19.3 fb^{-1}. The following requirements are imposed:

- Double muon trigger, with threshold on muon transverse momentum of 7 and 18 MeV respectively;
- At least 2 muons with $p_T > 20$ GeV/c in $|\eta| < 2.4$; in addition the same quality cuts adopted in main analysis have been applied to identify these muons;
- The invariant mass of the signal muon is required to be in the range [60, 120] GeV/c^2, compatible with Z peak range; in case of a third muon matches this requirement, the di-muon pair is the one which has the invariant mass closer to the nominal Z mass (91.2 GeV).
- One additional loose lepton (muon or electron); the same quality cuts adopted in main analysis have been applied respectively for muons and electrons, except for isolation, which is set to $I^{\ell}_{\mathrm{Rel}} < 0.2$.

The contribution of genuine leptons coming from WZ events is estimated from simulation, and it must be subtracted since it's not negligible (it corresponds to 30% of the events observed in data for tight lepton identification). ZZ contribution instead is suppressed by the cross section times branching ratio, and by the requirement of no extra lepton.
The Fake rate is then computed as:

$$f(p_T) = \frac{N_{tight}}{N_{loose}} \tag{4.16}$$

being N_{loose} the total number of selected loose leptons and N_{tight} the number of such leptons satisfying tighter requirements.
Since the probability to reconstruct a fake lepton is a function of transverse momentum, f has been calculated for in different bins of loose lepton p_T. In Figure 4.6 muon and electron fake rate have been reported as a function of the lepton p_T.

Muon Channel: QCD yield in basic

The aforementioned approach works well to compute QCD in tight region, where QCD can be estimated directly from a control region. Indeed for 2012 electron analysis the same strategy can be extended to estimate QCD yield in basic, thanks to the relatively small uncertainties on the yields in data and non-qcd samples in control region.

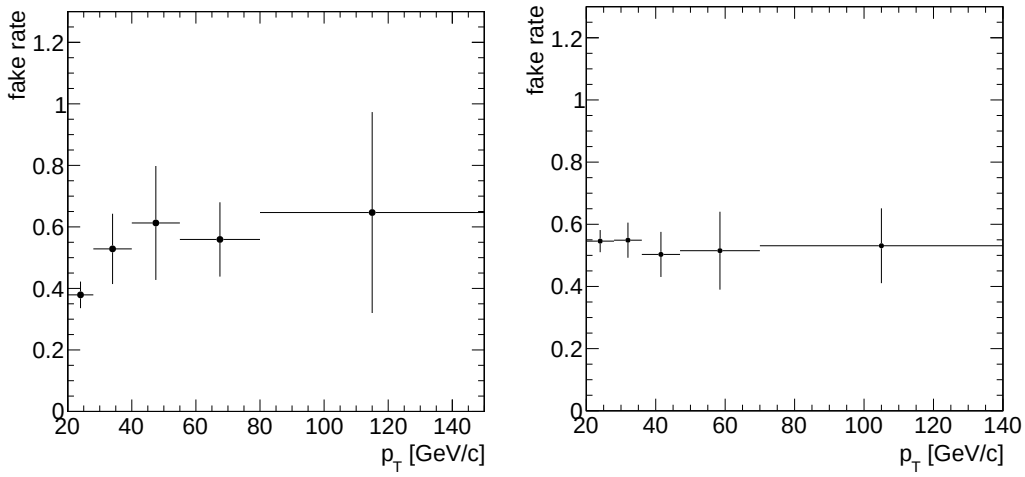


Figure 4.6: 2012 analysis: Fake rate F as function of p_T for muon (left) and electron channel (right). Error bars are statistical uncertainty only.

However, some issues rises when the same procedure is applied to compute QCD in basic selection for 2012 muon channel (where the systematic uncertainty of factor $N_{non-qcd}^{anti-iso}$ yields large uncertainty in QCD yield), or in 2011 (where large statistic uncertainties characterize muon channel due to the low QCD contribution).
For this reason an alternative approach has been used to estimate QCD in basic. Such a strategy relies on the approximated assumptions that efficiency of the χ^2 and E_T^{miss} cuts for events that pass the basic selection are uncorrelated, and can therefore be factorized. The number of QCD events in basic would read:

$$N_{qcd}^{B} = \frac{N_{qcd}^{T}}{\epsilon_{E_T^{miss}} \epsilon_{\chi^2}} \tag{4.17}$$

where N_{qcd}^{T} is the number of QCD events in tight, while $\epsilon_{E_T^{miss}}$ and ϵ_{χ^2} are the efficiencies to pass $\chi^2 < 20$ and $E_T^{miss} < 20$ GeV/c. These efficiencies are calculated from simulations as shown in Figure 4.5.

$\sqrt{s}$	$\epsilon_{E_T^{miss}}$	ϵ_{χ^2}	N_{qcd}^{T}	N_{qcd}^{B}
7 TeV	0.47 ± 0.09	0.55 ± 0.03	9 ± 5.1	35 ± 21
8 TeV	0.33 ± 0.06	0.45 ± 0.04	118 ± 41	790 ± 320

Table 4.5: Muon Analysis: Efficiency of χ^2 and E_T^{miss} cuts in 2011 and 2012 analysis; the total number of QCD events in basic N_{qcd}^{B} for muon analysis has been calculated from Equation 4.17. Only statistical uncertainties have been reported.

In 2011 using a semi-data driven approach, due to the large statistical uncertainties on QCD sample, a total uncertainty of 100% has been assumed

Table 4.6: Muon and electron channels: numbers relevant for the estimate of the QCD multijet yield based on simulated samples. The reported uncertainties are statistical only.

Muon channel (MC method)				
Selection	$N_{\text{data}}^{\text{antiiso}}$	$N_{\text{nonQCD}}^{\text{antiiso}}$	R	N_{QCD}
Basic	6019 ± 78	5428 ± 69	0.63 ± 0.09	360 ± 80
Tight	412 ± 20	258 ± 15	0.63 ± 0.09	92 ± 20
Electron channel (MC method)				
Selection	$N_{\text{data}}^{\text{antiiso}}$	$N_{\text{nonQCD}}^{\text{antiiso}}$	f	N_{QCD}
Basic	7162 ± 84	4400 ± 54	0.51 ± 0.13	1410 ± 360
Tight	542 ± 23	223 ± 12	0.51 ± 0.13	160 ± 40

Table 4.7: Muon and electron channels: numbers relevant for the estimate of the QCD multijet yield based on the misidentification rate measurement (Equation 4.13). Only the average value of f is reported, while values computed in bins of p_T are used in the analysis.

Muon channel (Fake rate mathod)				
Selection	$N_{\text{data}}^{\text{antiiso}}$	$N_{\text{nonQCD}}^{\text{antiiso}}$	f	N_{QCD}
Tight	412	257 ± 61	0.44 ± 0.09	119 ± 60
Electron channel (Fake rate method)				
Selection	$N_{\text{data}}^{\text{antiiso}}$	$N_{\text{nonQCD}}^{\text{antiiso}}$	f	N_{QCD}
Basic	7162	4400 ± 950	0.51 ± 0.10	2900 ± 1400
Tight	542	222 ± 53	0.51 ± 0.10	330 ± 160

for muon channel in basic and tight, while for electron channel a 50% error has been set. In 2012 analysis in addition to the 50% uncertainty assumed for N_{qcd}^{T}, an uncertainty of 50% has been assumed on the product of $\epsilon_{E_T^{miss}}$ and ϵ_{χ^2}. From the square sum of these contributions one gets an overall uncertainty of 70%. The partial correlation between tight and basic selection has been taken into account when calculating observed and expected limits.
Results for QCD yield in the electron and muon channels for 2012 analysis have been reported in Tables 4.6 and 4.7.

The alternative 2011 approach to QCD estimation in electron channel

In 2011 analysis a further approach has been proposed to give and estimate of the total QCD contribution, whose core relies on the fit of QCD shape, extracted by the difference between the isolation variable distribution observed in data and that from non-QCD processes obtained from MC.
One must setup a selection as closer as possible to the main selection, at the

same time reducing as much as possible statistical uncertainties. To this aim one can observe that the isolation variable distribution I^{ℓ}_{Rel} does not depend on the number of jets 4.7 or on the particular process considered (Figure 4.8).

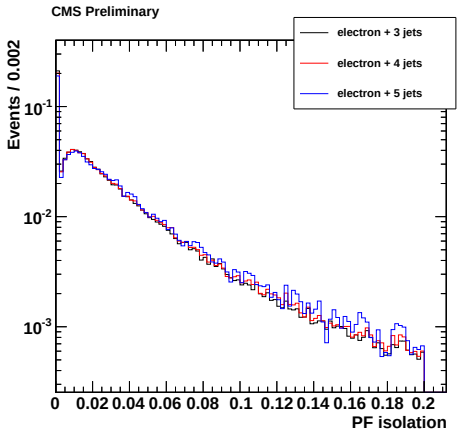


Figure 4.7: Distribution of I^{ℓ}_{Rel} for a simulated $t\bar{t}$ sample with at least 3,4,5 jets with $p_T > 30$ GeV/c: changes in shape are almost negligible.

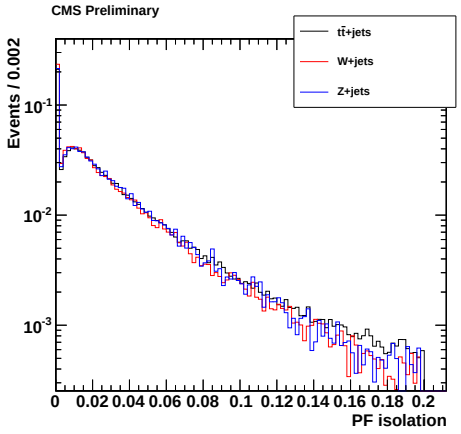


Figure 4.8: Comparison among I^{ℓ}_{Rel} distributions for simulated $t\bar{t}$, $W + jets$, $Z + jets$ samples passing basic selection.

Thus the shape of lepton isolation variable for non-QCD events can be obtained from simulated $t\bar{t}$ contributions, selecting events with at least one electron which meets the requirements of main selection and at least three jets (instead of five) with $p_T > 30$ GeV/c. The distributions in Figure 4.9 are produced.
The template shape of QCD contributions in the range between [0.06, 0.2] is obtained from data after subtraction of the I^{ℓ}_{Rel} distribution from non-QCD processes (obtained from MC). Since for low values of the isolation variable the QCD contribution is small, the first bin of non-QCD distribution is imposed

Model	N^B_{qcd}	$N^B_{non-qcd}$	$\chi^2/ndf(B)$	N^T_{qcd}	$N^T_{non-qcd}$	$\chi^2/ndf(T)$
poly 3	553 ± 14	7795	1.8	120 ± 4	711	0.56
poly 4	505 ± 10	7845	1.8	109 ± 3	723	0.55
exp	490 ± 10	7859	2.0	115 ± 3	717	0.55

Table 4.8: 2011 electron analysis: Relevant numbers for the estimate of the QCD multi-jet yield in the basic and tight selection for the 2011 electron analysis using the fit method. Total number of events Polynomial of third (poly3) and fourth (poly4) degrees and exponential ($A(1-exp(-x/B))$) have been tested. For each tested function form, for both basic and tight selection reduced χ^2 of the fit has been reported. Uncertainties are only statistical

to be equal to the first bin in data. The resulting correction is very small, being the normalization factor 1.007 (the isolation distribution changes less than 1%).
Some functional form (reported in Table 4.8) are used to fit the distribution. Fit results are also shown in Figure 4.10. Uncertainties in the non-qcd process over the range [0.06, 0.2] are dominated by shape uncertainties (total contribution of 20%), even though the yield in this range are definitely dominated by the data.
Once a template functional has been defined, one can fit the observed with the sum of non-QCD and QCD contribution, introducing the parameter pair (A, B), namely the weights for QCD and non-QCD contributions:

$$N^{anti-iso}_{data}(x_i) = Af(x_i) + BN^{anti-iso}_{non-qcd}(x_i) \tag{4.18}$$

where x_i represents the isolation variable. The fit results are shown in Table 4.8, and they look in agreement with the first QCD estimate (Table 4.4). The statistical error on the QCD yield, provided by the fitting procedure, is in the order of few percents. The isolation distribution, where the contribution of the fourth degree polynomial and the exponential fits are shown, is reported in Figure 4.11 and 4.12.
The main issue concerning this method is due to the indetermination on the functional form in proximity of very low isolation values. Indeed it's not trivial to guess exactly features of the function (as the possible presence of flexes). For this reason this method appears not to be completely reliable and it has been taken just as a further cross check of values calculated before.

4.5.2 γ+jets contribution

In electron channel isolated electrons can arise from γ conversion coming from γ + jets processes, which has not been taken into account yet (the isolated photon convert before reaching calorimeters). In 2011 the following method has been used to provide an estimate of such a contribution: a γ+jets-enriched

region has been defined, reversing all the photon conversion criteria applied to reconstructed electrons; such region has been defined "control region" and enclose a complementary region of the phase space with respect to the "signal" one. Given the sum of QCD and γ+jets, the net fraction of this latter can be calculated as

$$f = \frac{N_{\gamma+jets}}{(N_{\gamma+jets} + N_{qcd})} \tag{4.19}$$

being $N_{\gamma+jets}$ the yield of $\gamma + jets$ processes. Indeed f can be extracted by the following expression:

$$N_{qcd} = (1 - f)(N_{data}^{anti-conv} - N_{non-qcd}^{anti-conv})R_{qcd}^{conv} \tag{4.20}$$

R_{qcd}^{conv} stands for the ratio between the yield of simulated multi-jet QCD in the signal and control region.
Similarly a coefficient $R_{\gamma+jets}^{conv}$ has been introduced to represent the ratio between the yield in signal and control region, in the γ+jets MC sample. Since N_{qcd} is known, one can extract f from Equation 4.21 and calculate the absolute contribution $N_{\gamma+jets}$ according to the following expression:

$$N_{\gamma+jets} = f \times (N_{data}^{anti-conv} - N_{non-qcd}^{anti-conv}) \times R_{\gamma+jets}^{conv} \tag{4.21}$$

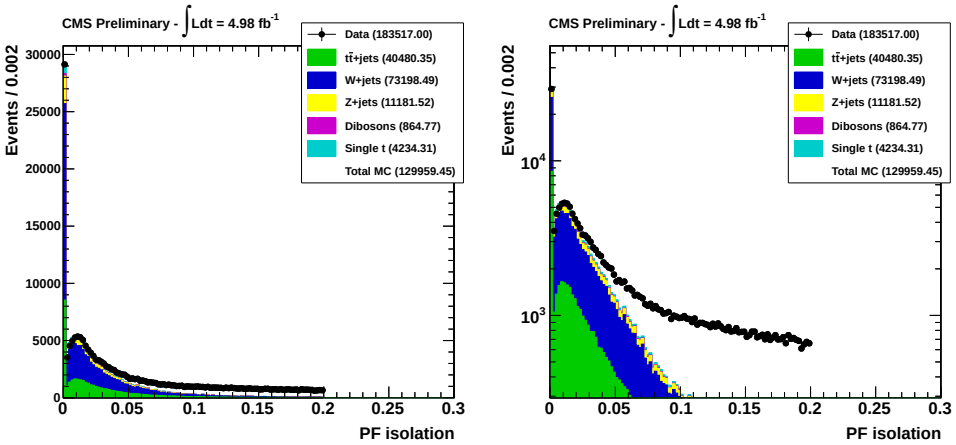


Figure 4.9: 2011 electron analysis: distribution in linear (left) and logarithmic (right) scale of I_{Rel}^{ℓ} in data and MC (only for non-QCD processes) for events with an electron and at least 3 jets with $p_T > 30$ GeV/c.

The γ+jets is found to be largely subdominant with respect to multijet QCD process, as shown in Table 4.9. In 2012, thanks to the higher statistics in MC samples and in the light of 2011 results, the contribution of γ+jets has been calculated from simulation as a relative contribution with respect to QCD multi-jet event yield. A full basic and tight selection has been applied to simulated dataset, and the ratio between the total QCD and $N_{\gamma+jets}$ yield has been calculated (Table 4.10).

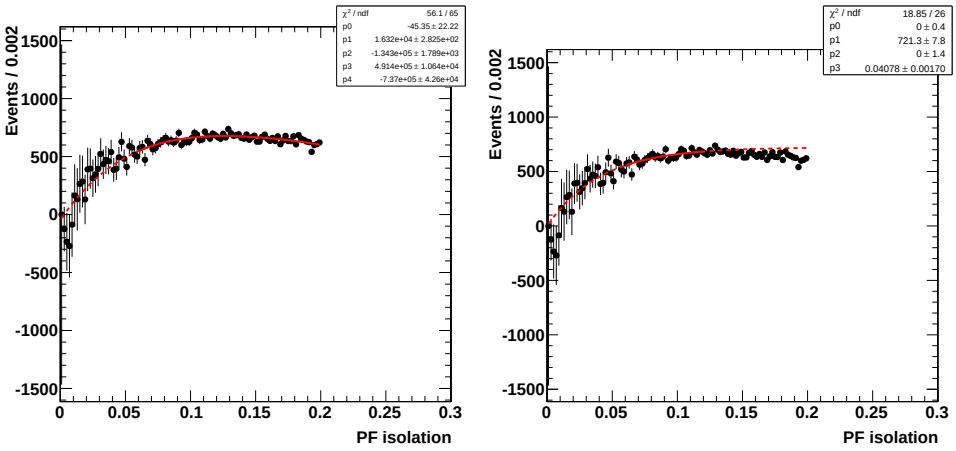


Figure 4.10: 2011 electron analysis: fourth degree polynomial (left) and exponential (right) fit of the difference between I^{ℓ}_{Rel} in data and MC (only for non-QCD processes) for events with an electron and at least 3 jets with $p_T >$ 30 GeV/c.

Selection	N_{qcd}	f	N^{conv}_{data}	$N^{conv}_{non-QCD}$	R^{conv}_{qcd}	$R^{conv}_{\gamma+jets}$	$N_{\gamma+jets}$
Basic	329 ± 68	0.19	219	16.6 ± 3.6	2.0 ± 1.1	1.2 ± 0.5	45
Tight	106 ± 22	0.045	57	1.5 ± 0.7	2.0 ± 1.1	1.2 ± 0.5	3

Table 4.9: 2011 electron analysis: Relevant number of the estimate of the QCD γ+jets yield in basic and tight selection for the electron analysis. Uncertainties are only statistical

Even though in tight selection the estimate is affected by large statistical uncertainties, it is compatible with estimate in basic. Moreover, R for $\gamma + jets$ events has been calculated using simulated samples, and it is found to be about 10 times larger than multi-jet QCD events. To take into account $\gamma + jets$ contribution the QCD yield estimated with fake rate method has been increased by 2%.

Sample	Yield
Multi-jet QCD	65.08 ± 26.71
γ+Jets	1.28 ± 0.02
Ratio	0.02 ± 0.01

Table 4.10: 2012 electron analysis: Relevant numbers for the estimate of the relative contribution of γ+jets with respect to the multi-jet QCD, applying the standard basic selection for both datasets. Uncertainties are only statistical.

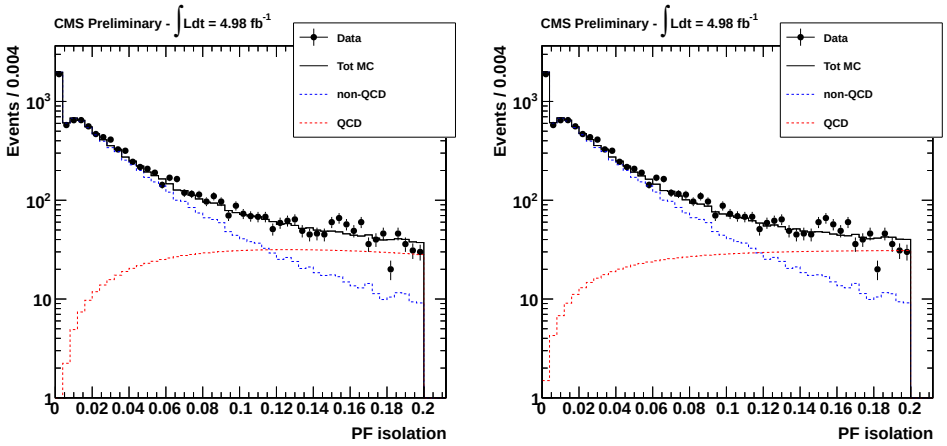


Figure 4.11: 2011 electron analysis: fourth degree polynomial (left) and exponential (right) fit to the observed distributions of I^{ℓ}_{Rel} for basic selection.

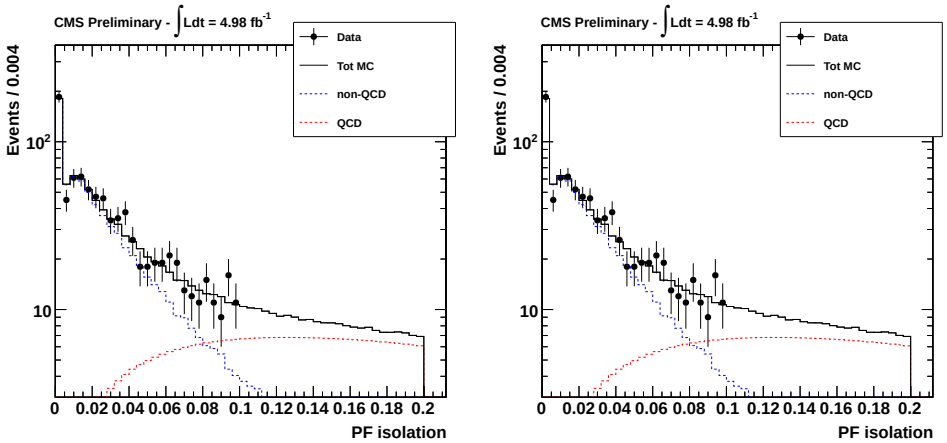


Figure 4.12: 2011 electron analysis: fourth degree polynomial (left) and exponential (right) fit to the observed distributions of I^{ℓ}_{Rel} for tight selection.

Chapter 5

Results

A counting experiment has been performed to quantify the presence of possible BNV processes: a significant excess in data can be interpreted as the existence of new physics processes compatible with the BNV decay. The best value of BR that fits the expected yield to data can be found minimizing a likelihood function.
If no significant excess is observed, an upper limit on decay branching ratio (BR) can be set: the experimental observable, namely the number of counts in tight N^T_{data}, is compared to the expected number of events N^T_{exp} (defined in Section 4.3). N^T_{exp} indeed can be read as the mean value of a poissonian distribution $P(x; N^T_{exp})$, while N^T_{data} represents the realization of x variable.

The nuisance parameters

Of course, as already explained in Section 4.3, N^T_{exp} is a function of BR and of a set of likelihood parameters θ (namely the selection efficiencies defined in Section 4.3, the cross sections, the non-top background yields), which affect directly the estimate of N^T_{exp}.
Such parameters are affected by systematic uncertainties: the jet energy scale, the jet energy resolution, the luminosity, the estimate of theoretical cross sections are just some of the quantities that may change the likelihood parameters values (the full list can be found in Section 5.2). The effect of systematic uncertainties on the likelihood parameters can be determined using stochastic variables, called "nuisance parameters": such variables are distributed according to PDFs, which basically describe the knowledge we have on them. PDFs parameters are assumed to be known.
Each likelihood parameter is function of different nuisance parameters: therefore it is not represented by a single value, but it is rather distributed according to a PDF as well.
The mathematical treatment of systematic uncertainties depends on the a statistical approach: two different methods have been used in 2011 and 2012,

which are described in detail in Section 5.1. Analysis results are presented in Section 5.3.
Statistical treatment of both 2011 and 2012 analyses have been implemented in RooStats framework [82][83].

5.1 Statistical methods

5.1.1 2011 approach

The full 2011 statistical method has been reported in the Appendix B. For brevity in the following lines just a summary is reported.
In 2011 analysis all the likelihood parameters affecting the estimated of N^T_{exp} are assumed to be distributed according to lognormal probability density functions, whose parameters (namely the median and the k factor [81]) are assumed to be know (they are estimated from simulation).
Since the parameters are affected by common sources of uncertainty, a (positive) correlation among them is expected. To account for any correlations, likelihood parameters have been split in subsets: correlations are expected to exists inside each subsets. Parameters belonging to different subsets are assumed to be independent of each other.
Following a conservative approach the uncertainty on each likelihood parameter is described as the combination of two distinct contributions: the first, shared by all parameters in its subset of correlated parameters, which account for the correlated part of the uncertainty; the second, describing the uncorrelated part, which provides independent and stochastic variations for each of them. From the combination of this two contributions a lognormal-distributed variable is obtained. All the details of the methods are given in Appendix B.
Even if correlations between non top background yield in basic and tight (N^B_{bck} and N^T_{bck}) and among efficiencies have been estimated, an exact computation of correlation for each pair of nuisance parameter cannot be provided in this approach. Due to this unsatisfactory feature, in 2012 this method has been dropped in favor of a template-morphing-oriented approach, described in the next section.

5.1.2 2012 approach

The 2012 statistical approach follows a strategy similar to the one adopted in [84].
Again, the likelihood contains all the parameters affecting the estimate of N^T_{exp}: they are the eleven efficiency values, $\sigma_{t\bar{t}}$ and σ_{tW} and the number of non-top events passing basic and tight selection (respectively N^B_{bck} and N^T_{bck}).
The efficiency values are all estimated from simulation and are expected to be highly correlated among each other. Treating them as fully correlated would be simple, but would lead to an underestimate the overall uncertainty

in N^T_{exp}. Thus a more accurate method, described generically in the following, is adopted.

Parameterization of efficiencies and V+Jets yields

As a starting point, likelihood parameters must be parameterized as function of nuisance parameters.
Among all the likelihood parameters θ_j, selection efficiencies and vector bosons plus jets yield (which contributes to the total non top background yield N^B_{bck} and N^T_{bck}), are both estimated directly from simulation and are expected to be correlated. Actually each source of uncertainty (listed in Section 5.2) may affect several parameters at the same time, introducing correlations.
Each source of uncertainty, measured in units of standard deviations, is associated to a random variable x_i (the nuisance parameter), distributed according to a unit gaussian probability density function $\mathcal{G}(x_i; 0, 1)$. The impact of x_i on each likelihood parameter θ_j is controlled by a transfer function $\Delta\theta_{ij}(x)$, which represents the shift of parameter θ_j from its measured value $\tilde{\theta}_j$, for every x_i value.
Given m sources of uncertainties, one can assume that the shifts sum up linearly:

$$\theta_j = \tilde{\theta}_j + \sum_{i=0}^{m} \Delta\theta_{ij}(x_i) \tag{5.1}$$

The shifts $\Delta\theta_{ij}$ observed in the j-th likelihood parameter when the $i-$th source of uncertainty is varied by plus or minus one standard deviation are noted as θ^+_i and θ^-_i. Therefore one can write:

$$\theta_j(x) = \tilde{\theta}_j(x) + \begin{cases} \Delta\theta_{ij}(x = -1) = \theta^-_{ij} \\ \Delta\theta_{ij}(x = 0) = 0 \\ \Delta\theta_{ij}(x = 1) = \theta^+_{ij} \end{cases} \tag{5.2}$$

θ^-_i and θ^+_i can be explicitly calculated with dedicated prescriptions typical of each source of uncertainty. These prescriptions will be described later in the text.
Once these values are known, three points (-1, θ^-_{ij}), (0, 0), (+1, θ^+_{ij}) determined in a 2D plane can be interpolated with a second order polynomial. The polynomial defines the function $\Delta\theta_{ij}(x_i)$ is expected to return, for any departure of the source of uncertainty from its central value, the shift experienced by the likelihood parameter.
In principle, the larger the number of interpolated points, the more accurate the dependence. Yet the punctual knowledge of the impact of many systematic uncertainty fluctuations is not always available.
In order to protect the function $\Delta\theta_{ij}(x)$ from steep, unlinear trends due to the quadratic behavior of the interpolated function, a linear extension $\Delta\theta^L_j(x)$ is

imposed outside the range [-1,+1] by requiring the continuity of the function and its first derivatives

$$\begin{cases} \Delta\theta_j(x_i = -1^-) = \Delta\theta_j^L(x_i = -1^-), & \Delta'\theta_j(x_i = -1^-) = \Delta'\theta_j^L(x_i = -1^-) \\ \Delta\theta_j(x_i = 1^+) = \Delta\theta_j^L(x_i = 1^+), & \Delta'\theta_j(x_i = -1^-) = \Delta'\theta_j^L(x_i = -1^-) \end{cases} \tag{5.3}$$

where $\Delta'\theta_j$ stands for the first derivative of each function.
To preserve the physical meaning of parameters, that by definition must be always positive, one imposes that efficiencies are always confined between 0 and 1 and both vector boson yields are always positive.

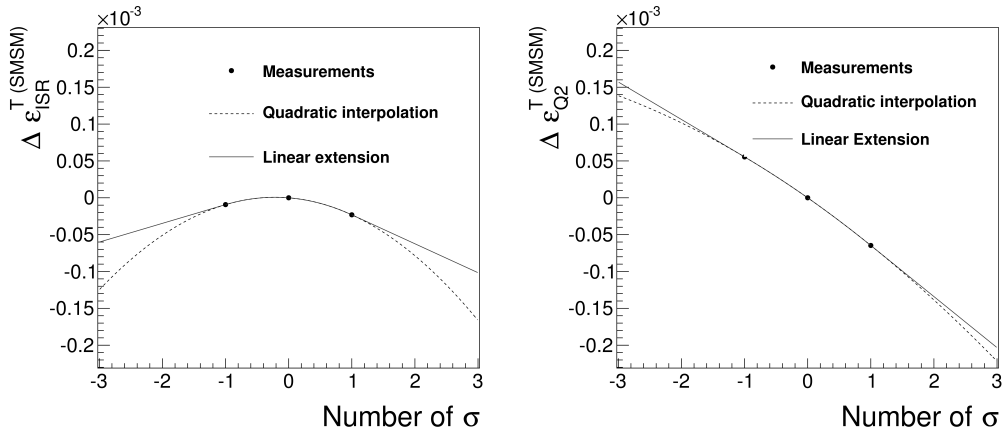


Figure 5.1: Absolute impact of initial/final state radiation (left) and factorization scales (right) on $\epsilon^T_{t\bar{t},SM\_SM}$ as function of the shift from the central value, expressed as number of sigmas. Each uncertainty source has been fluctuated of plus (minus) one sigma, then the efficiency to pass basic selection has been recalculated. The three points have been interpolated by a quadratic functional form ($y = ax^2 + bx + c$); a regularization has been made extending the interpolating function in a linear fashion outside [-1, 1] sigma interval.

Parametrization of QCD yields and cross section

The treatment of systematic uncertainties for QCD yield, $t\bar{t}$ and tW cross sections is explained in the following.
Non-top background events in basic and tight selection are split into three contributions: vector bosons plus jets background, QCD and other (subdominant) backgrounds. The treatment of V+jets yield has been already described in the previous section. The remaining contributions of non-top background different from multi-jet QCD are very small (less than 2% of the total expected yield), and therefore their uncertainties have been neglected.
The QCD yield, σ_{tW} and $\sigma_{t\bar{t}}$ are quantities that are expected to be independent of any other and can be treated individually. They are simply assumed to be

distributed according to lognormal distributions, namely:

$$L(x, m0, k) = \frac{1}{\sqrt{(2\pi)}\ln(k)x} \times e^{-ln^2\left(\frac{x/m0}{2\ln^2(k)}\right)} \tag{5.4}$$

m_0 standing for the nominal value, and k for the relative uncertainty on each quantity, added to one.
Concerning the QCD yield, which is estimated with a data-driven method, correlations need to be taken into account only between the estimate in basic and tight. In order to achieve this goal a lognormal distribution is defined via a change of variable. A unit-Gaussian distributed variable x_G is introduced:

$$x_{logn} = \exp\left(m_G + x_G\sigma_G\right) \tag{5.5}$$

By definition x_{logn} is distributed according to a lognormal pdf with the form shown in Expression , with $m_0 = \log(m_G)$ and $k = \log(\sigma_G)$.
Such a strategy, which is basically the one adopted in the 2011 analysis, allows to take easily into account partial QCD correlation between the basic and tight selection. In fact the gaussian variable x_G can be written as function of two gaussian-distributed unitary variable x_u, x_c:

$$x_G = \frac{\sigma_c \times x_c + \sigma_u \times x_u}{\sqrt{\sigma_c + \sigma_u}} \tag{5.6}$$

being σ_c and σ_u respectively the correlated and uncorrelated contribution to QCD uncertainty. x_c control at the same time the contribution of QCD yield in tight selection, accounting for correlations between basic and tight region. As long as it is applied to just 2 correlated variables, which is the case of the QCD basic and tight yields, this method does not introduce any approximation.

The likelihood

The full list of nuisance parameters is then composed by the x_i variables, associated to the systematic uncertainty sources, and by all the gaussian unitary variables, listed in the previous paragraph.
Once all the likelihood parameters are expressed as function of nuisance parameters, one can write the expression of the final likelihood which reads:

$$\mathcal{L}(\text{data} \,|\, BR, \theta_j(x_i), \theta_l(\tilde{\theta}_l, \tilde{\sigma}_l)) = \tag{5.7}$$

$$\mathcal{P}\Big(N^T_{obs} \,|\, N^T_{exp}(BR, \theta_j(x_i), \theta_l(\tilde{\theta}_l, \tilde{\sigma}_l))\Big) \cdot \tag{5.8}$$

$$\cdot \prod_i \mathcal{G}(x_i) \cdot \prod_l \rho(\theta_l, \tilde{\theta}_l, \tilde{\sigma}_l)\,, \tag{5.9}$$

where $\mathcal{P}\Big(N^T_{obs} \,|\, N^T_{exp}(BR, \theta_j(u_i), \theta_k(\tilde{\theta}_k, \tilde{\sigma}_k))\Big)$ is the Poissonian probability density function with expectation value N^T_{exp}. θ_j represents the set of all the

likelihood parameters depending on unitary variables x_i; θ_k stands for all the parameters log-normal distributed, having nominal value $m_0 = \tilde{\theta}_l$ and relative uncertainty $k = \tilde{\sigma}_l$.

5.2 Systematic uncertainties

In 2011 and 2012 analyses a number of uncertainty sources have been considered, having the same impact on both muon and electron channel. For each source of systematic uncertainty a quick description is given in Section 5.2.1.
To give an estimate of the impact of each uncertainty source, one has to calculate the relative variation for every parameters, in correspondence to (minus) 1-sigma fluctuation of each source. In such a way, also θ_j^+ and θ_j^- can be calculated.
The final relative uncertainties for 2012 analysis are reported in Tables 5.1 and 5.2. In case a significant asymmetry between variations around central values is observed, both changes are reported, with their sign; if just one number is given, one must assume a symmetric fluctuation of a given parameter around its central value; if no numbers are reported, the corresponding source has no impact on a given parameter; the presence of a symbol "-" indicates that the impact of the source cannot be quantified (e.g. due to the lack of dedicated samples).
Similar tables have been filled also for 2011 analysis, but for the sake of shortness will not be displayed.

In Table 5.5 and 5.9 are reported central values of the relevant parameters appearing in the likelihood respectively for 2011 and 2012 analyses, with their overall uncertainties. Likelihood parameters assumes their central values when nuisance parameters assumes their nominal values as well.

5.2.1 Sources of systematic uncertainties

The following sources of systematic uncertainties have been taken into account:

- **Jet energy scale**: as explained in Section 3.3.4, jet energy scale is not exactly reproduced by MC samples, thus jet energy scale (JES) in simulation it has to be corrected to fix these undesired effects. Jet energy corrections can be parameterized in different ways, as function of dynamical quantities (η, p_T etc.) or with constant offsets [89]. Their uncertainties generally have complex dependence on the jet p_T and η and is smaller than 3% [85]. The jet energy scale is one of the sources with the largest impact on likelihood parameters.

- **Jet energy resolution**: it is known to be underestimated [85] of about 10% in MC samples. To take into account this effect, the reconstructed

Par	JES	JER	PU	BTAG	UNCL	lepton eff	LUMI
$\epsilon^{B}_{t\bar{t}}$_SM_SM	11	< 1	1	< 1	0	5	
$\epsilon^{T}_{t\bar{t}}$_SM_SM	-9/11	2	2	< 1	1	5	
$\epsilon^{B}_{t\bar{t}}$_BNV_SM	11	1	1	< 1	0	5	
$\epsilon^{T}_{t\bar{t}}$_BNV_SM	14	5	2	< 1	0/-4	5	
$\epsilon^{B}_{t\bar{t}}$_BNV_BNV	-12/14	1	1/-2	< 1	0	5	
$\epsilon^{T}_{t\bar{t}}$_BNV_BNV	-11/13	-2/0	4/-5	< 1	2/-4	5	
$\epsilon^{B}_{\mathrm{tW}}$_$SM$	10	< 1	0/-4	3	< 1	5	
$\epsilon^{T}_{\mathrm{tW}}$_$SM$	10	-2/0	2/-3	< 1	-3/0	5	
$\epsilon^{B}_{\mathrm{tW}}$_$BNV$	-11/14	1	2/-1	1	0	5	
$\epsilon^{T}_{\mathrm{tW}}$_$BNV$	11	3	3	4	1	5	
N^{B}_{bck} (V+jets)	-10/12	1	1	4	5	5	4.5
N^{T}_{bck} (V+jets)	9	2	2	3	5	5	4.5

Table 5.1: Relative uncertainties (%) in the likelihood parameters in the muon and electron analyses arising from experimental sources. In those cases where a significant asymmetry is observed in the change of the parameters, both changes, with their sign, are reported. Whenever there is no number reported, the corresponding source has no impact on the parameter.

jet p_T^R in simulated data are first oversmeared, defining a new transverse momentum p_T as follows:

$$p_T = \max\left[0.0, \frac{p_T^R + \Delta p_T}{p_T^R}\right] \tag{5.10}$$

where Δp_T, given the jet momentum p_T^{GEN} at generation level, reads

$$\Delta p_T = (p_T^R - p_T^{GEN}) \times f \tag{5.11}$$

being f a smearing factor, depending on reconstructed jet pseudorapidity. Systematic uncertainties are taken into account substituting in Equation 5.11 a value for f which corresponds to +1/-1 sigma variation of jet energy resolution. Finally such corrections are propagated for the calculation of E_T^{miss}. Uncertainties on muon and electron energy resolution are negligible.

- **B-tagging efficiency**: in this analysis a b-tagging CSV algorithm has been used (described in Section 3.3.4); however the model of b-tagging is not perfect. The bias can be corrected reweighting events on the basis of their actual content in terms of b-, *light* quark and gluons, inferred by using MC truth. Indeed weights are function of jet p_T, η and flavor. The residual uncertainty on the b-tag efficiency is of the order of few percents [86].

Par	PDF	ISR	scales	x-s	Tot
$\epsilon^{B}_{t\bar{t}}$_SM_SM	2	0/-3	14/-13		19
$\epsilon^{T}_{t\bar{t}}$_SM_SM	4	-2/-5	12/-14		20
$\epsilon^{B}_{t\bar{t}}$_BNV_SM	3	-	-		12
$\epsilon^{T}_{t\bar{t}}$_BNV_SM	3	-	-		17
$\epsilon^{B}_{t\bar{t}}$_BNV_BNV	3	-	-		15
$\epsilon^{T}_{t\bar{t}}$_BNV_BNV	3	-	-		15
$\epsilon^{B}_{\mathrm{tW}}$_$SM$	1	-	-		12
$\epsilon^{T}_{\mathrm{tW}}$_$SM$	2	-	-		12
$\epsilon^{B}_{\mathrm{tW}}$_$BNV$	3	-	-		15
$\epsilon^{T}_{\mathrm{tW}}$_$BNV$	3	-	-		14
N^{B}_{bck} (V+jets)	2	-	-	50	52
N^{T}_{bck} (V+jets)	3	-	-	50	52
N^{B}_{bck} (QCD-Muon)	-	-	-	-	70
N^{B}_{bck} (QCD-Ele)	-	-	-	-	50
N^{T}_{bck} (QCD)	-	-	-	-	50
N^{B}_{bck} (Tot-Muon)	-	-	-	-	41
N^{T}_{bck} (Tot-Muon)	-	-	-	-	34
N^{B}_{bck} (Tot-Ele)	-	-	-	-	34
N^{T}_{bck} (Tot-Ele)	-	-	-	-	30
σ_{st}					8
$\sigma_{t\bar{t}}$					15

Table 5.2: Relative uncertainties (%) on the likelihood parameters in the muon and electron analyses arising from theoretical sources. In those cases where a significant asymmetry is observed in the change of the parameters, both changes, with their sign, are reported. Whenever there is no number reported, the corresponding source has no impact on the parameter. The presence of a symbol "-" indicates that the impact of the source cannot be studied due to lack of dedicated samples. In the case of the efficiency values, the value obtained on a similar sample is adopted. The last column reports the total uncertainty.

- **Muon/electron trigger**: the efficiency on the muon and electron trigger, isolation and identification, is biased. A unique value of 5% has been used to describe relative uncertainty in the efficiency of the muon end electron trigger, identification and isolation, as estimated from MC.

- **Unclustered tracks and energy deposits**: they represents objects not associated to reconstructed lepton and jets, with a minimum p_T of 10 GeV. They are used to compute E_T^{miss} [108]; an uncertainty at MC level on the modeling of such quantities implies an uncertainty on the final measurement. The systematic uncertainty can be evaluated varying unclustered energy contribution to E_T^{miss} by 10%.

- **Pileup**: MC is produced overlaying pileup interactions to each events; however each sample must be reweighted to match the number of vertex distribution found in data. The weight uncertainty in the is dominated by the uncertainty on the total inelastic cross section and on the measured luminosity (both quantities are taken into account in the reweighting procedure). The total uncertainty however can be obtained shifting the overall mean of interactions distribution in data and then performing the re-weighting. One can do it just recalculating the actual vertex distribution by varying the total inelastic cross section of 5%, around the central value.

- **Parton Distribution Functions (PDFs)**: they affect the BR measurements, as they affect, for instance, the jet p_T spectrum. The simulated samples are generated using CTEQ 6.6 PDFs. The uncertainty on PDFs are described by 22 parameters, which are varied one by one, giving 22 pairs of additional PDFs [88]. For each nuisance parameter, the sum in quadrature of the largest shift (up and down) of each PDF is assumed as an estimate of uncertainty deriving from PDF.

- **Theoretical uncertainties**: The modeling of initial and final state radiation (ISR/FSR) as well as the determination factorization and renormalization scales (Q^2) are studied just for $t\bar{t}$ process. Selection efficiencies can be recalculated for +/- 1 sigma variation of factorization, renormalization and merging scale on different dedicated simulated samples.

- $t\bar{t}$ **and** tW **Cross section**: They are assumed to be 15% and 8% respectively and, conservatively, uncorrelated (because they appear as a ratio). [77]

- **Integrated luminosity**: It has an impact just in calculation of V+jets contributions. An uncertainty of 4.5% has been assumed on integrated luminosity [90].

- **V-Jets cross sections**: This analysis uses the theoretical values of cross sections, affected by an uncertainty of about 5%. The CMS measurements of the ratio between the W+4 jets and the inclusive W+jets

cross section is in agreement with MadGraph predictions at the level of 30%. In addition, a further statistical uncertainty of 10%, due to small size of simulated samples, must be accounted at least in tight selection. Conservatively, an uncertainty of 50% on the W+Jets and Z+Jets cross section has been assumed.

- **QCD yield**: As explained in Section 4.5.1 an uncertainty of 70% (50%) has been assumed for muon analysis in basic (tight) selection during 2012, while an uncertainty of 50% has been assumed for electron QCD yield.

5.3 Results

In Table 5.3, 5.4 and 5.7, 5.8 the expected yields for all the SM processes for 2011 and 2012 analysis are reported, as well as the observed yields in data for muon and electron channel. In these tables BR is assumed to be zero. In the first column the raw yield in basic selection is reported; in the second column basic yield has been corrected for $t\bar{t}$ and tW backgrounds, as a consequence of assuming $N^B_{obs} = N^B_{exp}$, as detailed in Equation 4.6. In fact, this correction consists of a common multiplicative factor, F, which allows the sum of all yields to be equal to the observed yield in basic selection.
The same factor F must scale also expectations of top-background yield for tight selection. Such an approach assumes that the shift between the total expected and the observed yield in basic is actually due to uncertainties on the dominant top background, namely $t\bar{t}$ and tW. The correctness of the normalization approach does not simply rely on the fact that the signal fraction in basic selection is negligible: actually the normalization of basic selection can be performed for any tested BR value. This method is valid as long as the expected signal yield in tight is much higher than the expected one in basic selection.
QCD has been estimated using the approach described in Section 4.5.1 and 4.5.1 respectively for muon and electron channel.
All the tables have been approximated keeping two significant digits for central values and its relative uncertainties (namely the squared sum of systematic and statistic uncertainty).
It's the worth to note that counting experiment strategy overlooks data and MC distributions of kinematic variables and doesn't provide any information about the shape of QCD contribution, whose contribution has been estimated as described in Section 4.5.1.

5.3.1 2011 results

Table 5.3, 5.4 show expected and observed yields for 2011 analysis.
In 2011 a correction factor F of 0.99 (0.92) for the muon (electron) channel

was obtained. Figure 5.2 shows the observed and expected distribution of some relevant quantities for basic corrected and tight selection, respectively for muon and electron analysis. The SM uncertainty band in the middle and bottom plots includes statistical and systematic uncertainties: the former is the statistical uncertainty on the simulated expected yield in the tight selection; the latter is computed under the assumption of constant and fully correlated relative uncertainty across all the bins; in Figure 5.2 (bottom) the signal distribution has been stacked assuming a fitted branching ratio of 0.003.

Dataset	Cross Section (pb)	BASIC	corrected BASIC	TIGHT
$t\bar{t}$	157.5	7800 ± 2000	7720 ± 940	584 ± 81
$W + jets$	31310	1290 ± 770	1290 ± 770	76 ± 42
$Z + jets$	3048	180 ± 110	180 ± 110	36 ± 20
tW	15.7	233 ± 61	230 ± 55	12.8 ± 1.8
$t - channel$	64.6	45 ± 27	45 ± 27	2.3 ± 1.3
$s - channel$	4.63	4.8 ± 2.9	4.8 ± 2.9	0.26 ± 0.14
WW	43.0	9.9 ± 6.0	9.9 ± 6.0	0.97 ± 0.53
WZ	18.2	6.7 ± 4.0	6.7 ± 4.0	0.92 ± 0.51
ZZ	5.9	1.24 ± 0.75	1.24 ± 0.75	0.32 ± 0.18
$t\bar{t}W$	0.16	26 ± 16	26 ± 16	2.0 ± 1.1
QCD	-	35 ± 35	35 ± 35	9.0 ± 9.0
Total Exp.	-	9600 ± 2200	9544 ± 98	724 ± 39
Data	-	9544 ± 98	9544 ± 98	796 ± 28

Table 5.3: 2011 Muon analysis: expected and observed yields for the basic and tight selection with an assumed BR value of zero. The "basic-corrected" and tight columns report the yields in the basic and tight selection, respectively, after the normalization procedure described in the text and applied only to the $t\bar{t}$ and tW processes.

In 2011 analysis the shape of QCD for E_T^{miss} and lepton p_T has been taken from a selection of lepton + 3 jets, whereas the shape of QCD contribution in the χ^2 and BNV top invariant mass distributions is taken from Z+jet sample. Looking at Table 5.3 and 5.4 a slight excess of data over expectation of about 10% (6%) in muon (electron) channel has been found. Such an excess has been interpreted as an hint of existence of BNV decay. However, since this analysis requires a good modeling of jet and E_T^{miss} behavior, many cross checks have been performed to test the reliability of simulation, exploring phase space regions as similar as possible to the signal one. Further details will be given in the next Chapter.
Central values of nuisance parameters and the associated statistic uncertainties have been reported in Table 5.5.

A 95% CL upper limit on BR for BNV decay process has been set using CLs construction [91].

Dataset	Cross Section (pb)	BASIC	corrected BASIC	TIGHT
$t\bar{t}$	157.5	6800 ± 1700	6390 ± 670	497 ± 72
$W + jets$	31310	1130 ± 510	1130 ± 510	88 ± 35
$Z + jets$	3048	280 ± 120	280 ± 120	82 ± 33
tW	15.7	188 ± 50	178 ± 42	14.6 ± 2.1
$t - ch$	64.6	38 ± 17	38 ± 17	3.2 ± 1.3
$s - ch$	4.63	3.9 ± 1.8	3.9 ± 1.8	0.30 ± 0.12
WW	43.0	8.4 ± 3.8	8.4 ± 3.8	0.80 ± 0.32
WZ	18.2	6.8 ± 3.1	6.8 ± 3.1	1.10 ± 0.44
ZZ	5.9	1.67 ± 0.75	1.67 ± 0.75	0.37 ± 0.15
$t\bar{t}W$	0.16	23 ± 10	23 ± 10	1.77 ± 0.71
QCD	-	370 ± 190	370 ± 190	109 ± 54
Total Exp.	-	8800 ± 1900	8425 ± 92	798 ± 66
Data	-	8425 ± 92	8425 ± 92	843 ± 29

Table 5.4: 2011 Electron analysis: expected and observed yields in the basic and tight selection. The assumed BR value is zero. The "basic-corrected" and tight columns report the yields in the basic and tight selection, respectively, after the normalization procedure described in the text and applied only to the $t\bar{t}$ and tW processes.

Limits

As mentioned before, a slight excess in data has been found in 2011 analysis, both in electron and muon channel. The upper limit on BNV decay BR in muon and electron channel have been set independently, maximizing the likelihood function in Equation B.1 . For the combination of muon and electron channel the product of likelihoods has been maximized, while any nuisance parameters has been considered fully correlated in the two analysis. The 95% upper and the expected limits have been set using an Hybrid-Bayesian prescription [94] and they are reported in Table 5.6.
The computed upper limit is 0.0076 (0.0072) for muon (electron) channel and 0.0067 for the combination.

5.3.2 2012 results

Table 5.7, 5.8 show expected and observed yields for 2012 analyses.
In 2012 a correction factor F of 1.04 (1.05) for the muon (electron) channel was obtained, while γ+jets contribution has been accounted in the total QCD yield. Figures 5.4, 5.5 and 5.6 show the observed and expected distribution of

Quantity	Muon analysis	Electron analysis
$\epsilon^B_{SM,SM}$	$(1.00 \pm 0.20) \times 10^{-2}$	$(8.6 \pm 1.7) \times 10^{-3}$
$\epsilon^T_{SM,SM}$	$(7.5 \pm 1.5) \times 10^{-4}$	$(6.7 \pm 1.4) \times 10^{-4}$
$\epsilon^B_{BNV,SM}$	$(6.6 \pm 1.3) \times 10^{-2}$	$(6.1 \pm 1.2) \times 10^{-2}$
$\epsilon^T_{BNV,SM}$	$(2.10 \pm 0.52) \times 10^{-2}$	$(1.96 \pm 0.49) \times 10^{-2}$
$\epsilon^B_{BNV,BNV}$	$(5.5 \pm 1.2) \times 10^{-2}$	$(5.9 \pm 1.3) \times 10^{-2}$
$\epsilon^T_{BNV,BNV}$	$(1.76 \pm 0.45) \times 10^{-2}$	$(1.76 \pm 0.45) \times 10^{-2}$
ϵ^B_{SM}	$(2.97 \pm 0.76) \times 10^{-3}$	$(2.40 \pm 0.62) \times 10^{-3}$
ϵ^T_{SM}	$(1.66 \pm 0.56) \times 10^{-4}$	$(1.98 \pm 0.62) \times 10^{-4}$
ϵ^B_{BNV}	$(2.20 \pm 0.57) \times 10^{-2}$	$(1.91 \pm 0.50) \times 10^{-2}$
ϵ^T_{BNV}	$(7.16 \pm 0.20) \times 10^{-2}$	$(6.61 \pm 0.19) \times 10^{-2}$
N^B_{bck}	1.600 ± 720	1900 ± 840
N^T_{bck}	130 ± 53	290 ± 120
N^B_{obs}	9544 ± 98	8425 ± 92
σ_{tW}	15.8 ± 1.3 pb	
$\sigma_{t\bar{t}}$	157.5 ± 23.6 pb	

Table 5.5: 2011 analysis: central values and associated overall uncertainties for the quantities appearing in the likelihood function.

	95% CL Upp. lim.	Exp. lim.	68% exp. lim. range
Muon ch.	0.0076	0.0044	[0.0028, 0.0057]
Electron ch.	0.0072	0.0054	[0.0035, 0.0087]
Combined	0.0067	0.0041	[0.0027, 0.0060]

Table 5.6: Observed 95% CL upper limit on BR, expected median 95% CL limit for the $BR = 0$ hypothesis and ranges that are expected to contain 68% of all observed excursions from the expected median for the muon and electron channels and for their combination.

some relevant quantities for basic, basic corrected and tight selection, respectively for muon and electron analysis. In basic and tight selection error bands accounting for total uncertainty have been overlaid. The relative systematic uncertainty on the total expected yield in basic (tight), namely corresponds to 21% (20%) for muons (electrons) in basic, and 5% (5%) in tight. Systematic uncertainty on simulated $t\bar{t}$ and tW yield, as well as the total expected yield have been calculated as explained in Section 5.3.2. The uncertainty arising from WW, WZ, ZZ, s-channel, t-channel, ttW and ttZ has been conservatively set to 50% of the respective yield. The signal shapes, assuming a branching ratio of 0.005 for BNV decay, have been superimposed. The QCD shape for E_T^{miss} distribution has been taken from the simulation; for the other plots, QCD shape has been taken from data in anti-isolation region, after subtrac-

tion of top and electro-weak component (estimated from simulation).
Looking at Tables 5.7 and 5.8 no significant excess can be seen both in basic and tight selection. Actually, the slight excess observed in 2011 disappears completely. The reason that could most likely explain these results could be found in a better control of systematic uncertainties in 2012, which has made possible a nicer agreement between data and simulation. Of course we cannot exclude completely that new possible physics may be suppressed by the tighter jet threshold, implemented in 2012.
Central values of nuisance parameters and the associated statistic uncertainties have been reported in Table 5.9.

Dataset	Cross Section (pb)	BASIC	corrected BASIC	TIGHT
$t\bar{t}$	234	36900 ± 8900	38600 ± 3600	2200 ± 220
$W + jets$	37500	6300 ± 3200	6300 ± 3200	230 ± 120
$Z + jets$	3500	380 ± 190	380 ± 190	32 ± 18
tW	22.2	1160 ± 180	1210 ± 280	51 ± 12
$t - channel$	87.1	250 ± 130	250 ± 130	5.7 ± 3.0
$s - channel$	5.55	31 ± 16	31 ± 16	0.84 ± 0.52
WW	54.8	86 ± 43	86 ± 43	3.1 ± 1.7
WZ	33.2	41 ± 21	41 ± 21	1.43 ± 0.78
ZZ	17.7	5.5 ± 2.8	5.5 ± 2.8	0.49 ± 0.28
$t\bar{t}W$	0.23	128 ± 64	128 ± 64	5.9 ± 3.0
$t\bar{t}Z$	0.17	79 ± 40	79 ± 40	4.1 ± 2.1
QCD	-	790 ± 550	790 ± 550	119 ± 59
Total Exp.	-	46000 ± 10000	47951 ± 220	2660 ± 130
Data	-	47951 ± 220	47951 ± 220	2614 ± 51

Table 5.7: 2012 Muon analysis: adopted cross section values, expected and observed yields for the basic and tight selection in the muon analysis with an assumed BR value of zero. The "basic-corrected" and "tight" columns report the yields in the basic and tight selection, respectively, after the normalization procedure described in the text and applied only to the $t\bar{t}$ and tW processes. Reported uncertainties are statistical plus systematic, added in quadrature.

Determination of systematic uncertainties on the expected yields

In Table 5.8 and 5.7 the expected yields for muon and electron channel have been reported with their full (statistical plus systematic, added in quadrature) uncertainty.
As already mentioned a relative total uncertainty of 50% has been conservatively assumed for non top-background yields.

Dataset	Cross Section (pb)	BASIC	corrected BASIC	TIGHT
$t\bar{t}$	234	36400 ± 8600	38200 ± 3600	2030 ± 210
$W + jets$	37500	6500 ± 3300	6500 ± 3300	240 ± 120
$Z + jets$	3500	760 ± 380	760 ± 380	85 ± 45
tW	22.2	1110 ± 170	1170 ± 220	37.3 ± 7.5
$t - channel$	87.1	230 ± 120	230 ± 120	6.6 ± 3.6
$s - channel$	5.55	27 ± 14	27 ± 14	0.70 ± 0.50
WW	54.8	78 ± 39	78 ± 39	3.7 ± 2.0
WZ	33.2	45 ± 23	45 ± 23	2.1 ± 1.1
ZZ	17.7	11.1 ± 5.6	11.1 ± 5.6	1.40 ± 0.70
$t\bar{t}W$	0.23	132 ± 66	132 ± 66	6.2 ± 3.1
$t\bar{t}Z$	0.17	86 ± 43	86 ± 43	4.4 ± 2.2
QCD	-	2900 ± 1500	2900 ± 1500	330 ± 170
Total Exp.	-	48300 ± 10400	50108 ± 220	2740 ± 160
Data	-	50108 ± 220	50108 ± 220	2703 ± 52

Table 5.8: 2012 Electron analysis: adopted cross section values, expected and observed yields for the basic and tight selection in the muon analysis with an assumed BR value of zero. The "basic-corrected" and "tight" columns report the yields in the basic and tight selection, respectively, after the normalization procedure described in the text and applied only to the $t\bar{t}$ and tW processes. Reported uncertainties are statistical plus systematic, added in quadrature.

Quantity	Muon analysis	Electron analysis
$\epsilon^{B}_{SM,SM}$	$(8.1 \pm 1.5) \times 10^{-3}$	$(8.0 \pm 1.5) \times 10^{-3}$
$\epsilon^{T}_{SM,SM}$	$(4.62 \pm 0.93) \times 10^{-4}$	$(4.24 \pm 0.85) \times 10^{-4}$
$\epsilon^{B}_{BNV,SM}$	$(7.37 \pm 0.89) \times 10^{-2}$	$(7.33 \pm 0.88) \times 10^{-2}$
$\epsilon^{T}_{BNV,SM}$	$(1.86 \pm 0.32) \times 10^{-2}$	$(1.62 \pm 0.27) \times 10^{-2}$
$\epsilon^{B}_{BNV,BNV}$	$(1.00 \pm 0.16) \times 10^{-2}$	$(1.55 \pm 0.25) \times 10^{-2}$
$\epsilon^{T}_{BNV,BNV}$	$(1.74 \pm 0.32) \times 10^{-3}$	$(2.64 \pm 0.55) \times 10^{-3}$
ϵ^{B}_{SM}	$(2.68 \pm 0.32) \times 10^{-3}$	$(2.57 \pm 0.31) \times 10^{-3}$
ϵ^{T}_{SM}	$(1.13 \pm 0.14) \times 10^{-4}$	$(8.21 \pm 0.99) \times 10^{-5}$
ϵ^{B}_{BNV}	$(2.72 \pm 0.42) \times 10^{-2}$	$(2.80 \pm 0.42) \times 10^{-2}$
ϵ^{T}_{BNV}	$(5.38 \pm 0.84) \times 10^{-3}$	$(5.84 \pm 0.82) \times 10^{-3}$
N^{B}_{bck}	8100 ± 3400	10800 ± 3800
N^{T}_{bck}	400 ± 140	680 ± 230
N^{B}_{obs}	47951 ± 220	50108 ± 220
σ_{tW}	22.2 ± 1.8 pb	
$\sigma_{t\bar{t}}$	234 ± 35 pb	

Table 5.9: 2012 analysis: central values and associated overall uncertainties for the quantities appearing in the likelihood function.

For top contributions, namely $t\bar{t}$ and tW, uncertainties must reflect the normalization method described in Section 4.3 and the approach to treat correlated systematic uncertainties described in Section 5.1.2. As show in the following, normalization method allows to reduce the uncertainty on the simulated yield. In the basic selection, the $t\bar{t}$ and tW yields are simply calculated as

$$N_X^B = L \times \sigma_X \times \epsilon_X^B(BR) \tag{5.12}$$

being L the total luminosity, σ the production cross section of the considered process (either $t\bar{t}$ or tW), $\epsilon_X^B(BR)$ the efficiency to pass basic selection (Table 5.9) and X the nature of the process ($t\bar{t}$ or tW). However $\epsilon_X^B(BR)$ is given by the linear sum of all the shift from the nominal efficiency value given by systematic uncertainties: thus the total uncertainty on N_X^B can be calculated sampling repeatedly nuisance parameters from their own pdfs, and calculating the new value of likelihood parameters. A distribution of N_X^B is obtained, whose width (gaussian σ) represents the total uncertainty on simulated yield. From Equation 4.6, one can easily obtain a distribution for $t\bar{t}$ corrected basic yield, whose value reads:

$$N_{corr}^B = \left(N_{obs}^B - N_{bck}^B\right) \times \left(\frac{1}{1 + \frac{\sigma_{tW}\epsilon_{tW}^B(BR)}{\sigma_{t\bar{t}}\epsilon_{t\bar{t}}^B(BR)}}\right) \tag{5.13}$$

While the $t\bar{t}$ tight yield reads:

$$N_{corr}^B = \left(N_{obs}^B - N_{bck}^B\right) \times \frac{1}{1 + \frac{\sigma_{tW}\epsilon_{tW}^B(BR)}{\sigma_{t\bar{t}}\epsilon_{t\bar{t}}^B(BR)}} \times \frac{\epsilon_{t\bar{t}}^T(BR)}{\epsilon_{t\bar{t}}^B(BR)} \tag{5.14}$$

The corrected basic and tight yield for tW can be found swapping $t\bar{t}$ and tW subscripts in Equation 5.14 and 5.13.
The systematic uncertainty on the total expected yield in basic and tight selection can be found accordingly: the former is calculated respectively summing all the top and non-top background yields; the latter can be calculated from Equation 4.6. In Figure 5.8 the total number of expected events in basic and tight selection has been reported for muon channel.

Limits

A 95% confidence level (C.L.) upper limit at has been set. For the combination of muon and electron channel the product of likelihoods has been maximized, while any nuisance parameter (except for QCD) have been considered fully correlated in the two analysis. The choice to correlate even lepton ID uncertainties has been proven to have a negligible impact on the final results, being the systematics largely dominated by jet energy scale uncertainties. The upper limits have been calculated using a Feldman-Cousins approach [92] (Table 5.10), using the prescription proposed in [93] . The computed upper limit is

	95% CL Upp. lim.	Exp. lim.	68% exp. lim. range
Muon channel	0.0016	0.0029	[0.0017, 0.0042]
Electron channel	0.0017	0.0029	[0.0017, 0.0044]
Combined	0.0015	0.0029	[0.0016, 0.0044]

Table 5.10: Observed 95% CL upper limit on BR, expected median 95% CL limit for the $BR = 0$ hypothesis and ranges that are expected to contain 68% of all observed excursions from the expected median for the muon and electron channels and for their combination. γ+jets contribution is not included.

0.0016 (0.0017) for muon (electron) channel and 0.0015 for the combination. Remarkably, even doubling the statistics, no improvements in the expected limit calculation have been observed: actually sources of systematic uncertainties have been considered fully correlated and any over- (under-)fluctuation in muon channel have been expected to be found in the electron channel. Being 2012 analysis dominated by systematic uncertainties, no significant improvements are expected increasing the available statistics.
Upper limits have been calculated also according to Hybrid-Bayesian, Frequentist and Asymptotic approach. In these methods the same "one sided LHC-like" test statistic [95] [96] has been employed. However calculated limits show a difference of about 25% across the various method: while Feldman Cousins and Frequentist approaches give a shift of about one sigma from expected limit, Hybrid-Bayesian and Asymptotic observed limits are quite close to the expected values. This has been found to be a consequence of the methods used to handle systematic uncertainties; a more detailed explanation will be given in the next paragraph.

	95% CL Upp. lim.	Exp. lim.	68% exp. lim. range
Muon-Asymptotic	0.0016	0.0018	[0.0010, 0.0031]
Muon-Hybrid	0.0021	0.0021	[0.0016, 0.0031]
Muon-Frequentist	0.0017	0.0026	[0.0018, 0.0040]
Muon-FC	0.0016	0.0029	[0.0017, 0.0042]
Electron-Asymptotic	0.0019	0.0021	[0.0014, 0.0033]
Electron-Hybrid	0.0023	0.0023	[0.0017, 0.0032]
Electron-Frequentist	0.0019	0.0028	[0.0020, 0.0046]
Electron-FC	0.0017	0.0031	[0.0018, 0.0045]
Combined-Asymptotic	0.0016	0.0018	[0.0010, 0.0030]
Combined-Hybrid	0.0021	0.0021	[0.0015, 0.0030]
Combined-Frequentist	0.0017	0.0026	[0.0019, 0.0040]
Combined-FC	0.0015	0.0029	[0.0016, 0.0042]

Table 5.11: Observed 95% CL upper limit on BR, expected median 95% CL limit for the $BR = 0$ hypothesis and ranges that are expected to contain 68% of all observed excursions from the expected median for the muon and electron channels and for their combination. Results obtained with the asymptotic, hybrid and Frequentist CLs variants are presented as well as those obtained with the Feldman-Cousins prescription. γ+jets contribution is not included.

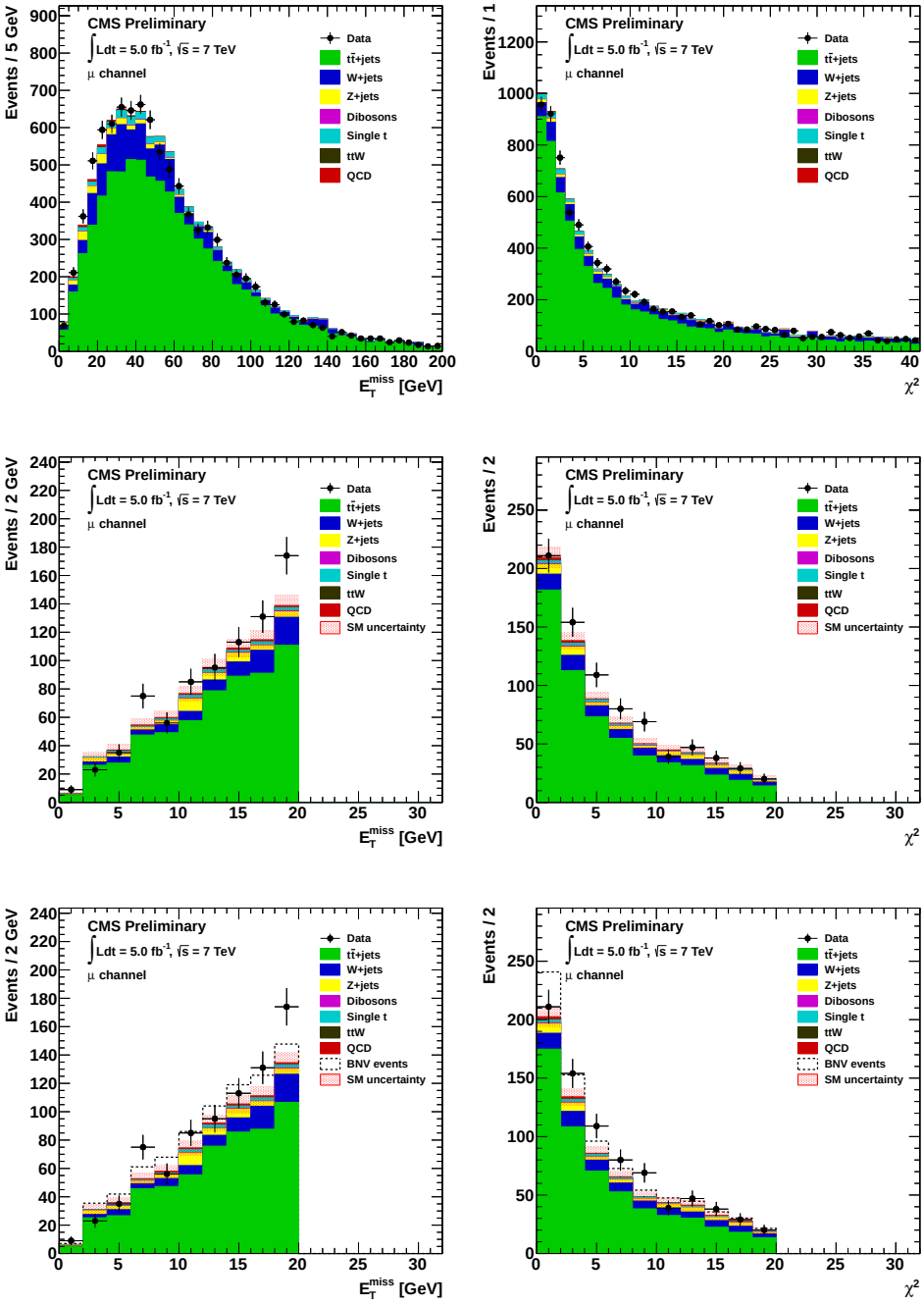


Figure 5.2: Muon analysis: distribution of E_T^{miss} (left) and χ^2 (right). Top: corrected basic selection with a null assumed BR value. Middle: tight selection with a null assumed BR value. Bottom: tight selection with an assumed BR value equal to 0.003.

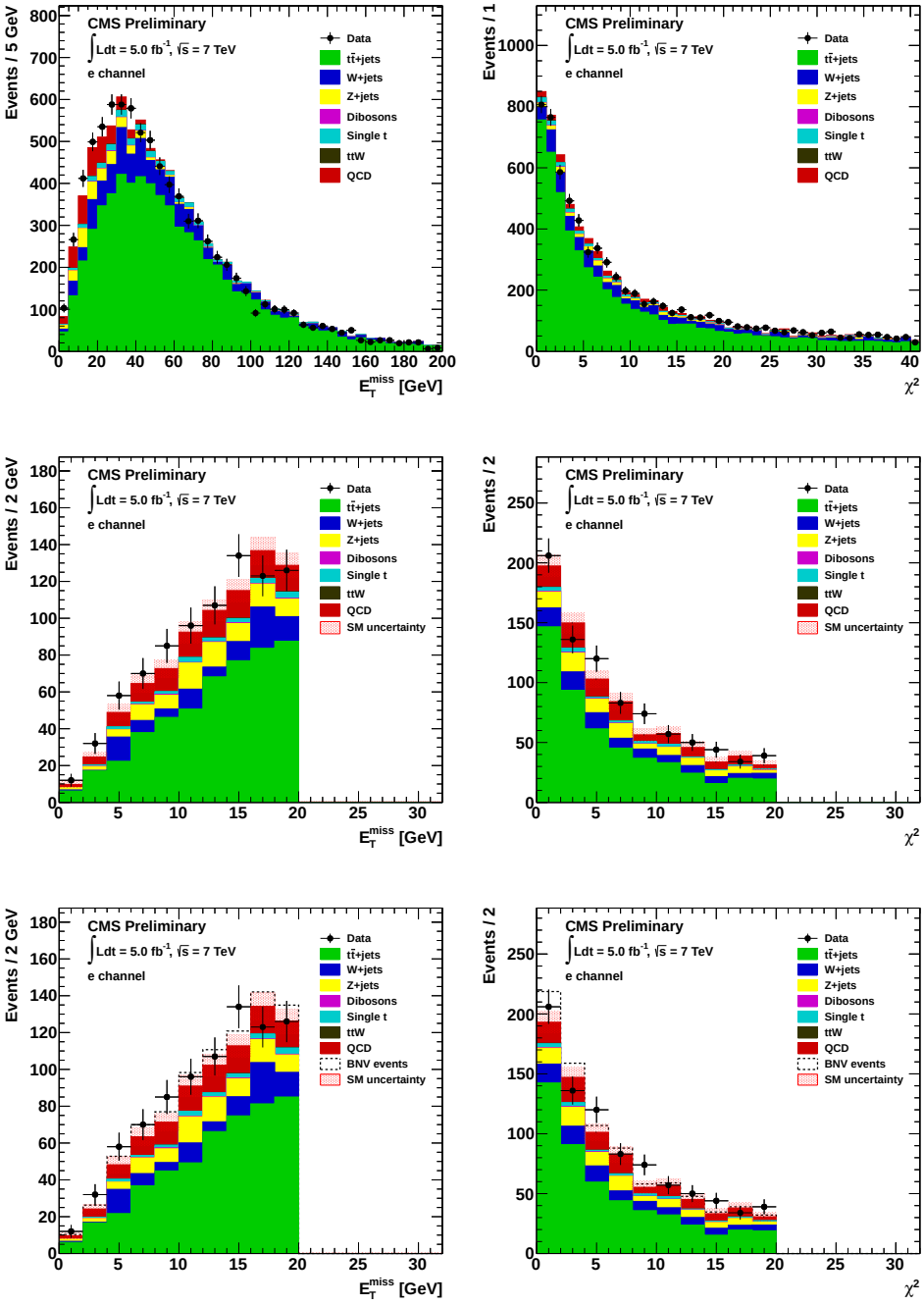


Figure 5.3: Electron analysis: distribution of E_T^{miss} (left) and χ^2 (right). Top: corrected basic selection with a null assumed BR value. Middle: tight selection with a null assumed BR value. Bottom: tight selection with an assumed BR value equal to 0.003.

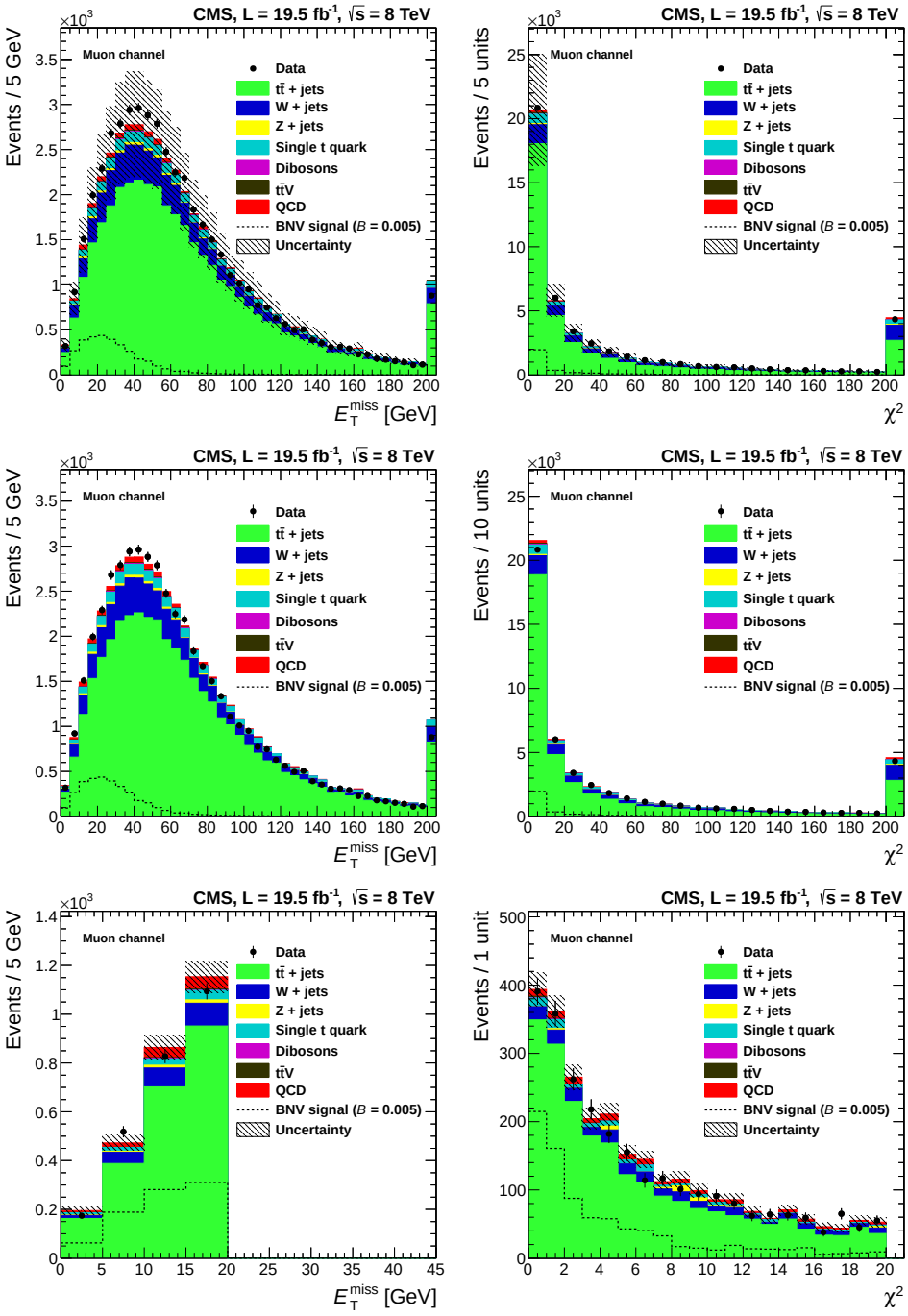


Figure 5.4: Muon analysis: distribution of E_T^{miss} (left) and χ^2 (right). for basic (top), basic corrected (middle) and tight selection (bottom). The signal contribution expected for $BR = 0.005$ is also shown.

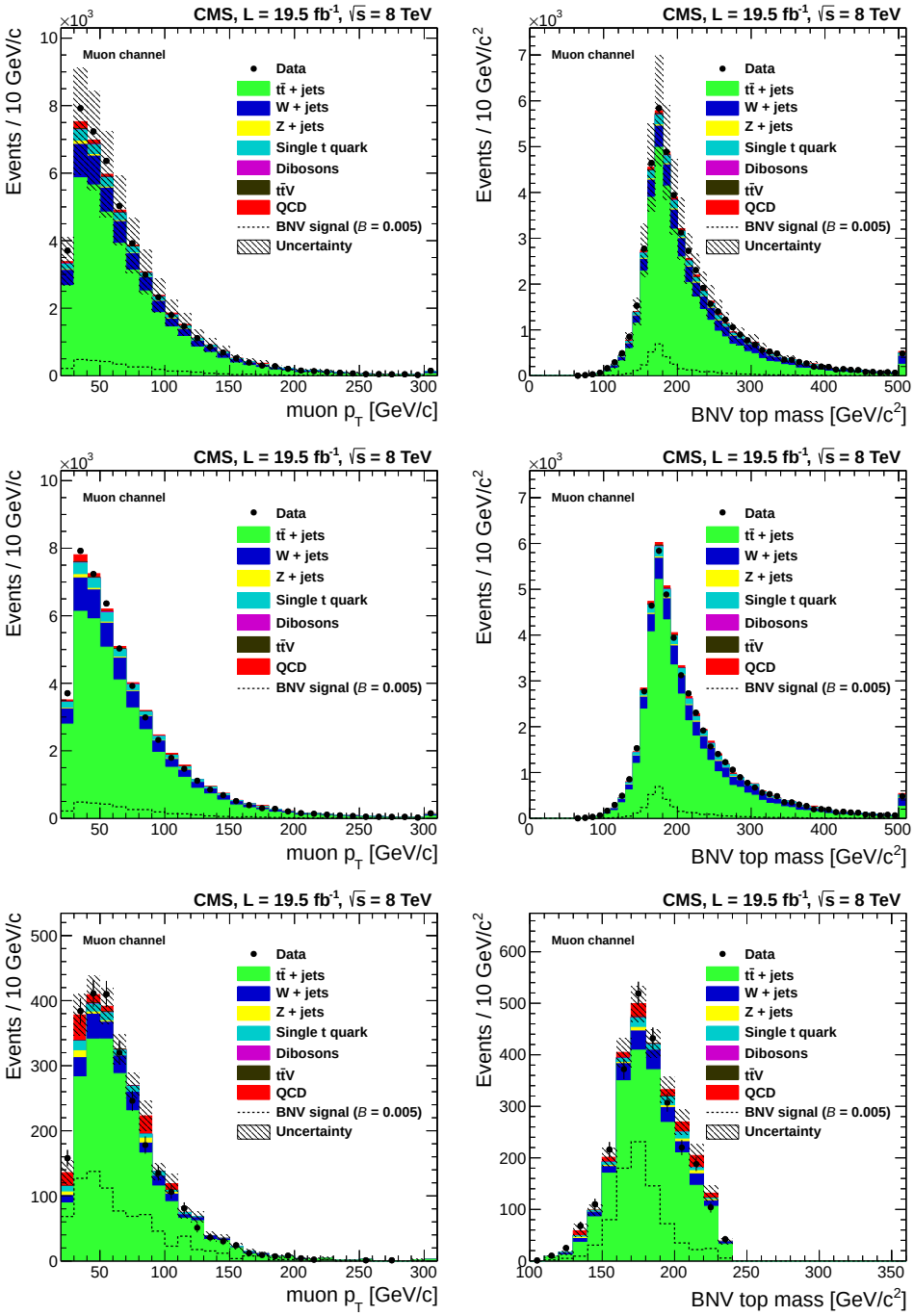


Figure 5.5: Muon analysis: distribution of muon p_T (left) and BNV top mass (right) for basic (top), basic corrected (middle) and tight selection (bottom). The signal contribution expected for $BR = 0.005$ is also shown.

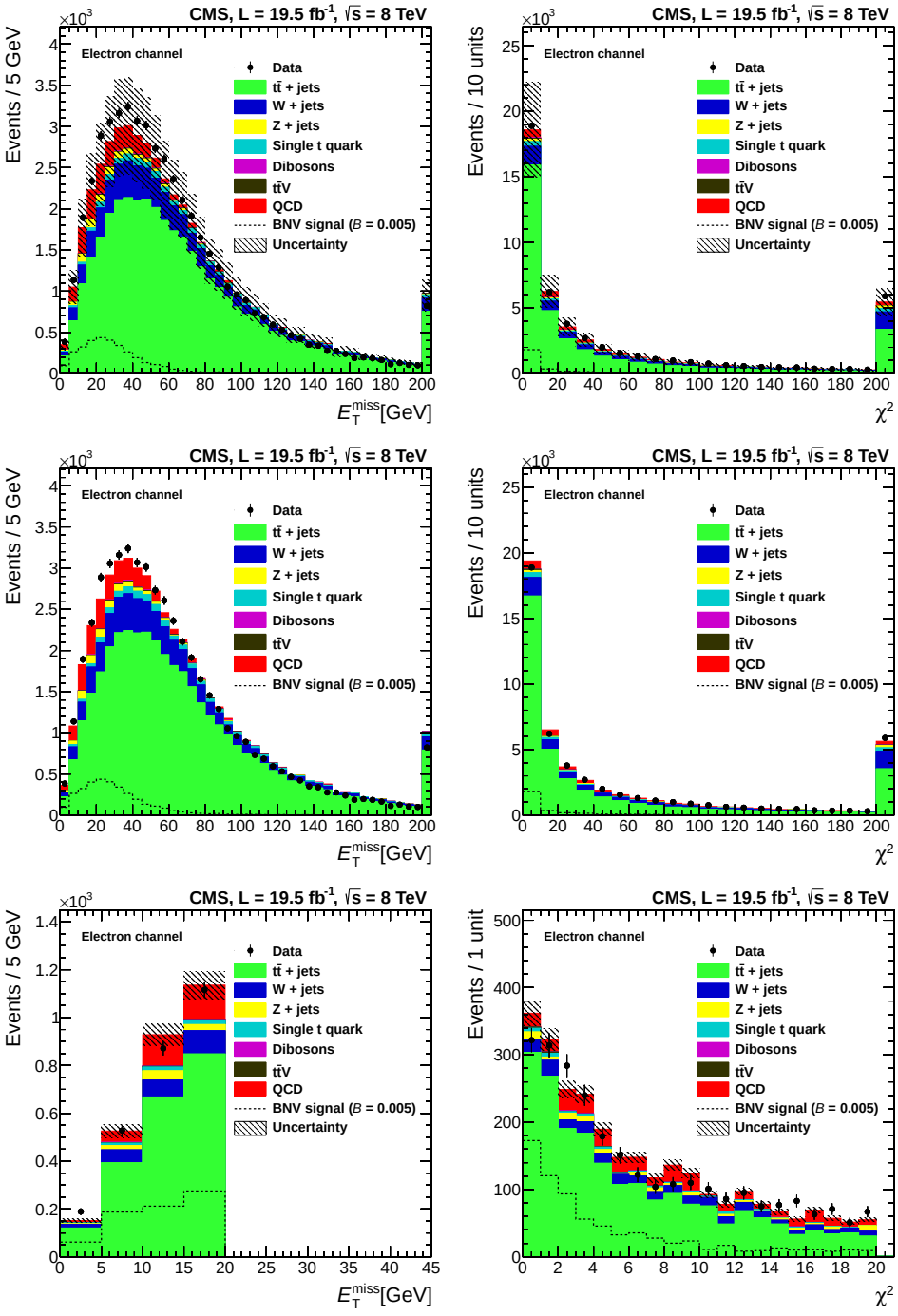


Figure 5.6: Electron analysis: distribution of E_T^{miss} (left) and χ^2 (right) for basic (top), basic corrected (middle) and tight selection (bottom). The signal contribution expected for $BR = 0.005$ is also shown.

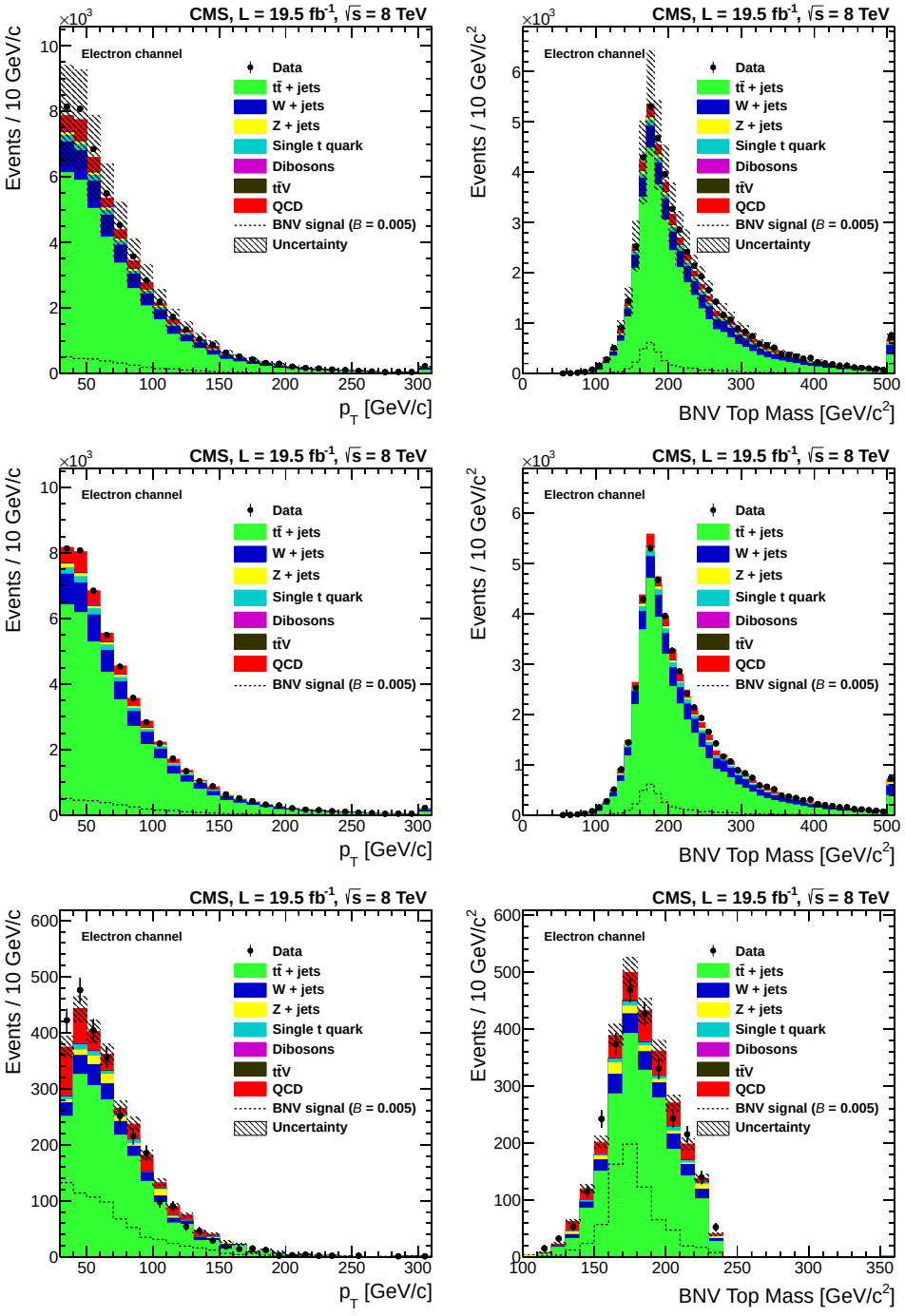


Figure 5.7: Electron analysis: distribution of electron p_T (left) and BNV top mass (right) for basic (top), basic corrected (middle) and tight selection (bottom). The signal contribution expected for $BR = 0.005$ is also shown.

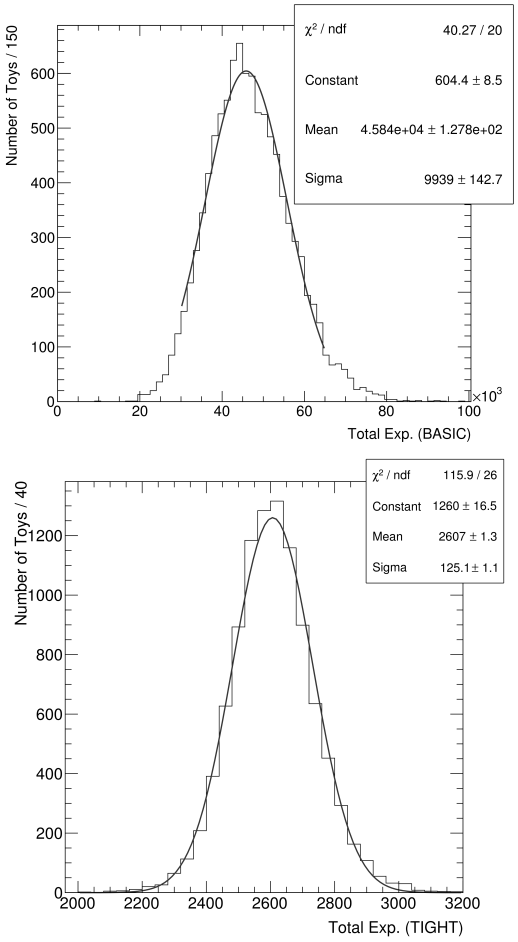


Figure 5.8: Distribution of the total number of expected events in basic (top) and tight (bottom) in muon channel. 10000 pseudo experiments have been generated, drawing nuisance values from their own pdf. The total uncertainty on the expected yield can be found summing in quadrature the distribution width and the statistical uncertainty.

Chapter 6

Cross checks

As shown in Chapter 5, a 2-sigma excess has been found in 2011 in data over SM expectations both in muon and electron channel. Many cross checks have been performed in order to analyze the origin of this excess. Even though no evident clues about possible sources of discrepancy have been found, these tests have been an interesting opportunity to stress the robustness of our analysis strategy against a number of systematic uncertainties.

The first cross checks were performed to test the robustness of the background yield estimate, with respect to cross section systematic uncertainties. The expected yields from $t\bar{t}$, tW, W+jets and Z+jets have been calculated using measured values of production cross sections at $\sqrt{s} = 7$ TeV, instead of theoretical ones. As shown in the following paragraphs, the differences found in the expected yield in the tight selection under these new conditions were totally negligible.

In addition, other test have been carried out, which can be classified as "low level" and "high level" cross checks. The former are the following:

1. In order to improve discrimination power between unphysical E_T^{miss} (due to QCD, jet energy resolution and geometrical acceptance effects) and physical E_T^{miss} (due to the presence of weakly interacting particles), the so-called E_T^{miss} Significance variable has been used as a discriminant variable. Negligible differences have been found with respect to our main strategy (more details will be given in Section 6.2.1).

2. Since the most significant discrepancy between expected and observed yield appears at low energy values, a mismodeling in the soft part of the E_T^{miss} spectrum, introduced by the E_T^{miss} *Type I* corrections, could be a possible cause of data excess. In Section 6.2.2 a cross check has been performed disabling such corrections, but no effect has been proven to justify the discrepancy between data and expectation.

In fact in the "low level" cross checks just statistical uncertainties have been reported. However one can quantify the excess of observed yield over SM expectations, assuming relative errors on the expected yield approximately equal to those estimated in Section 5.3.1 (namely about 8% in tight selection and about 20% in the basic).
On the other hand "high level" tests have been performed looking for any systematic uncertainties that were not under control. Actually a similar discrepancy should have been present even in signal free selections or whether the signal had a shape very similar to the background one. Phase space regions featuring such requirements have been analyzed to check the agreement between simulation and data for low E_T^{miss} values. Results are presented in the following paragraphs. No evident clues of E_T^{miss} mismodeling have been identified.

6.1 Cross section tests

In order to prove that analysis results are just weakly affected by cross section systematic uncertainty, in this section simulated yields have been recalculated with alternative choices of cross section values. These results refer to muon analysis and they use an older value of integrated luminosity (4.98 fb^{-1}), whose systematic uncertainty was 4.5%.

- **Theoretical cross section**: In Table 6.1 observed and simulated yields have been reported for $BR = 0$. For all the samples theoretical cross sections have been used. In this case a multiplicative factor F has been used to rescale $t\bar{t}$ and tW yield in basic, as we did in the main analysis (Section 5.3), under the assumption $N_{obs}^B = N_{exp}^B$. Such parameter, which basically gives a feeling of the agreement between observation and expectation in basic selection, turned out to be 1.06.

- **Measured cross section for** $t\bar{t}$: The yields reported in Table 6.1 have been reproduced in Table 6.2, using the CMS measured $t\bar{t}$ cross section. The yields of with the exception of W+jets and Z+jets, have been calculated using the theoretical cross section. The obtained F value is 1.01 and it has been used to rescale $t\bar{t}$ and tW contributions, as in the previous case. From a comparison between Table 6.2 with Table 6.1, negligible differences can be noted, proving the stability of the analysis strategy against $t\bar{t}$ cross section fluctuations.

- **Measured cross sections for dominant backgrounds**: In Table 6.3 observed and simulated yields have been reported for $BR = 0$. The cross sections of dominant backgrounds, namely $t\bar{t}$ [97], W+jets and Z+jets [98], have been set to their measured values (instead of the theoretical ones), while for the other backgrounds the cross sections shown in Section 5.3.1 have been employed. The computed F value was equal to 0.98

Table 6.1: Muon analysis cross-check: expected and observed yields for the basic and tight selection with an assumed BR value of zero and using the theory predicted cross section values for the SM processes. An older value of integrated luminosity with respect to the one employed in Section 5.3.1 has been used. The "corrected BASIC" and "TIGHT" columns report the yields in basic and tight selection, respectively, after the normalization procedure described in the text. Reported uncertainties are only statistical.

Dataset	BASIC	corrected BASIC	TIGHT
$t\bar{t}$	7351 ± 45	7806 ± 48	590 ± 13
W + jets	1211 ± 56	1211 ± 56	71 ± 13
Z + jets	171 ± 11	171 ± 11	33.4 ± 4.8
tW	219.1 ± 5.2	232.6 ± 5.6	13.0 ± 1.3
$t-channel$	42.5 ± 2.1	42.5 ± 2.1	2.14 ± 0.48
$s-channel$	4.47 ± 0.71	4.47 ± 0.71	0.24 ± 0.17
WW	9.34 ± 0.80	9.34 ± 0.80	0.91 ± 0.26
WZ	6.28 ± 0.47	6.28 ± 0.47	0.87 ± 0.17
ZZ	1.17 ± 0.12	1.17 ± 0.12	0.30 ± 0.06
ttW	24.63 ± 0.13	24.63 ± 0.13	1.89 ± 0.04
QCD	35 ± 21	35 ± 21	9.0 ± 5.1
Total Exp.	9076 ± 73	9544 ± 74	723 ± 19
Data	9544 ± 98	9544 ± 98	796 ± 28

Table 6.2: Muon analysis cross-check: expected and observed yields for the basic and tight selection with a BR assumed value of zero and using the theory predicted cross section values for all SM processes, with the exception of the $t\bar{t}$ cross section for which the measured value is used. The "corrected BASIC" and tight columns report the yields in the basic and "TIGHT" selection respectively, after the normalization procedure described in the text. Reported uncertainties are only statistical.

Dataset	BASIC	corrected BASIC	TIGHT
$t\bar{t}$	7739 ± 47	7817 ± 48	591 ± 13
W + jets	1211 ± 56	1211 ± 56	71 ± 13
Z + jets	171 ± 11	171 ± 11	33.4 ± 4.8
tW	219.1 ± 5.2	221.3 ± 5.3	12.4 ± 1.2
$t-channel$	42.5 ± 2.1	42.5 ± 2.1	2.14 ± 0.48
$s-channel$	4.47 ± 0.71	4.47 ± 0.71	0.24 ± 0.17
WW	9.34 ± 0.80	9.34 ± 0.80	0.91 ± 0.26
WZ	6.28 ± 0.47	6.28 ± 0.47	0.87 ± 0.17
ZZ	1.17 ± 0.12	1.17 ± 0.12	0.30 ± 0.06
ttW	24.63 ± 0.13	24.63 ± 0.13	1.89 ± 0.04
QCD	35 ± 21	35 ± 21	9.0 ± 5.1
Total Exp.	9464 ± 74	9544 ± 74	724 ± 19
Data	9544 ± 98	9544 ± 98	796 ± 28

and it has been used to rescale both $t\bar{t}$ and tW yield in basic. From a comparison between the total yield in tight selection reported in Table 6.3 with Table 6.1, a negligible differences can be noted (about 1%), proving the stability of the analysis strategy against the main backgrounds cross section fluctuations.

Table 6.3: Muon analysis cross-check: expected and observed yields for the basic and tight selection with an assumed BR value of zero and measured cross section values for $t\bar{t}$, W+jets and Z+jets. For the other SM processes the assumed cross section values are those reported in Section 5.3.1 The "corrected BASIC" and "TIGHT" columns report the yields in the basic and tight selection respectively, after the normalization procedure described in the text and applied only to the $t\bar{t}$ and tWprocesses. Reported uncertainties are only statistical.

Dataset	BASIC	corrected BASIC	TIGHT
$t\bar{t}$	7739 ± 47	7566 ± 46	572 ± 13
W + jets	1438 ± 66	1438 ± 66	85 ± 16
Z + jets	203 ± 13	203 ± 13	39.6 ± 5.7
tW	219.1 ± 5.2	214.2 ± 5.1	12.0 ± 1.2
$t-channel$	42.5 ± 2.1	42.5 ± 2.1	2.14 ± 0.48
$s-channel$	4.47 ± 0.71	4.47 ± 0.71	0.24 ± 0.17
WW	9.34 ± 0.80	9.34 ± 0.80	0.91 ± 0.26
WZ	6.28 ± 0.47	6.28 ± 0.47	0.87 ± 0.17
ZZ	1.17 ± 0.12	1.17 ± 0.12	0.30 ± 0.06
ttW	24.63 ± 0.13	24.63 ± 0.13	1.89 ± 0.04
QCD	35 ± 21	35 ± 21	9.0 ± 5.1
Total	9722 ± 83	9544 ± 82	724 ± 21
Data	9544 ± 98	9544 ± 98	796 ± 28

- **Normalization on W+jets**: In Table 6.4 the yield reported in Table 6.1 have been reproduced, using theoretical cross sections for all the simulated contributions. However just W+jets yield (representing the second dominant contribution in the expected yield) has been renormalized. Such an approach is equivalent to assume that the mismatch between data and expectation in basic is just due to a systematic error on the estimate of W+jets contribution.
  The computed F value turns out to be 1.386 and it is used to rescale only W+jets yield. Even in this case, there are no significant differences in the tight selection yield, with respect Table 6.1

Table 6.4: Muon analysis cross-check: expected and observed yields for the basic and tight selection with a BR assumed value of zero and using the their predictions for the SM cross section values. The "corrected BASIC" and "TIGHT" columns report the yields in the basic and tight selection, respectively, after the normalization procedure described in the text, but modified in such a way that it is applied to the W+jets process rather than to the $t\bar{t}$ and tW processes. Reported uncertainties are only statistical.

Dataset	BASIC	corrected BASIC	TIGHT
$t\bar{t}$	7351 ± 45	7351 ± 45	556 ± 12
W + jets	1211 ± 56	1679 ± 77	99 ± 18
Z + jets	171 ± 11	171 ± 11	33.4 ± 4.8
tW	219.1 ± 5.2	219.1 ± 5.2	12.2 ± 1.2
$t-channel$	42.5 ± 2.1	42.5 ± 2.1	2.14 ± 0.48
$s-channel$	4.47 ± 0.71	4.47 ± 0.71	0.24 ± 0.17
WW	9.34 ± 0.80	9.34 ± 0.80	0.91 ± 0.26
WZ	6.28 ± 0.47	6.28 ± 0.47	0.87 ± 0.17
ZZ	1.17 ± 0.12	1.17 ± 0.12	0.30 ± 0.06
ttW	24.63 ± 0.13	24.63 ± 0.13	1.89 ± 0.04
QCD	35 ± 21	35 ± 21	9.0 ± 5.1
Total Exp.	9076 ± 73	9544 ± 90	716 ± 23
Data	9544 ± 98	9544 ± 98	796 ± 28

6.2 E_T^{miss} Cross checks

6.2.1 Cut on MET significance

A proper reconstruction of E_T^{miss} depends on several factors [108], namely detector resolution, reconstruction inefficiencies, instrumental defects, improper pattern recognition etc. Many of these contributions represents undesirable, instrumental contributions that one would like to suppress. A discriminating variable, named E_T^{miss} *significance*, can be calculated in order to quantify on an event by event basis the probability that the measured E_T^{miss} is a fluctuation from zero, arising from finite measurement resolution.
For two reconstructed object with a transverse missing energy ϵ_1 and ϵ_2 one can define the significance in term of the likelihood function for the total reconstructed transverse energy $\vec{\epsilon}$.

$$\mathcal{L}(\vec{\epsilon}) = \int \mathcal{L}_1(\vec{\epsilon_1})\mathcal{L}_2(\vec{\epsilon_2})\delta(\vec{\epsilon} - (\vec{\epsilon_1} + \vec{\epsilon_2}))d\vec{\epsilon_1}\vec{\epsilon_2} \tag{6.1}$$

The formulation in terms of likelihood is completely general: however, even if any likelihood distributions $\mathcal{L}_i(\vec{\epsilon_i})$ can be used, one can usually assume a gaussian to model errors for any measured quantities. A similar expression can be defined recursively if more than two reconstructed objects are present in the single event. The significance is then defined as

$$\mathcal{S} \equiv 2\ln\left(\frac{\mathcal{L}(\vec{\epsilon} = \sum \vec{\epsilon_i})}{\mathcal{L}(\vec{\epsilon} = 0)}\right) \tag{6.2}$$

If one chooses a coordinate system with the x axis aligned with E_T^{miss} instead of the CMS horizontal axis, it can be shown that $\mathcal{S} = E_T^2/\sigma^2$ (being sigma the uncertainty on E_T measurement). More details can be found [108].
Of course when particle flow algorithms are used, the definition of E_T^{miss} significance should consider the momentum of the reconstructed PF jets and the remaining PF particles, reconstructed but unclustered, as the main ingredient to compute PF E_T^{miss}.
Thus, a further cross check has been performed modifying tight selection cuts. A cut on $\mathcal{S}$, instead of E_T^{miss} value, has been applied in tight selection, in addition to the usual χ^2 cut. A scan over several significance thresholds has been performed, retaining the value ($\mathcal{S} = 2$) which provided the expected limit closest to the main analysis one. The observed and expected yields for basic and tight selection under the assumption of $BR = 0$ and requiring $\mathcal{S} < 2$, have been reported in Table 6.5; in Figure 6.1 E_T^{miss} significance distribution with and without cut at $\mathcal{S} = 2$ has been shown. Even exploiting the new definition on MET cut, a discrepancy of about 10% is still visible between data and expectation.

Table 6.5: Muon analysis: expected and observed yields for the basic and tight selection under the assumption of null BR, using the E_T^{miss} significance algorithm, with cut value at 2 in TIGHT. Theoretical cross sections values have been used for all MC samples. Reported uncertainties are only statistical.

Dataset	BASIC	corrected BASIC	TIGHT
$t\bar{t}$	7920 ± 48	7796 ± 47	740 ± 15
W + jets	1316 ± 60	1316 ± 60	95 ± 16
Z + jets	182 ± 12	182 ± 12	43.0 ± 5.7
tW	236.9 ± 5.6	233.2 ± 5.5	18.4 ± 1.6
$t-channel$	45.8 ± 2.3	45.8 ± 2.3	2.95 ± 0.59
$s-channel$	4.75 ± 0.75	4.75 ± 0.75	0.26 ± 0.18
WW	10.07 ± 0.86	10.07 ± 0.86	0.80 ± 0.25
WZ	6.88 ± 0.51	6.88 ± 0.51	0.95 ± 0.19
ZZ	1.24 ± 0.12	1.24 ± 0.12	0.39 ± 0.07
ttW	26.87 ± 0.14	26.87 ± 0.14	2.55 ± 0.04
QCD	35.00 ± 0.00	35.00 ± 0.00	10.80 ± 0.00
Total Exp.	9786 ± 78	9658 ± 77	915 ± 22
Data	9658 ± 98	9658 ± 98	1004 ± 32

6.2.2 Results with uncorrected E_T^{miss}

As already mentioned, the transverse missing energy is computed as the negative vectorial sum of the transverse momenta associated to all the particles [108]. Since jets must be corrected accounting for scale uncertainties, also E_T^{miss} should be corrected, too. A *Type I* particle flow E_T^{miss} must be recalculated, taking into account further corrections on jet p_T to ensure a flat response and matching simulated jet response to the observed one.
The presence of any systematic effects introduced by energy corrections has been investigated, switching off *Type I* corrections and using uncorrected PF E_T^{miss}. In Table 6.6 observed and expected yield have been reported. However it is evident that even the use of such corrections is not responsible for the observed discrepancy between data and simulation.

6.3 High level cross checks

Looking at tight selection Figures 5.3 and 5.2, a slight discrepancy between observed and expected counts affects low energy region of E_T^{miss} distribution. Such a discrepancy is the cause of the small data excess in tight selection, where events with $E_T^{miss} < 20$ GeV are selected. Indeed, since the dominant background is rejected by the E_T^{miss} cut, a possible source of excess could be given by a non accurate modeling of E_T^{miss} distribution at low energy. Therefore, phase space regions where the signal is suppressed have been studied

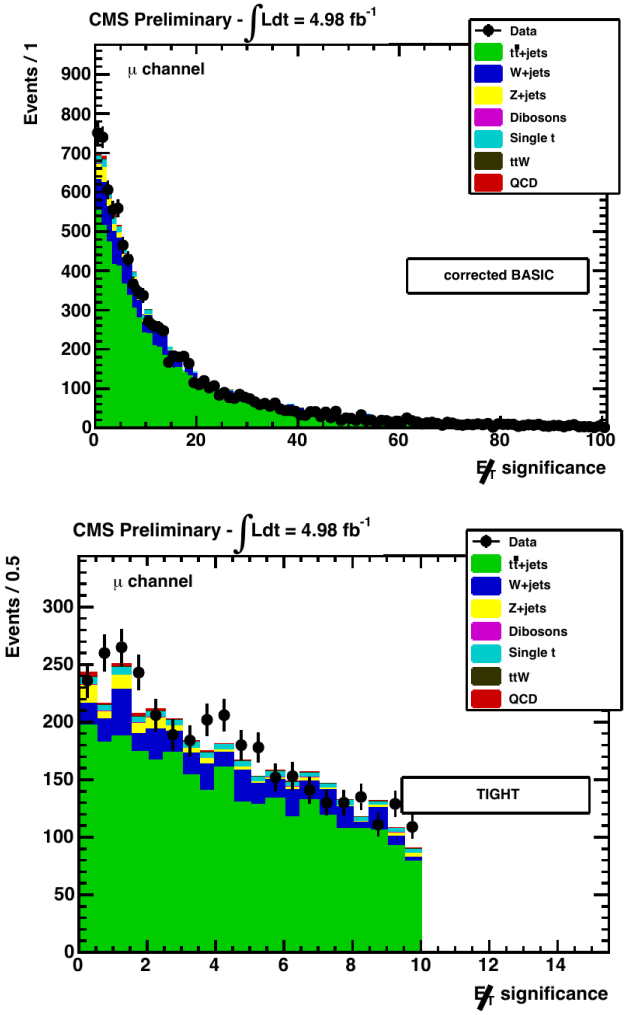


Figure 6.1: Muon analysis: E_T^{miss} significance distributions for the corrected BASIC (top) and TIGHT (bottom) selections under the assumption null BR, using the E_T^{miss} significance algorithm with cut value at 10.

comparing data and expectation, aiming to demonstrate that E_T^{miss} modeling was under control. Indeed, first high level cross checks results were performed in 2011, and were confirmed by further cross checks in 2012.

6.3.1 Single muon plus 3/4 jets selection

As a first cross check, a very similar phase space to the one described in the main analysis has been selected, performing a selection on 2011 data, at $\sqrt{s} = 7$ TeV.
One isolated muon with $p_T > 30$ GeV/c, imposing the same quality cuts used in the main analysis, has been required. The minimum number of jets with $p_T > 30$ GeV/c has been decreased from 5 to 3 (4). Being this selection dom-

Table 6.6: Muon analysis: expected and observed yields for the basic and tight selection under the assumption of null BR, using the uncorrected PF E_T^{miss}. Theoretical cross sections values have been used for all MC samples. Reported uncertainties are only statistical.

Dataset	BASIC	corrected BASIC	TIGHT
$t\bar{t}$	7844 ± 48	7742 ± 47	635 ± 14
W + jets	1301 ± 60	1301 ± 60	84 ± 15
Z + jets	182 ± 12	182 ± 12	41.2 ± 5.5
tW	234.2 ± 5.6	231.2 ± 5.5	15.8 ± 1.4
$t-channel$	45.7 ± 2.3	45.7 ± 2.3	2.34 ± 0.52
$s-channel$	4.75 ± 0.75	4.75 ± 0.75	0.26 ± 0.18
WW	10.01 ± 0.86	10.01 ± 0.86	0.88 ± 0.26
WZ	6.80 ± 0.50	6.80 ± 0.50	0.82 ± 0.17
ZZ	1.24 ± 0.12	1.24 ± 0.12	0.36 ± 0.07
ttW	26.37 ± 0.14	26.37 ± 0.14	2.07 ± 0.04
QCD	35.00 ± 0.00	35.00 ± 0.00	9.00 ± 0.00
Total Exp.	9691 ± 78	9586 ± 77	792 ± 21
Data	9586 ± 98	9586 ± 98	860 ± 29

inated by W+jets, the $t\bar{t}$ contribution is reduced
In Figure 6.2 E_T^{miss} distribution have been reported for a 3 jets selection, respectively in absence (left) and presence of signal (right); the other top contributions have been rescaled, taking into account a signal with $BR = 0.003$. Indeed for the 3 jets selection, an overall agreement between data and SM expectation has been found, being the presence of the signal not significant in this case. In Figure 6.3 the same distributions have been reported for a 4 jets selection. In the first bin at low E_T^{miss} a small excess in data seems visible instead. The QCD does not fill the gap between data and simulation and given the estimated order of magnitude of BR, the signal contribution is not totally negligible in this case.
Cross section values used for the normalization of simulated samples match the ones used in the main analysis, and the overall spectra have been normalized in order to appreciate the agreement between expected and observed distribution shapes.

6.3.2 Di-muon plus jets selection

A mismodeling of E_T^{miss} at low energy can be studied looking at processes where physical missing transverse momentum is absent, namely reducing any backgrounds which could provide neutrino emissions. In this scenario energy resolution, detector acceptance and objects misreconstruction are the unique ingredients that contribute to E_T^{miss} determination.

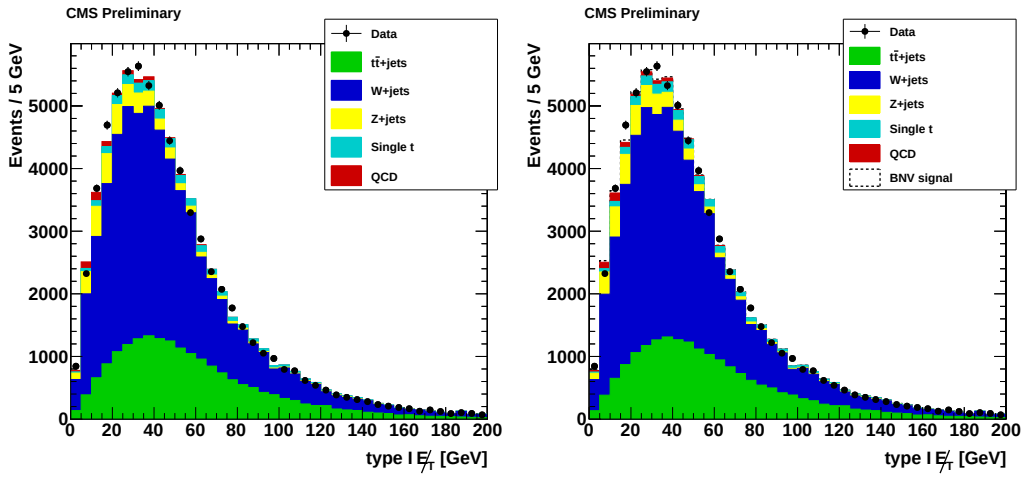


Figure 6.2: Distribution of E_T^{miss} for a muon + 3 jets selection in data and simulation for null hypothesis (left) and for $BR = 0.003$ (right). Signal has been stacked to SM simulation

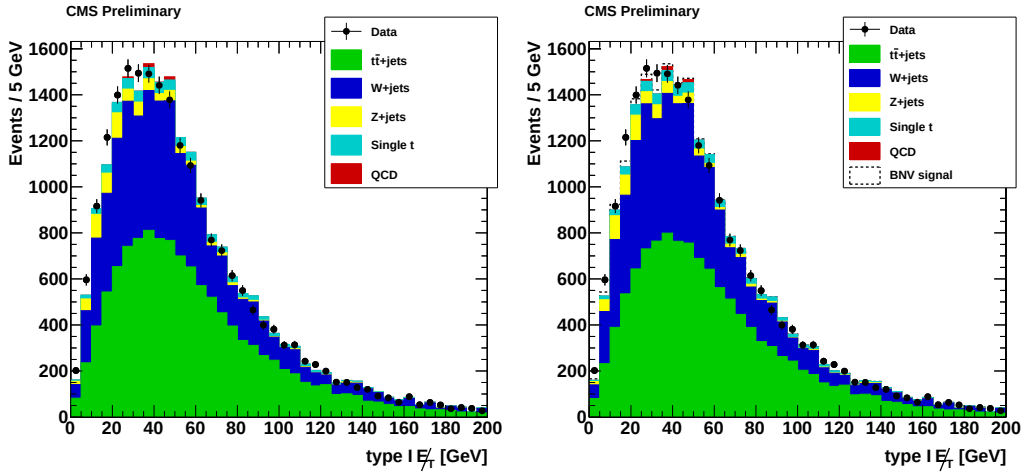


Figure 6.3: Distribution of E_T^{miss} for a muon + 4 jets selection in data and simulation for null hypothesis (left) and for BR = 0.005 (right). Signal has been stacked to SM simulation

The accuracy of E_T^{miss} description in simulated samples has been studied in 2011 using an integrated luminosity of more than 4.6 pb^{-1} ($\sqrt{s} = 7$ TeV). Drell-Yan events $Z \to \mu^+\mu^-$+jets have been selected. In such a selection exactly two isolated muons, with the same quality features specified in Section 4.4.1 have been required, though a tighter isolation cut ($I_{rel}^{\updownarrow} < 0.05$) has been set.
In order to select muons pairs coming from Z boson mass peak, di-muon invariant mass has been constrained in the range [76, 106] GeV/c^2. Moreover a

lepton veto has been applied, disregarding events having an additional electron (muon) with $p_T > 15$ (10) GeV/c in $|\eta| < 2.5$ (2.4) and $I^{\updownarrow}_{rel} < 0.2$. On top of that events having at least 3 jets, with $p_T > 30$ GeV/c in $|\eta| < 2.4$ have been selected. This selection is expected to be dominated by Z + jets selection, strongly reducing the impact of $t\bar{t}$ contribution.
In Figure 6.4 muon p_T distribution and E_T^{miss} distribution have been shown, this latter peaked at low energy values, as expected. In such a selection any contribution coming from a BNV signal is totally negligible. Cross section values used to normalize the different contributions are the same employed in the 2011 analysis. Finally simulated spectrum have been normalized to data, to emphasize possible shape deviations. The agreement between data and expectation is good.

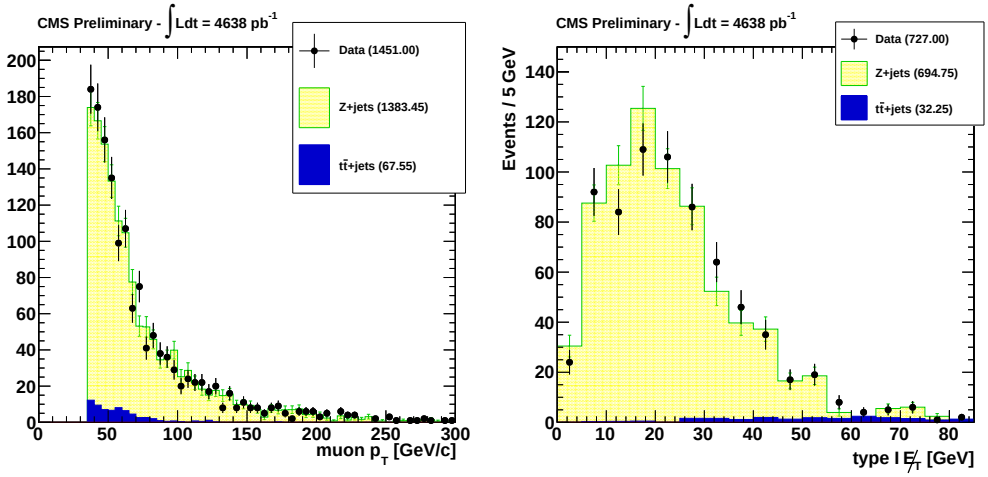


Figure 6.4: 2011 Analysis: distributions of muon p_T (left) and E_T^{miss} miss (right) in both data and simulation for a sample events enriched of Z+4 jets $\rightarrow \mu^+\mu^- + 4jets$ events. For both data and simulation the error bars indicate the statistical uncertainty.

In 2012 a very similar selection has been setup, applying the lepton quality cuts defined for 2012 muon analysis ($\sqrt{s} = 8$ TeV), and already described in Section 4.4.1. In this case four jets have been required, with a $p_T > 30$ GeV/c and $\eta < 2.4$. E_T^{miss} distribution has been shown in Figure 6.5. Even in this case a good agreement between observed and simulated E_T^{miss} distribution has been found even at low E_T^{miss}.

6.3.3 Di-leptonic $t\bar{t}$ selection

All the cross checks performed so far were mainly focused on low E_T^{miss} phase space region. However, $t\bar{t}$ contribution was never prevailing. A good way to study E_T^{miss} features in a region where $t\bar{t}$ fraction is dominant and the signal contribution is not expected to be significant, can be achieved performing a

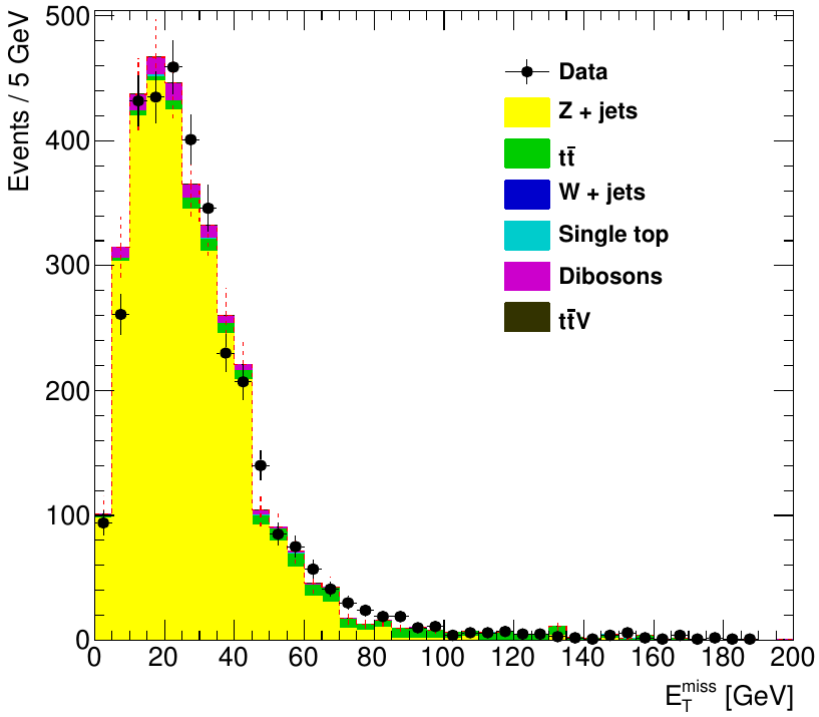


Figure 6.5: 2012 Analysis: distributions of E_T^{miss} miss in both data and simulation for a sample events enriched of Z+4 jets $\rightarrow \mu^+\mu^- + 4jets$ events. For both data and simulation the error bars indicate the statistical uncertainty.

di-lepton plus three jets selection. In principle such topology would be largely dominated by the presence of Drell Yan events, providing a Z boson decaying into two, opposite sign muon (electrons). However selection can be setup in order to maximize $t\bar{t}$+jets contribution: one can thus study directly properties of the main BNV background. Moreover, a b-tag discrimination can be applied, as well as in the main analysis, to reject non $t\bar{t}$ contributions.
Such a selection has been setup in 2012, on a subset of the total integrated luminosity (11.2 fb^{-1}) at $\sqrt{s} = 8$ TeV.
First, just dileptonic events with one isolated electron and one isolated muon have been required. This choice has been made to reject Drell-Yan production, without applying any cut on dilepton invariant mass. Events including one muon with $p_T > 30$ GeV/c in $|\eta| < 2.4$, plus one electron with $p_T > 30$ GeV/c have been selected.
Events with at least one additional loosely isolated lepton ($I_{rel}^{\updownarrow} < 0.2$) with $p_T > 20$ GeV/c in $\eta < 2.5$ have been discarded. At least three jets per event have been required, with $p_T > 30$ GeV/c and $|\eta| < 2.4$, including one loosely b-tagged jets (CSV algorithm threshold 0.266). The results of such selection is shown in Figure 6.6-6.8. BNV signal has been overlaid, assuming a branching fraction of 0.003, close to the value fitted in 2011.

As expected, the $t\bar{t}$ contribution is largely dominant: the first, second, third leading jet p_T distribution in data and in simulated samples are shown in Figure 6.6, while muon and electron p_T distributions have been shown in Figure 6.7. Observed and simulated overall spectra have been normalized in order to appreciate eventual deviations in shape. A good agreement has been observed between jets and leptons p_T distribution shapes, even though W+jets contribution is exposed to quite relevant fluctuations due to the low statistics.

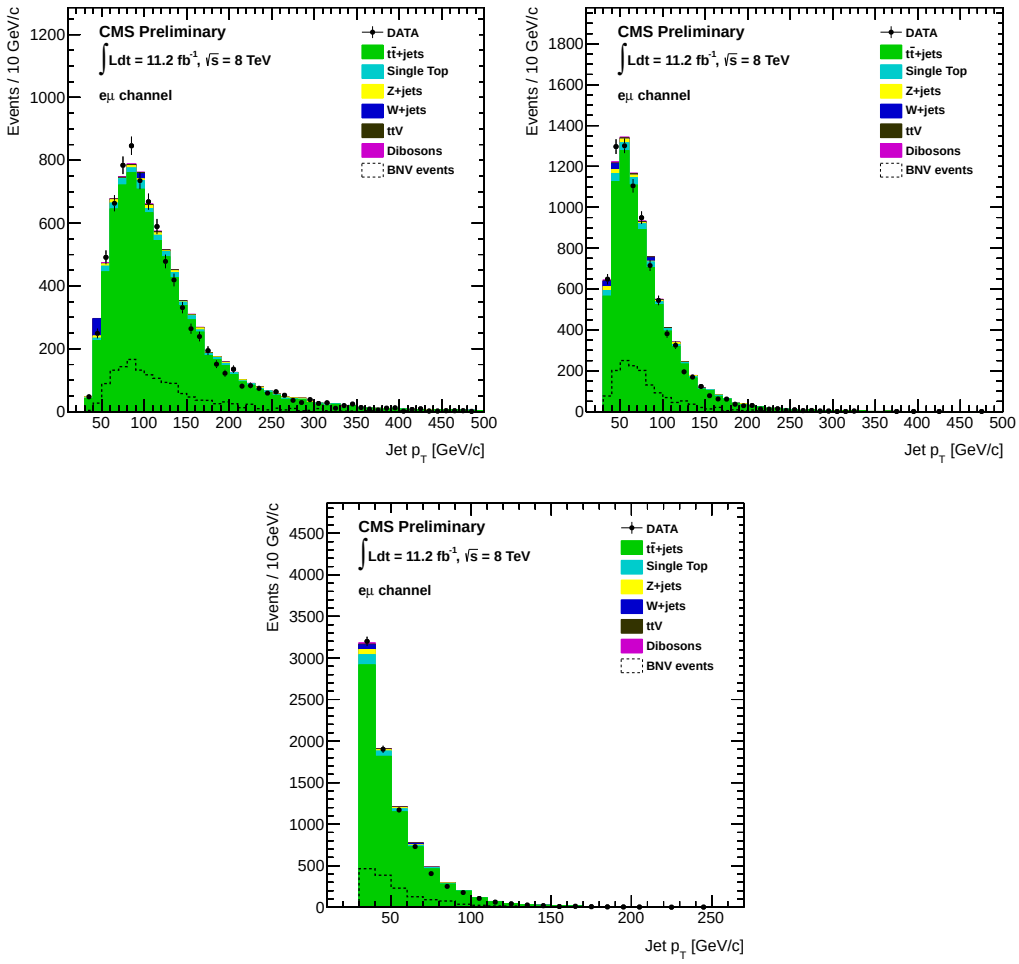


Figure 6.6: Dileptonic cross check: p_T distribution of the first, second and third leading jet in data and simulation, under the null hypothesis. Signal has been overlaid, assuming a branching ratio of 0.003.

In Figure 6.8 E_T^{miss} distribution has been shown. QCD contribution, expected to be small, is missing. Indeed the agreement between shapes is pretty good, even though a slight excess of 6.4% is still visible in the first two bins in E_T^{miss} distribution. Unfortunately, assuming a branching ratio of 0.003, the BNV contribution is not totally negligible even in this case and it is quite difficult to draw a conclusion about a possible mismodeling of simulated samples

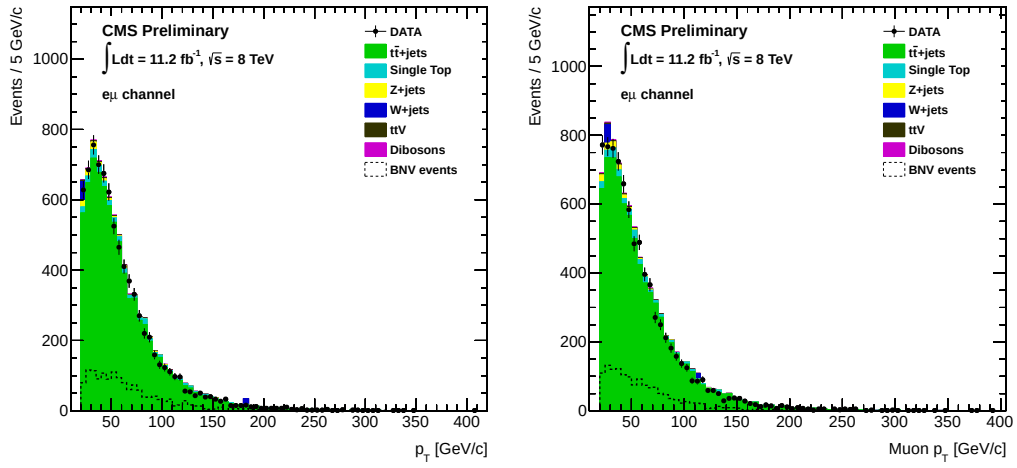


Figure 6.7: Dileptonic cross check: p_T distribution of the isolated muon and the isolated electron under the null hypothesis. Signal has been overlaid, assuming a branching ratio of 0.003.

for low E_T^{miss} values.

6.3.4 Isolated muon trigger selection

As explained in Section 4.2.1, in the 2012 analysis a HLT trigger requiring one isolated lepton and three jets with $p_T > 30, 40, 50$ GeV/c has been applied on data and simulated samples.
As already noticed, the set of cuts applied on the isolated leptons has not changed significantly in 2012 with respect to 2011 analysis. However in the first case an increase of jet transverse momentum offline thresholds was needed to match trigger jet p_T scale. Actually, as explained in Chapter 4.4.1, a minimum transverse momentum of 70, 55, 40 GeV/c has been required for the first three leading jet, beyond the minimum requirement of 5 jets with $p_T > 30$ GeV/c (as well as in 2011 analysis it has been done).
In such a way the analysis working point could be confined in the trigger efficiency plateau and the muon trigger efficiency could be considered constant and independent from jet p_T. However in principle an increase of jet p_T threshold might potentially suppress the presence of new physics resulting in any excess of soft jets.
In 2012 a further check has been performed on a subset of the total integrated luminosity (14.3 fb^{-1}) at $\sqrt{s} = 8$ TeV to exclude this possibility.
A selection identical to the main analysis one has been setup in muon channel. In fact a simple isolated muon trigger (muon $p_T > 25$ GeV/c) has been applied, without any requirements on jet transverse momentum (a lower muon p_T threshold of 30 GeV/c has been applied in the offline selection to match trigger threshold). This solution allowed to reproduce a selection even more

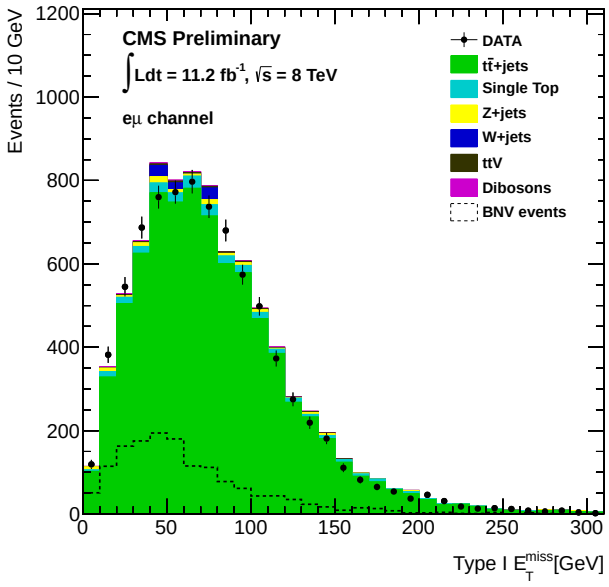


Figure 6.8: Dileptonic cross check: E_T^{miss} distribution of simulated and observed transverse missing energy in the null hypothesis. Signal has been overlaid, assuming a branching ratio of 0.003. An excess of 6.4% can be observed in the first two bins. Yet the contribution of BNV signal in the low energy bins is not negligible and does not allow to draw a definitive conclusion about the agreement between simulated and observed E_T^{miss} for very low missing energy values.

similar to the 2011 one (5 jets with $p_T > 30$ GeV/c), with no need of tighter cuts on jet energy scale. The numerical results have been reproduced in Table 6.7, where just statistical uncertainties have been reported. A very good agreement has been found in tight selection between the observed yield over the SM expectations. According to these results the hypothesis of a suppression of new physics processes yielding low momentum jets looks strongly unlikely.

6.3.5 Conclusion

For sake of shortness just the must significant cross checks performed in 2011 and 2012 have been shown in the previous paragraphs. The remaining - not shown - tests essentially confirmed the results that have been presented here. It is a matter of fact that the small excess observed in 2011, and confirmed by several cross checks, disappeared in 2012 analysis even though no significant changes have been performed in 2012 selection. The higher jet energy thresholds, that have been applied in 2012 analysis to match trigger requirements on jet energy scales, cannot explain the disappereance of the excess observed in 2011. On the other hand, a statistical overfluctuation of data over SM expectation could hardly explain our observations.

Dataset	Cross Section (pb)	BASIC	corrected BASIC	TIGHT
$t\bar{t}$	234	25840 ± 110	25970 ± 110	1693 ± 29
W + jets	37509	5070 ± 210	5070 ± 210	287 ± 49
Z + jets	3503.71	694 ± 34	694 ± 34	125 ± 14
tW	22.2	837 ± 16	841 ± 16	47.6 ± 3.9
t-channel	87.1	185.8 ± 6.4	185.8 ± 6.4	8.0 ± 1.3
s- channel	5.55	23.1 ± 2.1	23.1 ± 2.1	0.77 ± 0.39
WW	54.838	60.6 ± 2.3	60.6 ± 2.3	3.74 ± 0.56
WZ	33.21	33.1 ± 1.3	33.1 ± 1.3	2.40 ± 0.35
ZZ	17.654	9.84 ± 0.52	9.84 ± 0.52	1.59 ± 0.21
$t\bar{t}$W	0.232	81.0 ± 1.2	81.0 ± 1.2	4.26 ± 0.27
$t\bar{t}$Z	0.1743	52.15 ± 0.79	52.15 ± 0.79	3.11 ± 0.19
QCD	-	390 ± 85	390 ± 85	71 ± 17
Total Exp.	-	33270 ± 240	33410 ± 240	2247 ± 59
Data	-	33411 ± 183	33411 ± 183	2301 ± 48

Table 6.7: Muon analysis cross check: adopted cross section values, expected and observed yields for the basic and tight selection in the muon analysis with an assumed BR value of zero. A single isolated muon trigger has been applied, and the lower thresholds of 70, 55, 40 GeV/c on jet have been dropped off. The "basic-corrected" and "tight" columns report the yields in the basic and tight selection, respectively, after the normalization procedure described in the text and applied only to the $t\bar{t}$ and tW processes. Reported uncertainties are only statistical.

Therefore, even thought the exact motivation of the slight excess remain still unexplained at the moment, we are pretty confident that an higher accuracy in the modeling of systematic uncertainties in 2012 data represents the most likely explanation for what we observed.

Chapter 7

Conclusions

7.1 Analysis outlook

For many years huge efforts have been spent to prove the existence of processes that could violate baryon number conservation law. Many experiments have been carried out, aiming to find any evidence of proton decay, exotic decays of known particles (mesons, bosons or leptons...), which could support - directly or indirectly - the violation of baryon number violation hypothesis. However, as baryon number violation remains a fundamental ingredient of many popular physics theories, a remarkable attention is still addressed to this intriguing issue.

LHC represents an extraordinary possibility to probe the validity of many conservation laws at unprecedented energy scale. Hence, in the light of low energy experimental results, in recent times some models predicting baryon number violation at high energy scales have been proposed. In particular, due to the high production rate, the clear lepton signature and the presence of heavy flavor jets in the final state, top channel is largely considered as the best *trait d'union* between SM and new Physics. This work of thesis is based on the effective model presented in [7], where a new particle interaction has been introduced, allowing top quark decaying in one isolated leptons and two jets (one of them being a b-jet) without any neutrino in the final state. No particular assumptions have been made about the existence of any possible mediator.
Due to the low expected branching ratio, a search for BNV top decay was performed looking at $t\bar{t}$ pairs, where one top decays hadronically, while the other undergoes a baryon number violating decay. A counting experiment was carried out to reduce the complexity of the analysis. A strategy that was as model independent as possible was worked out and proved to be robust against a number of systematic uncertainties.
An analysis was performed both in 2011 (using 5 fb^{-1} of p-p collisions collected

by CMS Experiment at $\sqrt{s} = 7$ TeV) and in 2012 (19.5 fb^{-1} at $\sqrt{s} = 8$ TeV). Even though in 2012 analysis the strategy basically remained the same as 2011, some changes were implemented for what concerns statistical treatment and QCD estimation. Both analysis have been approved by CMS Collaboration.
The analysis started targeting a BASIC region, where the contribution of the signal was totally negligible and where the non top-background was reduced. As a second step, the expected yield was normalized to the observed one. Assuming that the mismatch between simulation and data was due to the uncertainties on top backgrounds, the equivalence between the expected and the observed number of events was reached just rescaling $t\bar{t}$ and tW contributions by a common factor. It's worth to note that such factor led to a variation of just few percents in $t\bar{t}$ and tW yield.
Finally a last TIGHT selection was performed to enhance the presence of the signal in data. A cut on the transverse missing energy and on a χ^2 variable (which was built from the difference between the nominal and the reconstructed top, anti-top and W boson mass values) was set. The total multi-jet yield was estimated in a data-driven way: in particular in 2012 the probability to reconstruct a genuine jet as a lepton was exploited, in order to find the contribution of QCD background.
All the cuts were tuned in such a way to maximize the expected limit on signal BR.

7.2 Results and discussion

Two independent analyses have been performed both in muon and in electron channel.
A small (not signficant) excess has been found in 2011 data over Standard Model expectations, while a very good agreement has been found in 2012 data.
Several cross checks have been performed to test the stability of analysis results against cross section systematic uncertainties or other effects due to a mismodeling of the missing energy. Indeed, calculated yields were proven to be quite stable and confirmed the robustness of our analysis strategy. Looking at phase space regions where either the signal contribution was expected to be suppressed, or background contributions had a shape similar to the signal one, no clear clues of mismodeling have been detected. In fact, in few cases the contribution of the signal was not negligible, and definitive conclusions about the agreement between observed and expected shapes *a priori* could not be drawn. Furthermore a cross check in 2012 has been setup to prove that, even setting lower jet p_T thresholds, no excess was visible in muon channel.
However, since no significant excess of data over SM expectation was found both in 2011 and 2012, for each analysis an upper limit on BR was set. An improvement on 2012 limit calculation was achieved, thanks to a more accu-

rate statistical approach and to a better control of systematic uncertainties. Indeed they have been proven to have a remarkable impact on our results: tighter limits on BR could be achieved just reducing the impact on the systematic effects, which are clearly dominant. In particular the overwhelming contribution of jet energy scales on the total uncertainty, which affected both muon and electron channels, jusitfied the choice to assume fully correlate (with very good approximation) the systematic uncertainties in muon and electron channels. As a direct consequence, any over-(under-) fluctuation in the limit computation for muon channel affected electron channel as well. Thus, an increase of available statistics was not expected to significantly improve limits calculation, while the combined limit did not provide an effective improvement in limit computation. In fact a reduction systematic uncertainties depends also on the overall performance of the whole CMS detector.

A Feldman-Cousins 95% CLs interval on decay branching fraction was computed both for muon and electron channel and for their combination as well. In the combined case a common value of BR was assumed for both channels, and the products of muon and electron likelihood was maximized. In 2012 analysis, limits have been calculated to be 0.0016 (0.0017) for muon (electron) channel, and 0.0015 for the combination.
Even though the experimental limit overlooks any theoretical assumption on BNV cross section, the effective model which our analysis relies on does not provide tight constraints on the expected branching fraction for any BNV process. Actually in [7] the authors even leave some room to make some phenomenological assumptions: since the expected branching ratio decreases as a function of Λ^{-4} (Section 1.2.3), a BNV branching fraction of 1.2×10^{-6} is expected choosing $\Lambda = 1$ TeV (being Λ the mass scale) and setting the fermion-flavor-dependent coefficients to 1 in Equation 4.1. On the contrary for lower mass scales (100 GeV) the expected branching ratio turned out to be even higher (about 1.2×10^{-4}). Even though different parameter values in principle could change the kinematical distributions of the top-quark decay products, the resulting impact on the final results was proven to be negligible. However since the accuracy of the effective model approximation improves as much as the mass scale increases, a description of BNV decay is expected to improve as well for Λ values much higher than experimentally accessible energies.
As already mentioned, the adopted analysis strategy is model-independent and one may consider to follow a similar approach also for future searches of other rare signals in $t\bar{t}$ final states. An interesting option could be (for instance) the search of flavor changing neutral currents in top decays like $t \to Zq$, where a small decay branching fraction is predicted and two leptons, with missing energy in the final states, are expected. However it is worth to note that Equation 4.6 is really effective just in the cases where the systematic uncertainties dominate on the statistical ones; hence cancellations of correlated terms and the overall reduction of systematic uncertainty may improve the upper limit

calculations. Therefore some further checks are required and a proper selections should be tuned on a case-by-case basis, in order to assess the actual benefit which this approach may provide.

Appendix A

Event selection

A.1 2011 selection

For each signal event exactly one muon with the following features has been required:

ID	Threshold
p_T	> 35 GeV/c
$\|\eta\|$	< 2.1
Transverse impact parameter	< 0.02 cm
Longitudinal impact parameter	< 1 cm
Global and particle-flow muon	-
Number of matched stations	> 2
χ^2 Global track fit	< 10

For each Signal event exactly one electron with the following features has been required:

ID	Thr. Barrel	Thr. Endcap
p_T	35 GeV/c	
$\|\eta\|$	< 2.5	
Particle Flow Electron	-	
Exclusion of η transition region	$1.442 < \|\eta\| < 1.5660$	
$I^{\updownarrow}_{Rel}$	< 0.15	
Transverse impact parameter	< 0.02 cm	
Longitudinal impact parameter	< 1 cm	
Electron ID	CiC HyperTight selection [114]	

Events with at least one further muon (electron) having $p_T > 10$ (20) GeV/c in $|\eta| < 2.4$ (2.5) and with $I^{\ell}_{\mathrm{Rel}} < 0.1$ are vetoed.

A.2 2012 selection

For each signal event exactly one muon with the following features has been required:

ID	Threshold
p_T	> 25 GeV/c
$\|\eta\|$	< 2.1
Transverse impact parameter	< 0.02
Longitudinal impact parameter	< 0.5
Global and particle-flow muon	-
Number of matched stations	> 2
χ^2 Global track fit	< 10

For each Signal event exactly one electron with the following features has been required:

ID	Thr. Barrel	Thr. Endcap
p_T	30 GeV/c	
$\|\eta\|$	< 2.5	
Particle Flow Electron	-	
Exclusion of η transition region	$1.442 < \|\eta\| < 1.5660$	
$I^{\updownarrow}_{Rel}$	< 0.15	
Transverse impact parameter	< 0.02 cm	
Longitudinal impact parameter	< 0.1 cm	
Number of missing reconstructed hits	< 1	
$\Delta\eta_{in}$	0.004	0.005
$\Delta\phi_{in}$	0.03	0.02
$\sigma_{i\eta i\eta}$	0.01	0.03
H/E	0.12	0.1
$\|1/E - 1/p\|$	0.05	0.05
Conversion Rejection Probability	1e-6	1e-6

Events with at least one further muon (electron) having $p_T > 10$ (20) GeV/c in $|\eta| < 2.4$ (2.5) and with $I^{\ell}_{\mathrm{Rel}} < 0.1$ are vetoed.

Appendix B

Statistical notes

B.1 2011 Statistical approach

In 2011 analysis the likelihood parameters θ_i are distributed according to a lognormal probability density function $\rho(\theta_i|\tilde{\theta}_i, \tilde{\sigma}_i)$, standing for the probability to measure a value θ for the i-th nuisance parameter, being its "true" value $\tilde{\theta}_i$ and its relative uncertainty $\tilde{\sigma}_i$.

The likelihood $\mathcal{L}$, given the data and the estimates $\tilde{\theta}_i, \tilde{\sigma}_i$ can be written as

$$\mathcal{L}(\text{data} \,|\, BR, \theta) = \mathcal{P}\big(N^T_{obs} \,|\, N^T_{exp}(BR, \theta)\big) \cdot \prod_i \rho(\theta_i|\tilde{\theta}_i, \tilde{\sigma}_i)\,, \tag{B.1}$$

$\mathcal{P}\big(N^T_{obs} \,|\, N^T_{exp}(BR, \theta)\big)$ stands for the Poissonian probability density function with mean $N^T_{exp}(BR, \theta)$, given by Equation 4.6 evaluated at the observed yield in tight selection N^T_{obs}.

Nuisance parameters in fact are affected by common sources of uncertainty: therefore a (positive) correlation among them is expected. Correlations are expected to exist for subsets of the likelihood parameters. Parameters belonging to different subsets are assumed to be independent of each other. Following a conservative approach one can assume that the uncertainty for each nuisance parameter consists of two distinct contributions: the first, shared by all parameters in its subset of correlated parameters, which account for the correlated part of the uncertainty; the second, describing the uncorrelated part, which provides independent and stochastic variations for each of them. Thus each nuisance parameter which belongs to the k-th correlated parameter subset can be expressed as a function of two gaussian-distributed variables, c_k and u_j with unit sigma, the former referring to the "correlated" part, the second driving the "uncorrelated" contribution. Thus for the k-th correlated subset, containing m_k nuisance parameters, the pdf of of the c_k and u_j variables reads:

$$\mathcal{G}(c_k) \cdot \prod_{j=1}^{m_k} \mathcal{G}(u_j)\,, \tag{B.2}$$

and the j-th nuisance parameters (e.g. the j-th efficiency value) can be written as:

$$\theta_j = \exp\left[\log(\tilde{\theta}_j) + x_j \times \log\left(1 + \sqrt{\sigma^2_{j,corr} + \sigma^2_{j,uncorr}}\right)\right] \tag{B.3}$$

where

$$x_j = \frac{c_k \times \sigma_{j,corr} + u_j \times \sigma_{j,uncorr}}{\sqrt{\sigma^2_{j,corr} + \sigma^2_{j,uncorr}}} \tag{B.4}$$

In Equation B.3, $\tilde{\theta}_j$ stands for the central value of j-th efficiency, $\sigma_{j,corr}$ and $\sigma_{j,uncorr}$ respectively as correlated and uncorrelated relative uncertainties, whose square sum gives the total uncertainty $\tilde{\sigma}$ defined before. One can easily show that x_j is a gaussian-distributed variable centered at zero with unit sigma, and consequently θ_i is distributed according to a lognormal distribution centered at $\tilde{\theta}$ and with uncertainty $\sqrt{(\sigma^2_{corr} + \sigma^2_{uncorr})}$.

B.2 Interpretation of 2012 statistical results

An interpretation has been given to justify the discrepancy between Hybrid and Frequentist predictions (Feldman Cousins approach can be treated as a particular case of Frequentist method). Even if such explanation is not completely rigorous, it may help the reader to gain a deeper understanding through the results.
It has already been noted the large impact of $\epsilon^B_{SM,SM}$ to $\epsilon^T_{SM,SM}$ ratio, whose uncertainties are dominant in the determination of upper limits. Remarkably, when the impact of systematic uncertainties on those efficiency (reported in Tables 5.1 and 5.2) is made symmetric and equal to the largest excursion for the related sources of systematic uncertainties (ISR and factorization scale with particular regard), the observed limit get closer to the expected, as it happens in hybrid approach. For sake of illustration, in Table B.1 the upper and expected limits have been recalculated, making symmetric the impact of systematic uncertainties on $\epsilon^B_{SM,SM}$ and $\epsilon^B_{SM,SM}$.

In principle one might check how the number of expected events and the observed events change, sampling several pseudo experiments in order to fulfill Hybrid-Bayesian and Frequentist prescription. Even though these quantities do not enter directly in the determination of limits, they may help to understand connections between expected and observed limits.
The main difference between Frequentist and Hybrid-Bayesian approach relies in the values that nuisance parameters assume when pseudo-experiments (toys) are sampled. Once all the toys are generated one can compare the distributions of the total pseudo-expected and pseudo-observed yield in the two methods, under the very same hypothesis [3].
In Hybrid Bayesian approach each nuisance pdf is interpreted as the "prior"

[3] Of course each method should implement the same test statistics and should use the

	95% CL Upp. lim.	Exp. lim.	68% exp. lim. range
Muon-Asymptotic	0.0017	0.0019	[0.0012, 0.0029]
Muon-Hybrid	0.0020	0.0021	[0.0016, 0.0030]
Muon-Frequentist	0.0018	0.0021	[0.0014, 0.0033]
Muon-FC	0.0017	0.0021	[0.0010, 0.0037]
Electron-Asymptotic	0.0019	0.0021	[0.0014, 0.0030]
Electron-Hybrid	0.0020	0.0022	[0.0016, 0.0032]
Electron-Frequentist	0.0019	0.0023	[0.0017, 0.0034]
Electron-FC	0.0018	0.0022	[0.0016, 0.0032]
Combined-Asymptotic	0.0015	0.0018	[0.0010, 0.0026]
Combined-Hybrid	0.0019	0.0020	[0.0015, 0.0029]
Combined-Frequentist	0.0018	0.0021	[0.0017, 0.0036]
Combined-FC	0.0015	0.0021	[0.0011 , 0.0041]

Table B.1: Observed 95% CL upper limit on BR, expected median 95% CL limit for the $BR = 0$ hypothesis and ranges that are expected to contain 68% of all observed excursions from the expected median for the muon and electron channels and for their combination. The uncertainties on the ratio of $\epsilon^{t}_{SM,SM}$ and $\epsilon^{B}_{SM,SM}$ have been made symmetric. Results obtained with the asymptotic, hybrid and Frequentist CLs variants are presented as well as those obtained with the Feldman-Cousins prescription. γ+jets contribution is not included

distribution of the true value of each parameter, whose mean match exactly the measured parameter, namely the best estimate of the true value. Whereas the expected number of events is fixed and remains the same as the "true" experiment, each pseudo-observation is calculated sampling nuisance parameter from its own prior "distribution". This process is equivalent to an MC integration over the nuisance parameters.
In Figure B.1 the distributions of pseudo-observed events for 15000 toys under $BR = 0$ and $BR = 0.002$ hypothesis are reported. The *expected number* of events for $BR = 0$ (BR = 0.002), is 2609 (2878). It's the worth to note that these values are smaller than the total number of events calculated using Equation 4.6 under the same hypothesis (2656 for $BR = 0$ and 2925 for $BR = 0.002$). Actually, when Tables 5.1 and 5.2 are made symmetric mean values matches almost exactly the ones computed by Equation 4.6.
If one look at the mean of *observed* counts obtained in toys (2609) and in the real experiment (about 2614), one finds that in both cases they are lower than the respective expected values; moreover they are very close each other (as expected values are).
Intuitively one may conclude that the observed and expected limits should be

same $CLs/CLs+b$ construction. In this paragraph however just the expected and observed number of events will be taken into account, since just a semi-quantitative picture of the problem will be given.

very close in Hybrid approach. Indeed this is what we observe.

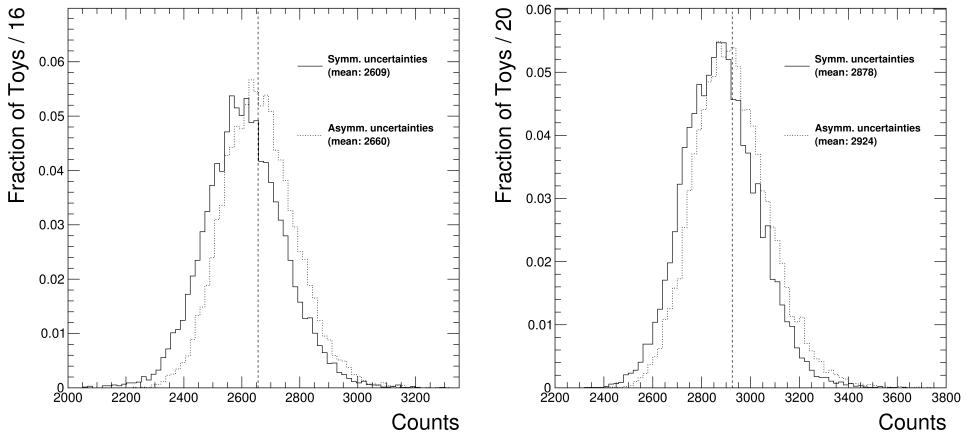


Figure B.1: Pseudo-observed number of events for $BR = 0$ (left) and for $BR = 0.002$ obtained for Hybrid-Bayesian prescription, computed with actual numbers in Table 5.2 and making impact on $\epsilon^{B}_{SM,SM}$, $\epsilon^{T}_{SM,SM}$ of systematic uncertainties sources symmetric ; the expected number of events is computed setting the correct value of BR into Equation 4.6

In Frequentist approach the nuisance parameters pdfs are interpreted as distribution of measured values, given a "true" value. Such a true value (which in principle should be computed using the Neyman band prescription), is estimated fitting parameters once for all. The observed number of events is sampled from a Poissonian, which is centered around the best fit value of the expected number of events, being all the nuisance parameters frozen at their fitted values. The expected number of counts can be obtained using nuisance parameters values drawn from the pdf centered around nuisance best fit values.

In this case one finds the mean of pseudo-expected counts distribution significantly lower than the pseudo-observed, under the same hypothesis (case for $BR = 0$ has been reported in Figure B.2). On the contrary in the real experiment observed counts (2614) are lower than expected ones (about 2660). Remarkably, this significant discrepancies is reduced when numbers in Tables 5.2 are made symmetric and expected (observed) limits get much closer. Thus one may expect an observed limit lower than the expected limit, even if a quantitative estimate is more difficult to be computed. This is what we observe: even in this case when the impact of systematic uncertainties get symmetric, expected and observed limits are pushed towards each others.
Finally, looking at the low value of expected counts in Frequentist approach, one may expect that an higher BR should be used to fit observed counts of events in Frequentist approach. Therefore one may conclude that the expected limit in Frequentist approach will be significantly higher than the expected in Hybrid. Actually this is exactly what we observe.

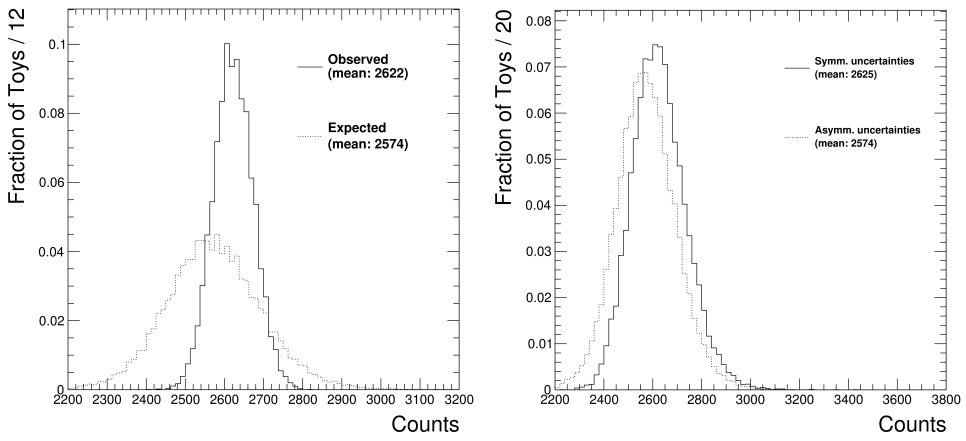


Figure B.2: Left: Distribution of expected (solid line) and observed counts for $BR = 0$ hypothesis under Frequentist approach, for actual values of Table 5.2. Pseudo-observed events are significantly higher than expected in toys; on the contrary in the real experiment observed (2614) is lower than expected (about 2660). This is exactly what we observe. Right: expected yield for Frequentist approach, with actual and "symmetrized" values of table 5.2; in the latter case, expected and observed limits get much closer to the best fit value for $BR = 0$.

In conclusion, the discrepancy between Hybrid and Frequentist approach is a joint effect of the difference i nuisance parameter treatment between the two methods and of the asymmetry in the impact of systematic uncertainty sources. Whereas in Hybrid approach observed and expected limits are pushed towards each other, in Frequentist approach they tend to ride off. Of course a quantitative approach should not overlook a full CLs construction, based on a full test statistic distribution.

Bibliography

[1] S. L. Glashow,Partial Symmetries Of Weak Interactions,Nucl. Phys. **22**, 579 (1961).

[2] A. Salam, *Weak And Electromagnetic Interactions*, Originally printed in *Svartholm: Elementary Particle Theory, Proceedings Of The Nobel Symposium Held 1968 At Lerum,Sweden**, Stockholm 1968, 367-377.

[3] S. Weinberg,*A Model Of Leptons*, Phys. Rev. Lett. **19** (1967) 1264

[4] D. Griffiths, *Introduction to Elementary Particles*,Second Edition, Wiley-VCH Verlag GmbH & Co.

[5] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett. **13**, (1964) 508 ;
F. Englert and R. Brout, *Broken symmetries and the mass of gauge vector mesons*, Phys.Rev. Lett. **13** (1964) 321.

[6] Aubert, J. J., et al. *Experimental observation of a heavy particle J.* Physical Review Letters 33.23 (1974): 1404-1406.
Augustin, J-E., et al. *Discovery of a Narrow Resonance in e^+ e^- Annihilation.* Physical Review Letters 33.23 (1974): 1406-1408.

[7] Dong, Zhe, et al. *Baryon number violation at the LHC: the top option* Physical Review D 85.1 (2012): 016006.

[8] Peskin, Michael E., Daniel V. Schroeder, and Emil Martinec. *An introduction to quantum field theory.* Physics Today 49 (1996): 69.

[9] J. Beringer et al. (Particle Data Group), Phys. Rev. D86, 010001 (2012)

[10] Aad, Georges, et al. *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC.* Physics Letters B (2012).

[11] Chatrchyan, Serguei, et al. *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC.* Physics Letters B (2012).

[12] Chakraborty, Dhiman, Jacobo Konigsberg, and David Rainwater. *Review of top quark physics.* arXiv preprint hep-ph/0303092 (2003)

[13] Liss, Tony M. *Top quark properties.* arXiv preprint arXiv:1212.0489 (2012).

[14] Czakon, Michal, et al. *Constraints on the gluon PDF from top quark pair production at hadron colliders.* arXiv preprint arXiv:1303.7215 (2013).

[15] Collaboration, A. T. L. A. S. *Search for FCNC single top-quark production at sqrt (s)= 7 TeV with the ATLAS detector.* arXiv preprint arXiv:1203.0529 (2012).

[16] Chatrchyan, Serguei, et al. *Search for flavor changing neutral currents in top quark decays in pp collisions at 7 TeV.* Physics Letters B (2012).

[17] Sullivan, Zack. *Supersymmetric QCD correction to top-quark production at the Fermilab Tevatron.* Physical Review, D 56.1 (1997).

[18] Appelquist, Thomas, Hsin-Chia Cheng, and Bogdan A. Dobrescu. *Bounds on universal extra dimensions.* Physical Review D 64.3 (2001): 035002.
Appelquist, Thomas, and George Triantaphyllou. *Technicolor enhancement of $t\bar{t}$ production at TeV colliders.* Physical review letters 69.19 (1992): 2750.

[19] Cho, Peter, and Elizabeth H. Simmons. *Searching for G^3 in $t\bar{t}$ production.* Physical Review D 51.5 (1995): 2360.

[20] Atwood, David, Alex Kagan, and T. G. Rizzo. *Constraining anomalous top quark couplings at the Fermilab Tevatron.* Physical Review D 52.11 (1995): 6264.

[21] Lane, Kenneth, and Estia Eichten. *Natural topcolor-assisted technicolor.* Physics Letters B 352.3 (1995): 382-387.
Eichten, Estia, and Kenneth Lane. *Multiscale technicolor and top production.* Physics Letters B 327.1 (1994): 129-135.

[22] Hill, Christopher T. *Topcolor assisted technicolor.* Physics Letters B 345.4 (1995): 483-489.
Balaji, Bhashyam. *Top decay in topcolor-assisted technicolor.* Physics Letters B 393.1 (1997): 89-93.

[23] Haber, Howard E., Gordon Kane, and Sally Dawson. *The Higgs hunter's guide.* Vol. 2. New York: Addison-Wesley, 1990.

[24] Weyl, Hermann. *Elektron und Gravitation. i.* Zeitschrift für Physik A Hadrons and Nuclei 56.5 (1929): 330-352.

[25] Cifuentes A., *Baryon and Lepton number violation in gauge unified theories*. Magister Thesis (1990)

[26] Reines, F., C. L. Cowan Jr, and M. Goldhaber. *Conservation of the Number of Nucleons*. Physical Review **96**.4 (1954) 1157.

[27] Sakharov, Andrei Dmitrievich. *Violation of CP in variance, C asymmetry, and baryon asymmetry of the universe.* Physics-Uspekhi 34.5 (1991): 392-393.

[28] Primakoff, Henry, and S. Peter Rosen. *Baryon number and lepton number conservation laws.* Seventy Years Of Double Beta Decay: From Nuclear Physics to Beyond-Standard-Model Particle Physics (2010): 279.

[29] 't Hooft, Gerard. *Symmetry breaking through Bell-Jackiw anomalies.* Physical Review Letters **37**.1 (1976): 8-11.

[30] Weinberg, Steven. *Supersymmetry at ordinary energies. Masses and conservation laws.* Physical Review D **26**.1 (1982): 287.

[31] Georgi, H., Helen R. Quinn, and Steven Weinberg. *Hierarchy of interactions in unified gauge theories.* Physical Review Letters **33**.7 (1974): 451-454.

[32] Bekenstein, Jacob D. "Nonexistence of baryon number for static black holes." Physical Review D 5.6 (1972): 1239.

[33] Barbier, Rémi, et al. *R-Parity-violating supersymmetry.* Physics reports **420**.1 (2005): 1-195.

[34] CMS Collaboration, *Search for anomalous production of multilepton events and R-parity-violating supersymmetry in* $\sqrt{s}$*= 7 TeV pp collisions.* CMS PAS EXO-11-045, http://cdsweb. cern. ch/record/1393758 (2011).

[35] Degrande, Celine, et al. *Effective field theory: a modern approach to anomalous couplings.* arXiv preprint arXiv:1205.4231 (2012).

[36] G. Durieux *Baryon number violation at the LHC: The top option*, CP3 lunch seminar presentation (2012)s

[37] Durieux, G., et al. *Three-generation baryon and lepton number violation at the LHC.* arXiv:1210.6598 (2013)

[38] Durieux, G. *Flavourful baryon and lepton number violation at the LHC.* arXiv:1305.3488 (2013)

[39] Learned, J.G. *Proton decay, review and future prospects*, Proceeding for 9th International Symposium on Neutrino Telescopes 6-9 Mar 2001. Venice, Italy (2001) 349-362

[40] Murayama, Hitoshi. *Nucleon Decay in GUT and Non-GUT SUSY models.* Arxiv preprint hep-ph/9610419 (1996).

[41] Regis, C., et al. *Search for proton decay via $p \to \mu^+ K^0$ in Super-Kamiokande I, II, and III.* Physical Review D 86.1 (2012): 012006.

[42] Kobayashi, K., et al. *Search for nucleon decay via modes favored by supersymmetric grand unification models in Super-Kamiokande-I.* Physical Review D 72.5 (2005): 052007.

[43] Nishino, H., et al. *Search for nucleon decay into charged antilepton plus meson in Super-Kamiokande I and II.* Physical Review D 85.11 (2012): 112001.

[44] Sanchez del Amo, P., et al. *Searches for the baryon-and lepton-number violating decays $B^0 \to \Lambda_c^+ l^-$, $B^- \to \Lambda l^-$, and $B^- \to \bar{\Lambda} l^-$* Physical Review. D, Particles Fields 83.9 (2011).

[45] Marciano, William J. *Tau physics 1994: A theoretical perspective.* No. BNL–61141; CONF-9409277–2. Brookhaven National Lab., Upton, NY (United States), 1994.

[46] Miyazaki, Y., et al. *Search for lepton and baryon number violating τ^- decays into $\bar{\Lambda}\pi^-$ and $\Lambda\pi^-$* Physics Letters B 632.1 (2006): 51-57.

[47] Godang, R., et al. *Search for baryon and lepton number violating decays of the τ lepton.* Physical Review D 59.9 (1999): 091303.

[48] Aaij, R., et al. *Searches for violation of lepton flavour and baryon number in tau lepton decays at LHCb.* arXiv preprint arXiv:1304.4518 (2013).

[49] CMS Collaboration, *Search for RPV supersymmetry with three or more leptons and b-tags.* CMS-PAS-SUS-12-027.

[50] Evans, Lyndon, and Philip Bryant. *LHC machine.* Journal of Instrumentation 3.08 (2008): S08001.

[51] Schopper, Herwig, and CH Llewellyn Smith. *LEP: A Historical Introduction [and Discussion].* Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences 336.1642 (1991): 179-189.

[52] Weiglein, Georg, et al. *Physics interplay of the LHC and the ILC.* Physics Reports 426.2 (2006): 47-358.

[53] Image courtesy of CERN

[54] Buon, Jean. *Beam phase space and emittance.* CERN European Organisation for Nuclear Research -REPORTS-CERN (1994): 89-89.

[55] ATLAS Collaboration, G. Aad et al., *The ATLAS Experiment at the CERN Large Hadron Collider.* Journal of Instrumentation 3.03 (2008) S08003.

[56] CMS Collaboration, and R. Adolphi. *The CMS experiment at the CERN LHC.* Journal of Instrumentation 3.04 (2008): S08004

[57] Ruggiero, G., et al. *The TOTEM Detector at LHC.* Nucl. Instrum. Methods Phys. Res., A 617: 62-66.

[58] LHCf Collaboration, O. Adriani et al., *The LHCf detector at the CERN Large Hadron Collider.* Journal of Instrumentation 3.06 (2008): S08006.

[59] LHCb Collaboration, J. Alves, A. Augusto et al., *The LHCb Detector at the LHC.* Journal of Instrumentation 3.05 (2008): S08005.

[60] ALICE Collaboration, K. Aamodt et al., *The ALICE experiment at the CERN LHC.* Journal of Instrumentation 3.02 (2008): S08002.

[61] CMS Collaboration, *CMS Luminosity - Public Results.* Twiki page, https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults

[62] CERN Press Office, *CERN releases analysis of LHC incident.* http://press.web.cern.ch/press-releases/2008/10/cern-releases-analysis-lhc-incident .

[63] CERN Courier, *Work for the LHC's first long shutdown gets under way.* http://cerncourier.com/cws/article/cern/52361 .

[64] Karimäki, V. *The CMS tracker system project: Technical Design Report.* Technical Design Report CMS. CERN, Geneva (1997).

[65] Khachatryan, Vardan, et al. *CMS tracking performance results from early LHC operation.* The European Physical Journal C 70.4 (2010): 1165-1192.

[66] Fabjan, Christian W., and Fabiola Gianotti. *Calorimetry for particle physics.* Reviews of Modern Physics 75.4 (2003): 1243.

[67] Lustermann, Werner. *The electromagnetic calorimeter of cms, summary and status.* Journal of Physics: Conference Series. Vol. 160. No. 1. IOP Publishing, 2009.

[68] CMS Collaboration, *Search for baryon number violating top quark decays in pp collisions at* $\sqrt{s} = 7$ *TeV* CMS-PAS-B2G-12-002.

[69] CMS Collaboration, *Search for baryon number violating top quark decays in pp collisions at* $\sqrt{s} = 8$ *TeV* CMS-PAS-B2G-12-023.

[70] Alwall, Johan, Simon de Visscher, and Fabio Maltoni. *QCD radiation in the production of heavy colored particles at the LHC.* Journal of High Energy Physics 2009.02 (2009): 017.

[71] Frixione, Stefano, Paolo Nason, and Carlo Oleari. "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method." Journal of High Energy Physics 2007.11 (2007): 070.

[72] Sjöstrand, Torbjörn, Stephen Mrenna, and Peter Skands. *PYTHIA 6.4 physics and manual.* Journal of High Energy Physics 2006.05 (2006): 026.

[73] Alwall, Johan, et al. "MadGraph 5: going beyond." Journal of High Energy Physics 2011.6 (2011): 1-40.

[74] Sjöstrand, Torbjörn, Stephen Mrenna, and Peter Skands. *A brief introduction to PYTHIA 8.1.* Computer Physics Communications 178.11 (2008): 852-867.

[75] M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau, A. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions* JHEP 0307:001,2003, hep-ph/0206293.

[76] Agostinelli, S., et al. "GEANT4—a simulation toolkit." Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment 506.3 (2003): 250-303.

[77] CMS Collaboration, *Standard Model Cross Sections for CMS at 8 TeV.*, Private CMS Twiki page

[78] P.Azzurri, B.Mangano, *Optimal filtering of fake tracks*, CMS IN 2008/017 (2011)
CMS Collaboration, *Tracking and Vertexing Results from First Collisions*, CMS-PAS-TRK-10-001
CMS Collaboration, *CMSSW Software Cross-Reference*, http://cmslxr.fnal.gov/lxr/source/RecoTracker/FinalTrackSelectors/python/selectHighPurity_cfi.py

[79] Chatrchyan, Serguei, et al. *Search for neutral Higgs bosons decaying to tau pairs in pp collisions at* $\sqrt{s} = 7$ *TeV.*, Physics Letters B 713.2 (2012): 68-90.

[80] Cacciari, Matteo, Gavin P. Salam, and Gregory Soyez. "The anti-kt jet clustering algorithm." Journal of High Energy Physics 2008.04 (2008): 063.

[81] Rene Brun and Fons Rademakers, *ROOT - An Object Oriented Data Analysis Framework*, Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. & Meth. in Phys. Res. A 389 (1997) 81-86. See

also http://root.cern.ch/.
ROOT Reference, *RooLognormal*, http://root.cern.ch/root/html534/RooLognormal.html

[82] Moneta, Lorenzo, et al. *The RooStats Project.* arXiv preprint arXiv:1009.1003 (2010).

[83] Schott, Grégory. *RooStats for Searches.* arXiv preprint arXiv:1203.1547 (2012).

[84] Conway, J. S. *Incorporating nuisance parameters in likelihoods for multisource spectra.* arXiv preprint arXiv:1103.0354 (2011).

[85] CMS Collaboration, *Determination of jet energy calibration and transverse momentum resolution in CMS.* arXiv preprint arXiv:1107.4277 (2011).

[86] CMS Collaboration, *Performance of b-jet identification in CMS.* CMS Physics Analysis Summary CMS-PAS-BTV-11-001 (2011).

[87] CMS Collaboration, "Missing transverse energy performance of the CMS detector." arXiv preprint arXiv:1106.5048 (2011).

[88] Nadolsky, Pavel M., et al. "Implications of CTEQ global analysis for collider observables." Physical Review D 78.1 (2008): 013004.
P. Biallass et al., "Parton Distribution Uncertainty Determination within CMSSW", CMS Note 2009/048 (2009).

[89] CMS Collaboration, *Jet energy scale uncertainty sources*, Private CMS Twiki page

[90] CMS Collaboration, *CMS luminosity based on pixel cluster counting-Summer 2012 Update.* CMS Physics Analysis Summary CMS-PAS-LUM-12-001 10 (2012).

[91] Read, *A. L. Modified frequentist analysis of search results (the CL_s method).* No. CERN-OPEN-2000-205. 2000.

[92] Feldman, Gary J., and Robert D. Cousins. *Unified approach to the classical statistical analysis of small signals.* Physical Review D 57.7 (1998): 3873.

[93] Moneta, L. *LHCb RooStats tutorial*, CERN, 27 April 2012

[94] Cousins, Robert D., and Virgil L. Highland. *Incorporating systematic uncertainties into an upper limit.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 320.1 (1992): 331-335.

[95] ATLAS, CMS. Collaborations, *LHC Higgs Combination Group, Procedure for the LHC higgs boson search combination in Summer 2011.* ATL-PHYS-PUB/CMS NOTE 2011-11, 2011/005, http://cdsweb. cern. ch/record/1379837, 2011.

[96] Cowan, Glen, et al. *Asymptotic formulae for likelihood-based tests of new physics.* The European Physical Journal C 71.2 (2011): 1-19.

[97] CMS collaboration. "Combination of top pair production cross section measurements." CMS-PAS-TOP-11-024 (2011).

[98] Chatrchyan, Serguei, et al. *Jet production rates in association with W and Z bosons in pp collisions at* $\sqrt{s} = 7$ *TeV.* Journal of High Energy Physics 2012.1 (2012): 1-46.

[99] CMS Collaboration, *Determination of jet energy calibration and transverse momentum resolution in CMS.* arXiv preprint arXiv:1107.4277 (2011).

[100] Santocchia, Attilio. *Jet energy correction in CMS.* Nuclear Science Symposium Conference Record (NSS/MIC), 2009 IEEE. IEEE, 2009.

[101] Frühwirth, R. *Application of Kalman filtering to track and vertex fitting.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 262.2 (1987): 444-450..

[102] M. Anderson et al., *Review of clustering algorithms and energy corrections in ECAL.*, CMS IN 2010/008 (2011)

[103] CMS Collaboration, *Muon Reconstruction in the CMS Detector* , CMS AN 2008/097 (2009)

[104] CMS Collaboration, *Performance of CMS muon reconstruction in pp collision events at sqrt (s)= 7 TeV.* arXiv preprint arXiv:1206.4071 (2012).

[105] Campbell, John M., J. W. Huston, and W. J. Stirling. *Hard interactions of quarks and gluons: a primer for LHC physics.* Reports on Progress in Physics 70.1 (2007): 89.

[106] CMS Collaboration, *Particle–Flow Event Reconstruction in CMS and Performance for Jets, Taus, and* E_T^{miss} , CMS PAS PFT-09-001 (2009)

[107] CMS Collaboration, *Electron reconstruction and identification at* $\sqrt{s} =$ *7 TeV.*, CMS PAS EGM-10-004 (2010)

[108] G. Landsberg, F. Moortgat, *MET Reconstruction, Performance and Validation.* CMS AN 2008/089

[109] S.Esen, G. Landsberg *Missing ET Performance in CMS* CMS AN 2007/041

[110] D. Barge et al., *Study of photon conversion rejection at CMS.*, CMS AN 2009/159

[111] G. L. Bayatian et al., *CMS technical design report, volume II: Physics performance.* Journal of Physics, G 34 (2007) : 995–1579

[112] P. Lenzi, *Study of Z/γ + jets production with the CMS detector using matched Matrix Element-Parton Shower event generators*, PhD thesis, Scuola di Dottorato in Scienze, Università degli Studi di Firenze.

[113] S. De Visscher, J. Alwall, *Introduction to jet-parton matching in MG/ME.*, https://cp3.irmp.ucl.ac.be/projects/madgraph/wiki/IntroMatching

[114] CMS Collaboration, *Electron reconstruction and identification at $\sqrt{s}$ = 7 TeV.* CMS Physics Analysis Summary CMS-PAS-EGM-10-004 12 (2010).
Baffioni, Stephanie, et al. *Electron reconstruction in CMS.* The European Physical Journal C 49.4 (2007): 1099-1116.

Acknowledgements

When I started my PhD, I wished to undertake an experience of human and professional growth.
Indeed I could not ask anything more than what I had. I will remember especially the faces of the friends, colleagues and collaborators who came along with me and provided an outstanding contribution to make me score this goal.
First, I would like to thank my "italian-side" supervisor, Paolo Vitulo, who supported me and gave me the opportunity to start this extraordinary experience. During these years I learnt a lot; I shared with Paolo happy and unhappy times, as well as the satisfactions that this work gave us. I address him my keenest human and professional consideration.
Thanks to Giacomo Bruno, my "belgian-side" supervisor, who accepted to bet on me. Thanks to have grown me up in this unique environment, so different from the ones I was used to; thanks to have provided me anything I needed during these years, and to have shown me what an amazing thing the human mind is.
Thanks to Davide Pagano, who gave an irreplaceable contribution to this work. I have grown up with Davide, from the human and professional point of view. Thanks to have trusted me! We shared working days and nights, anxiety and enthusiasm, success and frustration. Say, eye bags. I forgot: dude, thanks to have introduced me *nel Mondo delle Scadenze Irrevocabili*!
Thanks to Fabio Maltoni, Jean-Marc Gerard and Gauthier Durieux, for the fruitful discussions we had and for all the clarifications they kindly gave us; thanks, Anna Colaleo, Marco Fraternali and Daniele Pedrini, to have red my thesis and for the interesting comments you sent me.
Finally, I would like to remind the members of the RPC background group, even thoguh they were not directly involved in this work of thesis: thanks to Silvia Costantini, for her patience and for her dedication to the Physicist's job, and to Mircho Rodozov, for his kindness and congeniality: I wish you all the best for your future. I'm really happy to have shared my time with you.
Now, I would like also to extend my acknowledgments to that universe of wonderful people who came along with me inside and outside the academia.
I will start from CP3 guys: Jerome, Vincent and Pavel, thanks for taking care

of the lenses of our telescopes; a huuuge *thanks!* to Céline, Johan, Olivier, Pierre, Priscila, Simon, and all the members of the (outstanding) MadGraph Team; thanks Lawrence, Djiogo, Gustavo, Jesus, Marcello, Roberto, Suzan and Tristan to have shared the European Football Championship, the Woke at Louvain-la-Neuve, the nightly beers and the interesting discussions on the train to Schuman; and thanks to Simone, too, who also tried to introduce me to Lambic Beers (with no success, anyway). I will never forget all the PhD student I knew during these years: Adrien (famous for his legendary teachings of contemporary Napoletano); Elisa (please, tell me the next time you'll cook the baked heart - which animal belonged to...?); Lucia (experimentalist not just in Physics, but in countless artistic and culinary disciplines); Marco (who has, among his several merits, the honour to have introduced me for the first time to pasta, bacon and chickpeas); Mathieu (great office mate; my deepest respect for your bravery, climbing the hill under the snow at -4°C, on your bike). And many thanks also to all the other CP3 guys, among them: Alexandre, Arnaud, Bob, Camille, Claude, Elvira, Federico, Gauthier, Laurent, Loic, Ludvine, Michael, Nicholas, Pavel. Finally thanks to the Administration staff, in particular to the irreplaceable Ginette, for his kindness and congeniality, who helped me from the beginning.
Thanks also to all the landlords and the house mates I found wandering from house to house and that I met during these odd years! They crossed my life and gave me extraordinary windows on other cultures, and other habits!
Poi.
Poi c'è Pavia.
La Pavia di Giulio, con la sua straordinaria creatività, umanità e gioia di vivere che si portava dentro, il cui pensiero mi ha costantemente accompagnato e mi accompagnerà per sempre.
La Pavia di Beppe e del pranzo che ha voluto organizzare per salutarci, prima di partire.
La Pavia di Roberto, con tutti i consigli che in questi anni mi ha dato, e a cui io auguro, un giorno, di poter assaggiare la famosa *Pecora in Callara*, perché sono sicuro che poi mi ringrazierà a sua volta.
La Pavia delle gruppo di CMS: Cristina, che ringrazio (nonostante accetti solo mate in filtro) per gli innumerevoli sforzi in cui si sempre prodigata per capirmi, Aurora (delegata ambasciatrice presso l'Amministrazione), Alice (anche se ormai compromessa dal Lato Oscuro della Forza), Ilaria (che non si muoverà dall'ufficio prima che la sua Tesi di Laurea sarà corretta); la Pavia di Luca (che spero non dimentichi più cartoni imbevuti di vernice secca in Lab, com'è ormai solito fare) e di Francesco (che potrebbe ritenere il mio commento meritevole di ulteriori approfondimenti).
Grazie anche a Paolo e Paola, per l'attenzione che hanno avuto nei miei confronti, e insieme ai quali lavorare è stato veramente un piacere.
Vorrei rivolgere un sentito ringraziamento a tutti gli amici, di Brescia e di tutta Italia, che mi hanno sostenuto e che non mi hanno mai fatto mancare il loro

supporto, nonostante gli impegni ci abbiano tenuti distanti. Grazie ai miei genitori, a mia sorella e a tutti i miei famigliari, che hanno sempre ascoltato (e sopportato) i miei (lunghissimi) racconti facendo di tutto per farmi sentire la loro presenza e sostenermi nelle difficoltà.
Grazie infine a Margherita, per la tenacia, la dolcezza, la fiducia e la pazienza con cui ha affrontato questi lunghi mesi di lontananza e ha saputo donarmi il suo amore e la sua energia ogni giorno.