

Università degli Studi di Torino
FACOLTÀ DI SCIENZE MATEMATICHE FISICHE E
NATURALI
Corso di Laurea Magistrale in Fisica

Tesi di Laurea Magistrale in Fisica



Study of $pp \rightarrow Z + b$ process with the CMS detector as a background in Higgs boson search

Candidato:
Stefano Casasso

Relatore:
Prof. Ernesto Migliore
Controrelatore:
Dott.ssa Stefania Beolè

Anno accademico 2010-2011

Contents

Introduction	vi
1 The Standard Model and the Higgs Boson	1
1.1 The Standard Model	1
1.1.1 Goldstone theorem and spontaneous symmetry breaking	3
1.1.2 The Higgs mechanism and the Glashow-Weinberg-Salam model of electroweak interaction	5
1.1.3 Yukawa coupling	8
1.2 Higgs properties	9
1.2.1 Higgs mass	10
1.2.2 Higgs production	11
1.2.3 Higgs decay channels	12
1.3 Experimental results in the SM Higgs boson search	15
1.3.1 Indirect Higgs search	15
1.3.2 Direct Higgs search at LEP	16
1.3.3 Direct Higgs search at Tevatron	17
1.3.4 Direct Higgs search at LHC	19
2 The Large Hadron Collider and the CMS Detector	21
2.1 The Large Hadron Collider	21
2.1.1 Design and performances	21
2.1.2 The LHC experiments	23
2.2 The CMS detector	27
2.2.1 Coordinate conventions	27
2.2.2 Overall design	28
2.2.3 Magnet	29
2.2.4 Muon system	30
2.2.5 Electromagnetic calorimeter	32
2.2.6 Hadron Calorimeter	33
2.2.7 Inner tracking system	35
2.2.8 Trigger and data acquisition	36
2.3 CMS performances from early data	38
2.3.1 Muon identification and reconstruction	38

2.3.2	Electron identification and reconstruction	40
2.3.3	Jet performances	44
2.3.4	B-tagging performances	46
3	Production of Z in association with one or more bottom-quark jets	51
3.1	Theoretical motivations for Z+b measurement	51
3.1.1	SM Higgs search	51
3.1.2	MSSM Higgs search	52
3.1.3	M2HDM Higgs search	55
3.2	Z+b cross section	58
3.2.1	Z + b in the fixed flavor number scheme	59
3.2.2	Z + b in the variable flavor number scheme	60
4	Z+b events selection with CMS detector	63
4.1	Monte Carlo and data samples	63
4.2	Reconstruction and selection of Z+b candidates	65
4.2.1	Di-lepton selection	65
4.2.2	Jet selection	66
4.2.3	B-tagging	67
4.3	Results	67
4.4	Event display	71
5	Z+b events with more than two leptons	75
5.1	Selection of additional leptons	75
5.1.1	Muon selection	76
5.1.2	Electron selection	77
5.2	Event Yields	79
5.3	Control plots	91
5.4	Additional leptons kinematics	95
5.4.1	Muons	96
5.4.2	Electrons	98
5.5	Data/MC comparison	100
5.6	Event display	107
	Conclusions	111
	Appendices	113
	A Factorization theorem and hadronic cross sections	113
	B B-tagging in CMS	115
	C The anti- k_T jet clustering algorithm	118
	D Event reweighting	121
	Bibliography	125

Acknowledgements

127

Introduction

The Standard Model is the most successful theory of subatomic elementary particles. It provides an elegant mathematical framework which describes how the fundamental constituents of the matter interact between each other, through the electromagnetic, weak and strong forces.

Some decades of experimental observations have validated this theory and successfully tested its predictions. Nevertheless, one particle predicted by the Standard Model has never been observed yet: the Higgs boson. The existence of the Higgs field is invoked within the framework of the spontaneous breaking of electroweak symmetry to explain the mass of elementary particles. The same mechanism is used in other theories, like Supersymmetry, which aim to solve some of the problems which Standard Model suffers. Such theories often predicts the existence of many Higgs bosons, thus increasing the number of channels where scientists should look at this (these) elusive particle(s). The "Higgs hunting" is one of the main objectives of the physics program at the Large Hadron Collider (LHC): the Compact Muon Solenoid (CMS), one of the two largest experiments, has been designed mainly for this reason, as well as for looking at signs of new physics. The Higgs decay $H \rightarrow ZZ^{(*)} \rightarrow 4l$ is often referred to as the "golden channel", because of its clean signature and because of the possibility to study the Higgs properties with high precision over a wide mass range.

The work presented in this thesis has been carried out within the Torino CMS group. It concerns the study of the production of the Z boson in association with one or more b quarks as a background in Higgs search in the golden channel of the Standard Model Higgs, through events selected with CMS detector.

In chapter 1, after a short introduction of the Standard Model, the theoretical framework of the scalar sector is described, the phenomenological properties of the Higgs boson are presented and the most recent results in the Higgs search from LEP, Tevatron and LHC are discussed. In chapter 2, after an introduction to the LHC and its experiments is given, the design of the CMS detector is presented with more details together with some performances estimated from the first collision data. A brief presentation of the theoretical aspects related to the Z+b measurement is given in chapter 3, where the main motivations for this measurement and the most recent

calculations of the $Z+b$ cross section are presented. In chapter 4 the selection of $Z+b$ candidates is described and the first preliminary results on $Z+b$ candidate events selected in the 2010 data are given. In chapter 5 the selection of additional leptons in the $Z+b$ candidate events is discussed and the analysis of these events is presented, based on a total integrated luminosity of 1.14 fb^{-1} .

Chapter 1

The Standard Model and the Higgs Boson

The Standard Model (SM) is the theory that most successfully describes subnuclear particles interactions. Over the years it passed many experimental precision tests, but still one particle, the Higgs boson, predicted in order to explain the mass of all the other observed particles has not yet been detected. The discovery or the rule out of the Higgs boson is one of the main physics programs for the Large Hadron Collider (LHC) experiments. In this chapter, after a summarized review of SM is given, the theoretical motivation and the mathematical construction of the Higgs sector and the mechanism of spontaneous symmetry breaking are overviewed and the latest results in the Higgs search at LEP, Tevatron and LHC are reviewed, with focus on the $H \rightarrow ZZ^{(*)} \rightarrow 4l$ channel.

1.1 The Standard Model

The Standard Model (SM) is the local gauge theory based on the symmetry group:

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$$

It is the direct product of the $SU(3)_C$, the color group upon which Quantum Chromo Dynamics (QCD) is built, the gauge groups of weak isospin, $SU(2)_L$, and hypercharge, $U(1)_Y$. Electromagnetic and weak interactions are unified in the *electroweak* gauge group $SU(2)_L \otimes U(1)_Y$, upon which the Glashow-Weinberg-Salam Model is built.

The particle content of the SM can be summarized as in the following.

- **12 fermions**, i.e. spin $\frac{1}{2}$ particles
According to spin-statistics theorem, fermions obeys Pauli exclusion principle. From Dirac's equation, it is well known that every spin $\frac{1}{2}$ particle has a corresponding antiparticle (electron-positron, up-quark

(u) and antiup-quark (\bar{u}) and so on) with opposite charges and same mass and mean life.

Fermions can be classified by their charge, i.e. by their interactions, in *quarks* (Table 1.1) which experience all the interactions (strong and electroweak) and *leptons* (Table 1.2) which have no color charge and do not interact strongly. Neutrinos do not have electric charge and they only interact weakly.

Table 1.1: Quarks

Charge (unit of e)	1 st gen.		2 nd gen.		3 rd gen.		Interact.
$+\frac{2}{3}$	Up	u	Charm	c	Top	t	all
$-\frac{1}{3}$	Down	d	Strange	s	Bottom	b	all

Table 1.2: Leptons

Charge (unit of e)	1 st gen.		2 nd gen.		3 rd gen.		Interact.
-1	Electron	e^-	Muon	μ^-	Tau	τ^-	E.M. + weak
0	Electron neutrino	ν_e	Muon neutrino	ν_μ	Tau neutrino	ν_τ	weak

The fermions are also divided in generations: corresponding fermions of different generations differ in mass and *flavor*. The lightest fermions are the one in the first generation, and they do not decay so the ordinary matter is made of electrons and protons and neutrons, which are made of different combinations of u and d quarks.

- **12 gauge bosons**, i.e spin 1 particles

They are the "force carriers" which mediate the interactions between elementary particles. These interactions, in quantum field theory, are seen as the exchange of vector bosons (i.e. spin 1) between the interacting particles. Each fundamental force has its vector bosons, and their properties are responsible for the phenomenology of the corresponding force. In table 1.3 the main features of the 3 fundamental interactions (gravity excluded) are listed. Among vector bosons, only W^\pm and Z are massive:

$$M_W = 80.399 \pm 0.023 \text{ GeV} \qquad M_Z = 90.1876 \pm 0.021 \text{ GeV}^1 \quad (1.1)$$

¹ $\hbar=c=1$ in the following

Table 1.3: Gauge bosons and fundamental interactions

	Electromagnetic	Weak	Strong
Field quantum	photon (γ)	W^\pm, Z	8 gluons g_α
Coupling constant	α	G_F	α_S
Range (m)	∞	10^{-18}	10^{-15}

- **Higgs boson**

It is an hypothetical scalar particle (spin 0) theorized independently by Robert Brout, François Englert, Peter Higgs, Gerald Guralnik, C. R. Hagen, and Tom Kibble in 1964. The motivation that led to the introduction of a scalar field in the SM (no elementary scalar particle has been observed so far) is related to the mechanism of spontaneous symmetry breaking, which will be discussed in details later. It plays a unique role in the SM, explaining why, for example, photon is massless and Z is massive or how the other elementary particles get their masses.

So far, no direct evidence of Higgs boson was found in the experiments. Probably LHC experiments will be able to solve definitively the question about its existence pretty soon.

1.1.1 Goldstone theorem and spontaneous symmetry breaking

The *Goldstone theorem* concerns spontaneous breaking of global symmetries and it states: every broken generator of a global symmetry group has a corresponding massless spin 0 *Goldstone boson*. The best way to illustrate the application of this theorem to field theories is to start with a very simple example, the broken global $U(1)$ symmetry.

Let's start from the lagrangian of a complex scalar field with mass m :

$$\mathcal{L} = \partial_\mu \phi^\dagger \partial^\mu \phi - m^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2 \quad (1.2)$$

where the last term is the potential. This lagrangian is invariant under global $U(1)$ transformation, $\phi \rightarrow e^{iQ\theta} \phi$, and Q is the conserved charged related to this symmetry. Setting $\mu^2 = -m^2 < 0$, we get the potential:

$$V(\phi^\dagger \phi) = -\mu^2 (\phi^\dagger \phi) + \lambda (\phi^\dagger \phi)^2 \quad (1.3)$$

The shape of this potential looks like a Mexican hat (figure 1.1). It has infinite minimum points at $\phi^\dagger \phi = \frac{\mu^2}{2\lambda}$, i.e. a infinite degeneracy of the minimum energy state, the vacuum. We can arbitrary choose one of them

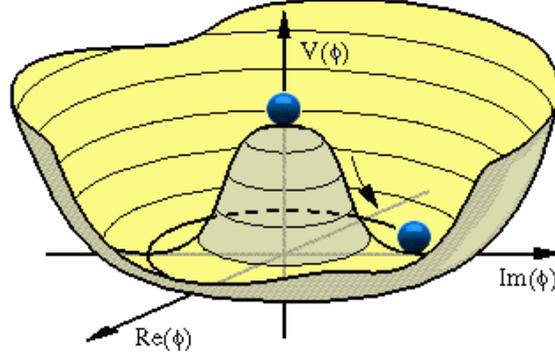


Figure 1.1: The shape of the potential (1.3) in a 3D view.

as the vacuum state, let's say $\text{Re}\phi = \frac{\mu}{\sqrt{2\lambda}}$ and $\text{Im}\phi = 0$. Then we define the vacuum expectation value (v.e.v.) v of the scalar field:

$$\langle \phi \rangle_0 = \langle 0 | \phi | 0 \rangle = \frac{\mu}{\sqrt{2\lambda}} = \frac{v}{\sqrt{2}} \quad (1.4)$$

To get the mass spectrum of the theory, we have to re-parametrize the field before, in order to have the usual definition of the field as the particle fluctuations around the vacuum. We write:

$$\phi(x) = \frac{1}{\sqrt{2}} (H(x) + v) e^{i\pi(x)/v} \quad (1.5)$$

where the "fluctuation fields" H and π have null v.e.v. Substituting eq. (1.5) into the lagrangian (1.2) we get:

$$\mathcal{L} = \frac{1}{2} (\partial_\mu H \partial^\mu H + \partial_\mu \pi \partial^\mu \pi) + \frac{1}{v^2} (H^2 + 2vH) \partial_\mu \pi \partial^\mu \pi - \lambda v^2 H^2 - \frac{1}{4} \lambda (H^4 + 4vH^3) + \frac{1}{4} \lambda v^4 \quad (1.6)$$

From eq. (1.6) we can see that the field H has acquired mass $m_H = \sqrt{2\lambda}v$, while π is massless. The π corresponds to the Goldstone boson cited in the Goldstone theorem, and it's easy to see that the broken generator is the $U(1)$ charge Q , in fact:

$$Q|0\rangle \neq 0 \quad (1.7)$$

i.e. the vacuum it's not invariant under $U(1)$ transformations generated by Q . No unbroken generators remain, so the symmetry breaking pattern is $U(1) \rightarrow \text{nothing}$.

The meaning of *spontaneous symmetry breaking* (SSB) is the following: the symmetry it's not explicitly broken by any term in the lagrangian, which is still invariant under the group transformations, but after the re-parametrization of the scalar field as variations around the vacuum then the

vacuum state does not realize the symmetry anymore.

To fully appreciate the role of Goldstone bosons in the SSB we have to consider local gauge symmetry, as it is done in the following.

1.1.2 The Higgs mechanism and the Glashow-Weinberg-Salam model of electroweak interaction

In contrast to the electromagnetic interaction, the weak interaction involves the exchange of massive vector bosons, namely W^\pm and Z , which are responsible respectively for weak charged and neutral currents. In building a field theory for weak interactions with local gauge invariance any mass term for vector bosons breaks explicitly the symmetry. To preserve gauge invariance and get mass for W and Z , the spontaneous symmetry breaking mechanism is invoked: the lagrangian preserves gauge invariance, while the vacuum state it's not invariant under gauge group transformations.

The model that correctly describes weak and electromagnetic interactions in a unified gauge theory is known as *Glashow-Weinberg-Salam model* (GWS), that is the a field theory with gauge group $SU(2)_L \otimes U(1)_Y$ spontaneously broken:

$$SU(2)_L \otimes U(1)_Y \rightarrow U(1)_{EM} \quad (1.8)$$

It gives the right masses to W , Z and γ through the Higgs mechanism.

Example of Higgs mechanism

The Higgs mechanism is the extension of the spontaneous symmetry breaking in order to give mass to the vector bosons of a local gauge theory. The most interesting application of it is the GWS model, as it will be discussed later in details. Here a simplified example is given for a local $U(1)$ symmetry.

Consider the lagrangian:

$$\mathcal{L} = (D_\mu \phi)^\dagger (D^\mu \phi) + \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2, \quad \mu^2 < 0, \lambda > 0$$

where ϕ is a complex scalar field and D_μ is the usual *covariant derivative* define as:

$$D_\mu = \partial_\mu + ieQA_\mu$$

The minimum of the potential is in $\langle \phi \rangle_0 = \frac{v}{\sqrt{2}}$ where $v^2 = -\frac{\mu^2}{\lambda}$. As we know from field theory, oscillations around the vacuum correspond to real particles, so we can rewrite the scalar field as:

$$\phi(x) = \frac{1}{\sqrt{2}}(v + H(x))$$

where H is the Higgs field and we have used a gauge transformations to eliminate the Goldstone boson π :

$$\begin{cases} \phi \rightarrow e^{iQ\theta(x)}\phi \\ A_\mu \rightarrow A_\mu - \frac{1}{e}\partial_\mu\theta(x) \end{cases}, \theta(x) = -\frac{\pi(x)}{v} \quad (1.9)$$

Now, looking at the kinetic term in the lagrangian, we get:

$$(D_\mu\phi)^\dagger(D^\mu\phi) = \frac{1}{2}\partial_\mu H\partial^\mu H + \frac{1}{2}e^2Q^2v^2A_\mu A^\mu + \frac{1}{2}e^2Q^2(h^2 + 2hv)A_\mu A^\mu$$

that means that A_μ (our "photon") has acquired mass: $m_A = eQv$. From the gauge transformation (1.9) is clear the role of Goldstone boson in the Higgs mechanism: it vanishes giving its degree of freedom to the gauge field, which in turn acquires a mass, hence a new polarization.

The GWS model

The $SU(2)$ symmetry group is related to the weak interactions, but in order to break correctly the symmetry it's necessary to start from an higher symmetry. Glashow was the first who worked with $SU(2) \otimes U(1)$, but only in 1967 - 1968 with the work of Weinberg and Salam a new gauge theory was built on this group, that led to the unification of electromagnetic and weak interactions under $SU(2)_L \otimes U(1)_Y$.

The general idea of the model is to start with massless gauge bosons, introduce a scalar field with non-vanishing vacuum expectation value and spontaneous symmetry breaking potential to give mass to W^\pm and Z and leave the photon massless.

Being W_μ^a ($a = 1,2,3$) the gauge fields and g the coupling constant of $SU(2)$ and B_μ and g' the ones for $U(1)$, let's write the covariant derivative as:

$$D_\mu\phi = \partial_\mu\phi + igT^aW_\mu^a\phi + ig'YB_\mu\phi \quad (1.10)$$

In equation (1.10) $T^a = \frac{1}{2}\sigma^a$ are the generators of $SU(2)$ (σ^a are the Pauli matrices) and $Y = 1/2$ is the hypercharge value for the scalar field. In eq. (1.10) is implicit the choice of the representation for the scalar field, which is a doublet: $\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$.

We now introduce the lagrangian for the scalar sector:

$$\mathcal{L} = (D_\mu\phi)^\dagger(D^\mu\phi) + \mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2, \mu^2 < 0, \lambda > 0 \quad (1.11)$$

The minimum of the potential is in:

$$\phi^\dagger\phi = \frac{v^2}{2} \quad (1.12)$$

which is an infinitely degenerate point. Hence we can arbitrarily choose:

$$\langle \phi \rangle_0 = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix} \quad (1.13)$$

Now it's easy to notice that all generators are broken with this choice of $\langle \phi \rangle_0$, except one linear combination of them:

$$Q = T^3 + Y \quad (1.14)$$

in fact $Q \langle \phi \rangle_0 = 0$. Q is the electric charge, and the charge values for ϕ are the one indicated from the beginning.

The spontaneous symmetry breaking pattern is therefore the one anticipated in eq. (1.8). With 3 broken generators we have three Goldstone bosons, which can be eliminated, as usual, taking advantage of gauge invariance (*unitary gauge*). We can then write:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ H(x) + v \end{pmatrix} \quad (1.15)$$

H is the *Higgs boson*.

Substituting eq. (1.15) into (1.11) we get from the kinetic term:

$$(D_\mu \phi)^\dagger (D^\mu \phi) = \frac{1}{2} \partial_\mu H \partial^\mu H + \frac{1}{8} (H + v)^2 \left[(-gW_\mu^3 + g'B_\mu)^2 + g^2 (W_\mu^+ W^{+\mu} + W_\mu^- W^{-\mu}) \right] \quad (1.16)$$

where W^\pm are the charged SU(2) fields.

$$W_\mu^\pm = \frac{1}{\sqrt{2}} (W_\mu^1 \mp iW_\mu^2) \quad (1.17)$$

From eq. (1.16) is clear that both W bosons acquire mass $M_W^2 = \frac{1}{4}g^2v^2$.

We see that also $-gW_\mu^3 + g'B_\mu$ is a massive field. It is convenient to do a rotation to get the proper linear combinations for the massless field (A_μ , the photon) and the massive one (Z_μ):

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_w & -\sin\theta_w \\ \sin\theta_w & \cos\theta_w \end{pmatrix} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix} \quad (1.18)$$

where θ_w is *Weinberg angle* or *weak mixing angle*. From the normalization of Z and A fields:

$$\begin{cases} Z_\mu = \frac{1}{\sqrt{g^2 + g'^2}} (gW_\mu^3 - g'B_\mu) \\ A_\mu = \frac{1}{\sqrt{g^2 + g'^2}} (g'W_\mu^3 + gB_\mu) \end{cases} \quad (1.19)$$

so θ_w can be written in terms of g and g' :

$$\cos\theta_w = \frac{g}{\sqrt{g^2 + g'^2}}, \sin\theta_w = \frac{g'}{\sqrt{g^2 + g'^2}} \quad (1.20)$$

In terms of Z the the second part of eq. (1.16) can be rewritten as:

$$(D_\mu\phi)^\dagger(D^\mu\phi) = \frac{1}{8}(H+v)^2 [(g^2 + g'^2) Z_\mu Z^\mu + g^2 (W_\mu^+ W^{+\mu} + W_\mu^- W^{-\mu})] \quad (1.21)$$

From eq. (1.21) we see that $M_Z^2 = \frac{1}{4}(g^2 + g'^2)v^2 \geq M_W^2$, more precisely:

$$M_W = M_Z \cos\theta_w \quad (1.22)$$

which is a relation valid only at tree level. M_W and M_Z can be related to well-known physical constants, such as the *Fermi constant* G_F , the fine structure constant α and $\sin^2\theta_w$, and its actual accepted values are those of eq. (1.1):

$$M_W = \left(\frac{\alpha\pi}{G_F\sqrt{2}} \right)^{1/2} \frac{1}{\sin\theta_w} \quad (1.23)$$

From (1.23) and from $M_W^2 = \frac{1}{4}g^2v^2$ we can get a prediction of Higgs field v.e.v. from the measurement of one electroweak constant, for example G_F :

$$v = \left(\frac{1}{\sqrt{2}G_F} \right)^{1/2} \simeq 246 \text{ GeV} \quad (1.24)$$

We have seen how we can get the correct masses from the spontaneous symmetry breaking of $SU(2)_L \otimes U(1)_Y$ group through the Higgs mechanism: again this can be interpreted as the massive bosons "eating" the Goldstone bosons (namely $\text{Re}\phi^+$, $\text{Im}\phi^+$ and $\text{Im}\phi^0$) and getting the longitudinal polarization while getting mass.

1.1.3 Yukawa coupling

Fermions fields can be split into a left-handed and right-handed part, that are chirality eigenstates, according to:

$$\psi = \psi_L + \psi_R, \psi_{L,R} = \frac{(1 \mp \gamma^5)}{2}\psi \quad (1.25)$$

Left and right part of fermion fields fill different multiplets of electroweak gauge group, to account for parity violation of weak interactions. For the first generation we can write:

$$q_L \equiv \begin{pmatrix} u_L \\ d_L \end{pmatrix}, u_R, d_R, l_L \equiv \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}, e_R \quad (1.26)$$

We cannot introduce in the fermion lagrangian an explicit mass term like $m\bar{\psi}\psi$ because it's not gauge invariant: from eq. (1.26) it's clear that ψ_L and ψ_R behave different under $SU(2)$ transformations. However, with a Higgs doublet as introduced in the GWS model there is a gauge invariant interaction that looks like a fermion mass term when the Higgs gets its vacuum expectation value, that is called *Yukawa coupling*. For the electron:

$$\mathcal{L}_{Yuk} = -Y_e \bar{L} \phi e_R + h.c. \quad (1.27)$$

where $h.c.$ is the hermitian conjugate. Using unitary gauge, eq. (1.15), we get, for the terms proportional to v :

$$-\frac{Y_e v}{\sqrt{2}} (\bar{e}_L e_R + \bar{e}_R e_L) = \frac{Y_e v}{\sqrt{2}} \bar{e} e \Rightarrow m_e = \frac{Y_e v}{\sqrt{2}} \quad (1.28)$$

From eq. (1.28) we can see that electron has acquired a mass, proportional to v . Like for vector bosons W and Z , also fermions get mass from spontaneous symmetry breaking: the mass is proportional to Y_f , which is the strength of the coupling of the fermion f to the Higgs.

For the down quark, the Yukawa coupling term is the same as for electron:

$$\mathcal{L}_{Yuk} = -Y_d \bar{q}_L \phi d_R + h.c. \quad (1.29)$$

but for the up quark we need something different, because the v.e.v. of Higgs field is placed in the down part of the doublet (see equation (1.13)). Defining:

$$\phi^C = \begin{pmatrix} \phi_0^* \\ -\phi_+^* \end{pmatrix} \quad (1.30)$$

we can get a new Yukawa invariant interaction for the up quark:

$$\mathcal{L}_{Yuk} = -Y_u \bar{q}_L \phi^C u_R + h.c. \quad (1.31)$$

Finally we can get the mass terms for u and d quarks too:

$$m_u = \frac{Y_u v}{\sqrt{2}}, \quad m_d = \frac{Y_d v}{\sqrt{2}}$$

1.2 Higgs properties

In the previous paragraphs the theoretical framework of Higgs boson was introduced, and the main consequences of the introduction of the scalar sector in the SM were illustrated. Although M_H is unknown, from eq. (1.21), (1.27), (1.29) and (1.31) Higgs couplings with gauge bosons and fermions are obtained and this make possible to set Feynman rules and get predictions of Higgs phenomenology as a function of its mass. In this section some aspects regarding Higgs mass, production mechanisms and decay channels are discussed, focusing on the Higgs physics program at the Large Hadron Collider.

1.2.1 Higgs mass

The tree-level relation for the Higgs mass reads $M_H = \sqrt{2\lambda}v$, thus Higgs mass depends on the unknown parameter λ that it's not possible to predict from the theory. However, theoretical arguments, such as the so-called *stability*, *triviality* and *unitarity* conditions set lower and upper bounds on M_H (figure 1.2). Higgs potential is affected by radiative corrections, i.e.

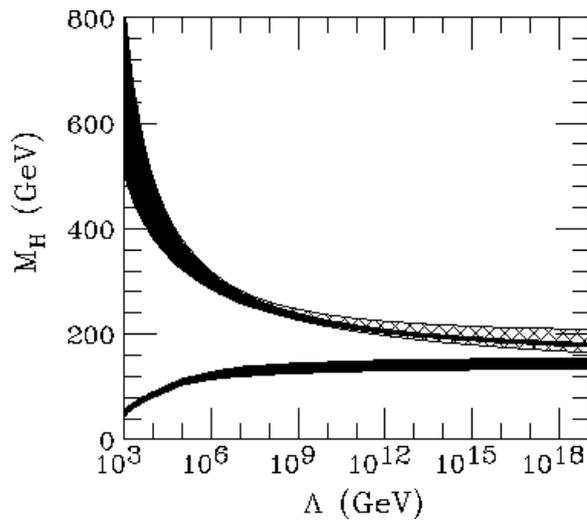


Figure 1.2: Lower and upper theoretical bounds on M_H

quantum loop corrections to the classical value. These corrections can lead to a potential unbounded from below, hence unphysical. This requirement of stability is equivalent to the requirement that the running coupling constant $\lambda(Q^2)$, which is related to Higgs mass, never becomes negative up to a scale Λ where new physics appear. This constraint can be inverted to set a lower limit on M_H which is a function Λ and the mass of the top quark m_t . Calculations show that for the nominal value of the top mass ($m_t = 172.9$ GeV) and $\Lambda = m_{Planck} \simeq 10^{19}$ GeV then $M_H \geq 130$ GeV.

From the requirement that no Landau pole appears in Higgs self couplings we can get an upper bound on M_H .

It can be shown that:

$$\lambda(\mu) \leq \frac{2\pi^2}{3 \ln\left(\frac{\Lambda}{\mu}\right)} \quad (1.32)$$

Eq. (1.32) shows that $\lambda \rightarrow 0$ when $\Lambda \rightarrow \infty$, i.e. if no new physics exists at scale Λ then the SM stops to make sense and become a non-interacting theory (a "trivial" theory). If instead we imagine that the SM is embedded in a more complete theory having new physics at an energy scale Λ the re-

quirement can be loosened to get an upper limit on M_H . If SM is valid up to $\Lambda \simeq m_{Planck}$, $M_H \lesssim 140$ GeV, while if we set $\Lambda \simeq 1$ TeV we get $M_H \lesssim 750$ GeV.

Another upper bound on Higgs boson mass comes from unitarity arguments. It can be shown that the scattering of longitudinal W and Z (W_L, Z_L) violates unitarity at high energy. Introducing the Higgs boson restores the unitarity and this procedure sets an upper limit on Higgs quartic coupling constant, which can be turned, as usual, into a constraint on Higgs mass, which for $W_L W_L$ scattering reads:

$$M_H \leq \sqrt{\frac{8\pi\sqrt{2}}{5G_F}} \simeq 780 \text{ GeV} \quad (1.33)$$

1.2.2 Higgs production

In proton-proton collisions at $\sqrt{s} = 7 - 14$ TeV, like those at the Large Hadron Collider, the main mechanisms of production of the Higgs boson are:

- *Gluon-gluon fusion: $gg \rightarrow H$*
This is the largest rate mode for all M_H , as it's clear from figure 1.4 and it's proportional to the Yukawa coupling Y_t .
- *Vector boson fusion (VBF): $qq \rightarrow qqH$*
This is the second rate mode and it's proportional to the WWH (ZZH) coupling.

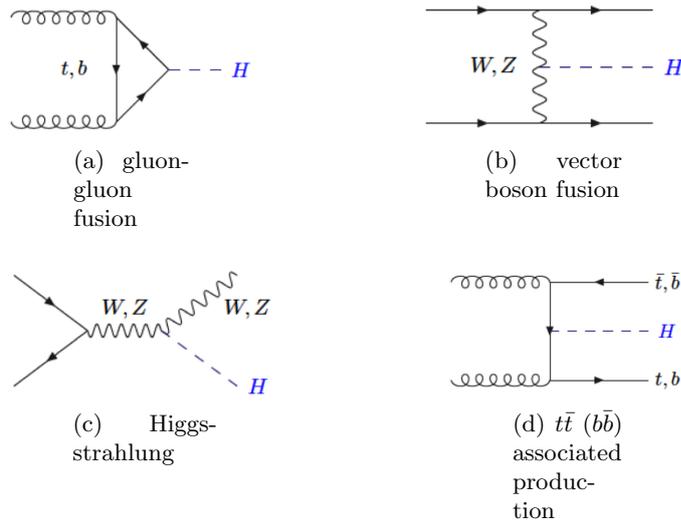


Figure 1.3: Higgs production mechanisms

- *Higgs-strahlung*: $q\bar{q} \rightarrow W(Z)H$
This is proportional to WWH (ZZH) coupling too, and it's the third largest rate for the production rate.
- $t\bar{t}(b\bar{b})$ associated production: $gg \rightarrow t\bar{t}(b\bar{b})H$
It's proportional to the Yukawa coupling of the heavy quark involved, and it's the lowest rate for every M_H .

In figure 1.4 the Higgs cross sections for the different production mechanisms are shown as a function of M_H .

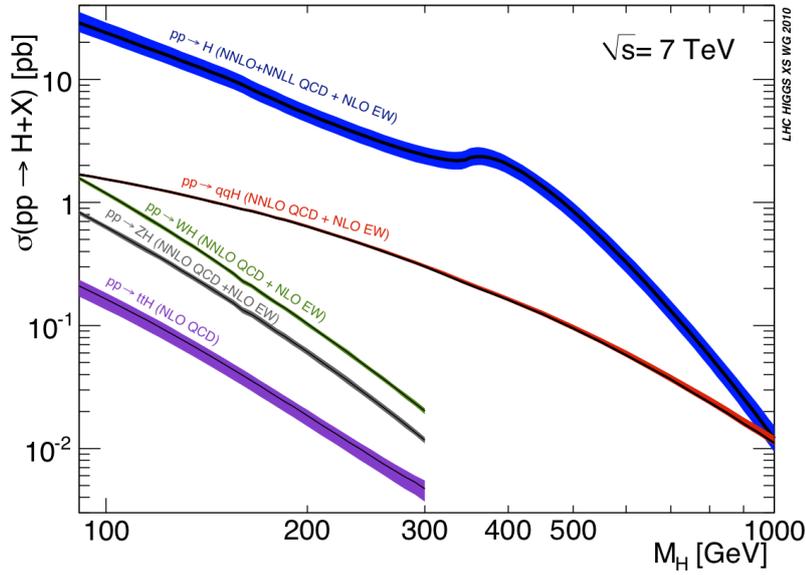


Figure 1.4: Higgs production cross section at $\sqrt{s}=7$ TeV.

1.2.3 Higgs decay channels

From eq. (1.21), (1.27), (1.29) and (1.31) we see that SM Higgs boson couples both to gauge bosons and fermions, with a strength depending on the mass:

$$g_{Hf\bar{f}} = \frac{m_f}{v} \quad g_{HVV} = \frac{M_V^2}{v^2} \quad (1.34)$$

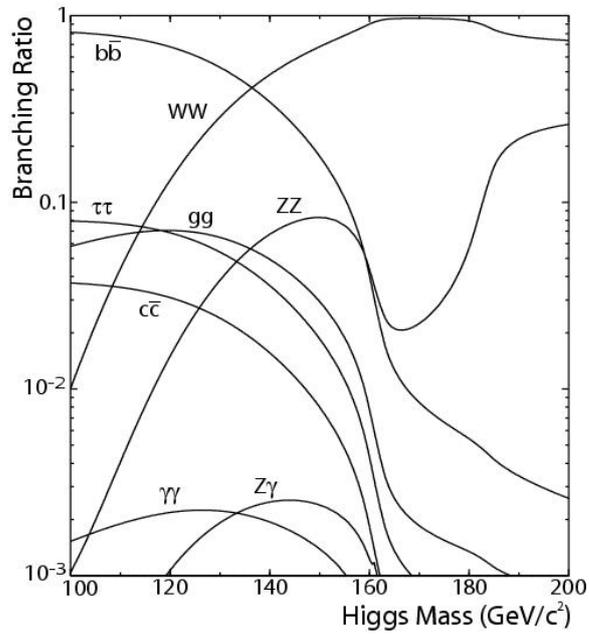
Thus Higgs tends to decay to the heaviest SM particle that is kinematically accessible.

Higgs branching ratios can be calculated as a function of the unknown M_H , as it is shown in figure 1.6.

It seems natural to divide the Higgs mass range in:



Figure 1.5: SM Higgs couplings

Figure 1.6: Higgs branching ratio as a function of m_H

- *Low mass region* ($M_H < 130$ GeV), where the dominating channel is $H \rightarrow b\bar{b}$. Other channels considered in this range are $H \rightarrow \tau^+\tau^-$ and $H \rightarrow \gamma\gamma$.
- *Intermediate mass region* (130 GeV $< M_H < 180$ GeV), where fermionic decays start to be less important and decays into a pair of vector bosons (W and Z) start to grow. The decrease of $H \rightarrow ZZ^{(*)}$ branching ratio around 160-170 GeV is due to the threshold of two W on shell, which increases the $H \rightarrow W^+W^-$ channel.
- *High mass region* ($M_H > 180$ GeV), where the most important channels are $H \rightarrow W^+W^-$ and $H \rightarrow ZZ$ with both vector bosons on shell.

The main decay channels under study correspond to the following inclusive signatures:

- $W(Z)H(\rightarrow b\bar{b})$
This analysis relies on the reconstruction of a W or Z boson in the associated production with the Higgs boson in order to reduce the background of dijet events, and looks for a resonant pair of b-tagged jets. This signature is important in the low mass region where the decay into a $b\bar{b}$ is dominant.
- $H \rightarrow \tau^+\tau^-$
For this channel, four different tau pairs final states can be considered, depending on whether there is one or two τ decaying leptonically: $e\tau_h$, $\mu\tau_h$, $\mu\mu$, or $e\mu$, where τ_h is a hadronically decaying τ . For the first and second mode, a reconstructed τ with opposite charge with respect to an isolated leptons is required. For the $\mu\mu$ and $e\mu$ final states, two isolated oppositely charged leptons are required.
- $H \rightarrow \gamma\gamma$
Despite its small branching ratio (see figure 1.6), this is one of the most promising channel for the low mass region. It provides a clean final state topology and a great precision in the mass peak reconstruction.
- $H \rightarrow W^+W^- \rightarrow 2l2\nu$ ($l = e, \mu$)
This channel requires both W decaying leptonically, resulting in the search for two high p_T isolated leptons and large missing transverse energy (MET).
- $H \rightarrow ZZ^{(*)} \rightarrow l^\pm l'^\mp l'^\pm l'^\mp$ ($l, l' = e, \mu$)
This search relies only on the measurements of leptons, requiring 2 pairs of same flavor and opposite sign leptons.
- $H \rightarrow ZZ \rightarrow 2l2\nu$ ($l = e, \mu$)
These events are characterized by the presence of a boosted Z boson decaying to an e^+e^- or $\mu^+\mu^-$ pair (high p_T isolated leptons) and large MET arising from the decay of the other Z boson decaying into neutrinos.
- $H \rightarrow ZZ \rightarrow l^+l^-2\text{jets}$ ($l = e, \mu$)
For these signature one Z is required to decay leptonically, the other one to a $q\bar{q}$ pair, which is reconstructed as 2 jets in the detector. This channel has a branching ratio 20 times larger than the fully leptonic one, and can lead to better sensitivity at higher masses, where background can be suppressed kinematically.

1.3 Experimental results in the SM Higgs boson search

In this section the results for direct and indirect search for the Higgs boson in the experiments at LEP and Tevatron are presented, together with the latest results coming from the experiments at LHC.

1.3.1 Indirect Higgs search

In the SM many parameters, such as $\rho = \frac{M_W^2}{M_Z^2 \cos^2 \theta_w} = 1$, $\sin^2 \theta_w$, M_W^2 and α receive radiative corrections from loop diagrams like those in figure 1.7.

$$\bar{\rho} = 1 + \Delta\rho \quad (1.35)$$

$$\sin^2 \theta_{eff} = (1 + \Delta\kappa) \sin^2 \theta_w \quad (1.36)$$

$$M_W^2 = \frac{\pi\alpha}{\sqrt{2} \sin^2 \theta_w G_F} (1 + \Delta r) \quad (1.37)$$

$$\alpha(M_Z^2) = \frac{\alpha(0)}{1 - \Delta\alpha}, \quad \Delta\alpha = \Delta\alpha_{lept} + \Delta\alpha_{top} + \Delta\alpha_{had}^{(5)} \quad (1.38)$$

In eq. (1.35) - (1.38) the radiative corrections $\Delta\rho$, $\Delta\kappa$, Δr , $\Delta\alpha$ are functions of m_t^2 and $\log M_H$.

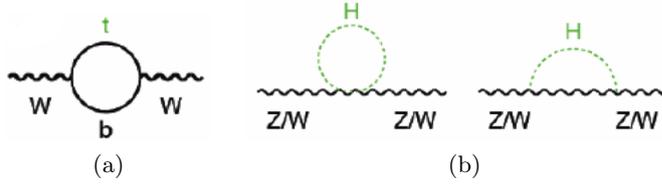


Figure 1.7: t-quark loop corrections $\propto m_t^2$ (a) and Higgs loop corrections $\propto \log \frac{M_H}{M_{Z,W}}$ (b)

In principle, it's possible to get predictions of the unknown parameters like the Higgs mass provided precision measurements of electroweak physical constants are performed. This is what was done at LEP to get predictions on the mass of top quark and the mass of the Higgs boson, before the discovery of the top quark in 1995. Including the measurement of the mass of the top quark, a global fit on the whole SM was performed with only M_H as a free parameter and the result is shown in figure 1.8.

This test tends to exclude the Higgs for $M_H < 154$ GeV at 95% C.L. and the value that minimize the χ^2 is $M_H = 84_{-26}^{+34}$ GeV.

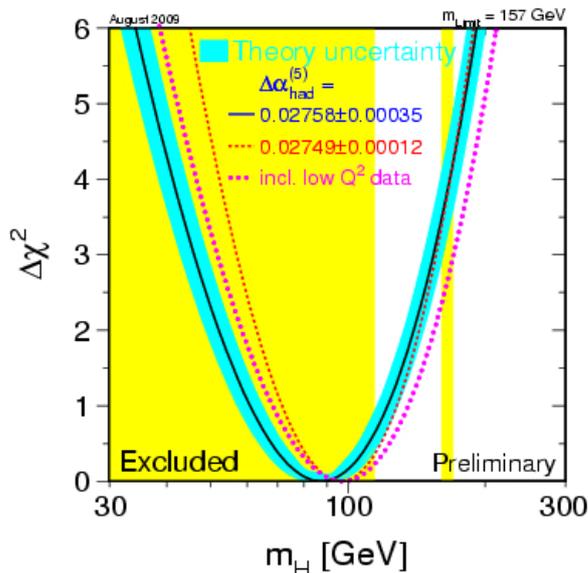


Figure 1.8: The global fit (χ^2) as a function of M_H

1.3.2 Direct Higgs search at LEP

The Large Electron Positron collider (LEP), built at CERN, collided e^+e^- from 1989 to 2000, first at $\sqrt{s} = 189$ GeV then at $\sqrt{s} = 209$ GeV. During this period four experiments (ALEPH, DELPHI, L3 and OPAL) collected data for a total integrated luminosity of about 600 pb^{-1} per experiment.

At LEP the main contribution for the production of the Higgs is expected to be due to the Higgs-strahlung process $e^+e^- \rightarrow HZ$. In the mass range relevant for these energies, Higgs is expected to decay mainly to a $b\bar{b}$ pair, about $\simeq 74\%$ for $M_H = 115$ GeV. Other contributions to the total width comes from $\tau^+\tau^-$, gg , WW^* and $c\bar{c}$.

A combination of the results of the four LEP experiments was published in 2003 (see [2]). The test statistic used is $-2 \ln Q$ where Q is the ratio of the likelihood function for the signal plus background ($s + b$) hypothesis to the likelihood function for the background only hypothesis (b):

$$Q = \frac{\mathcal{L}_{s+b}}{\mathcal{L}_b} \quad (1.39)$$

For the definition of the likelihood functions we address to [2].

In the modified frequentist approach in order to set exclusion limits the

confidence level CL_s is used, which is defined as:

$$CL_s = \frac{CL_{s+b}}{CL_b} \quad (1.40)$$

where, as usual, CL_b is the p-value for the h hypothesis, i.e. the integral of the probability density function (PDF) from $-\infty$ to the observed value of the statistic. A mass hypothesis is excluded at 95% confidence level when the corresponding value of CL_s is less than 0.05.

In figure 1.9 CL_s is shown as a function of the tested Higgs mass. The

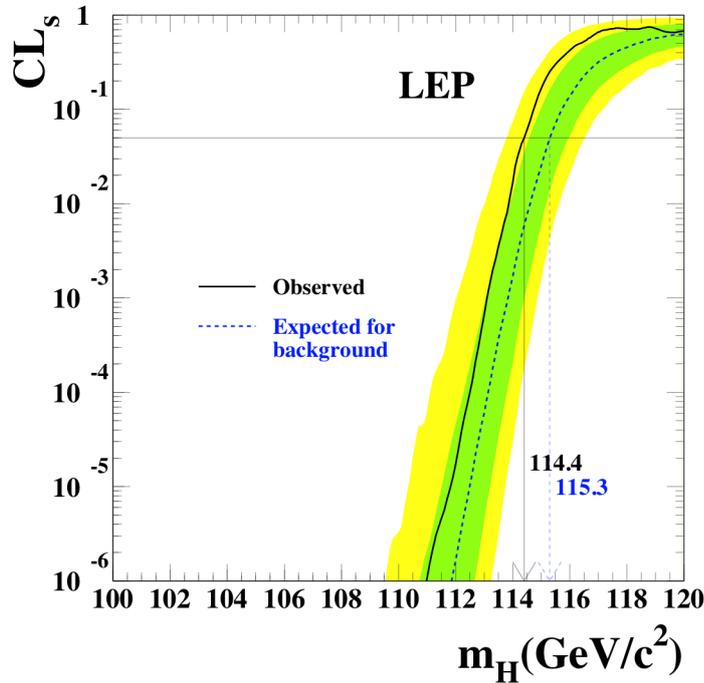


Figure 1.9: CL_s as a function of M_H . The green and yellow bands correspond to 68% and 95% probability bands.

intersection of the horizontal line for $CL_s = 0.05$ with the observed curve is used to define the 95% confidence level lower bound on the mass of the Standard Model Higgs boson. Thus LEP experiments exclude the Higgs boson for $M_H \lesssim 114.4$ GeV at 95% C.L.

1.3.3 Direct Higgs search at Tevatron

The Tevatron is a proton-antiproton collider in the United States (Illinois), at the Fermi International Laboratories (Fermilab), working at $\sqrt{s} =$

1.96 TeV. It started taking data in 1983 and recently the collaborations at Fermilab announced that it will be shut down in September 2011. It is the second world largest particle collider after the LHC at CERN, and it hosts two main experiments, CDF and D0. Mostly known for the discovery of the top quark in 1995, both experiments gave great contributions in the direct Higgs search too.

The most recent publication on the combined search for the SM Higgs boson, based on an integrated luminosity of 8.6 fb^{-1} for both experiments can be found at [3]. In this analysis, the following channels were studied: $WH \rightarrow l\nu b\bar{b}$, $ZH \rightarrow \nu\bar{\nu} b\bar{b}$, $WH + ZH + VBF \rightarrow b\bar{b}2\text{jets}$ and all $t\bar{t}H$ channels, for a Higgs mass range from 100 GeV to 200 GeV. A modified frequentist method is used too, as in LEP combination (see previous paragraph) and also a Bayesian method is applied. For the former the result is shown in figure 1.10. Another way to present the results is to show the ratio

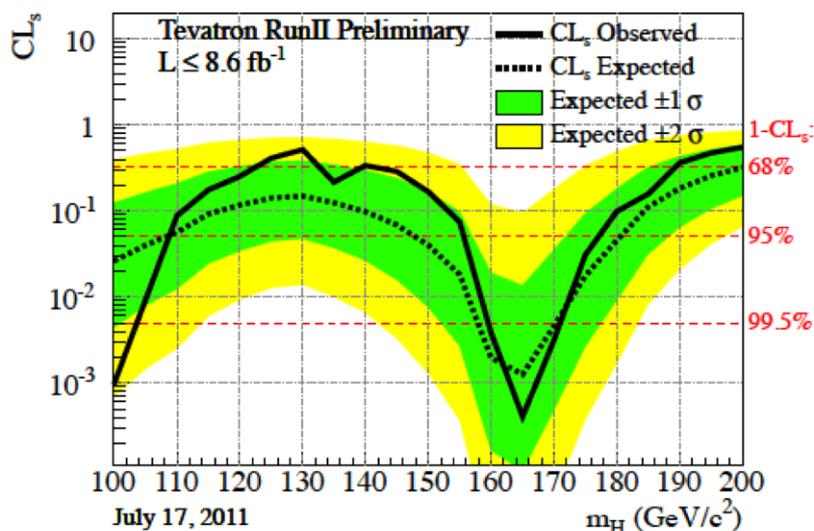


Figure 1.10: CL_s as a function of M_H for Tevatron combined Higgs search.

of the observed limit to the SM Higgs cross section, as a function of M_H (see figure 1.11): a value of the ratio which is less or equal to 1 indicates that the particular mass hypothesis is excluded at 95% C.L. Tevatron combination excludes at 95% C.L. the mass ranges $156 \text{ GeV} < M_H < 177 \text{ GeV}$ and $100 \text{ GeV} < M_H < 108 \text{ GeV}$. These results refer to the use of the Bayesian method, however the two methods give very similar results.

From figure 1.11 it's clearly visible a small excess in the region $145 < M_H < 155 \text{ GeV}$ ($\simeq 1\sigma$) that didn't allow the exclusion in this mass interval.

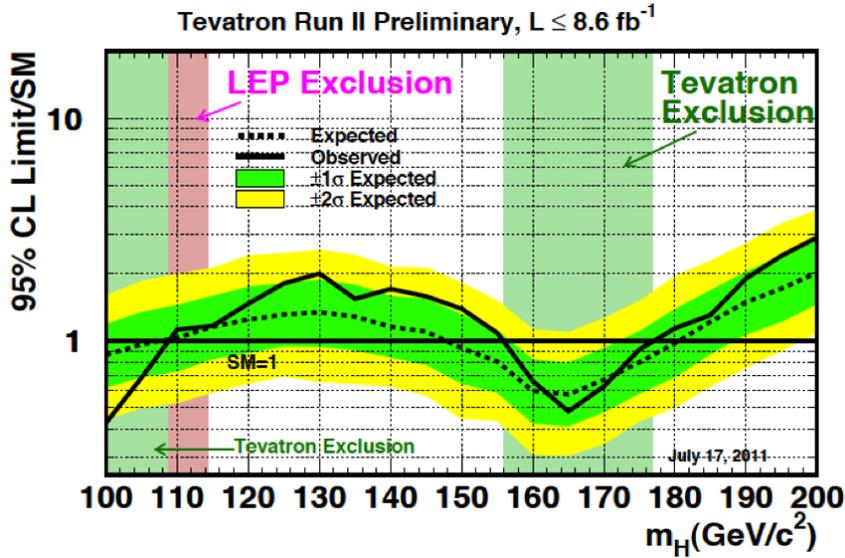


Figure 1.11: The ratio of the observed limit to the SM Higgs cross section as a function of M_H for Tevatron combination.

1.3.4 Direct Higgs search at LHC

The Large Hadron Collider (LHC) is the largest particle accelerator ever built, colliding protons and heavy ions at unprecedented energy and luminosity. It hosts three main experiments: CMS, ATLAS and ALICE, the last being especially dedicated to heavy ion physics. For a more detailed description of the LHC machine and CMS detector see chapter 2.

Here we present a summary of the analysis about the combination of eight channels of the SM Higgs boson studied within the CMS collaboration (see [4]). These 8 signatures are: $H \rightarrow \gamma\gamma$, $H \rightarrow \tau\tau$, $H \rightarrow bb$, $H \rightarrow WW \rightarrow 2l2\nu$, $H \rightarrow ZZ \rightarrow 4l$, $H \rightarrow ZZ \rightarrow 2l2\tau$, $H \rightarrow ZZ \rightarrow 2l2\nu$ and $H \rightarrow ZZ \rightarrow 2l2q$. The amount of data is different from channel to channel and varies from 1.1 - 1.7 fb^{-1} . In this analysis a frequentist method is used, which is based on the definition of the confidence level CL_s as said above for LEP and Tevatron results. In figure 1.12 the CL_s value for the SM Higgs hypothesis is shown as a function of Higgs boson mass.

In figure 1.13 the ratio of the observed limit to the SM Higgs cross section, also known as "signal strength modifier", is shown as a function of M_H .

The SM Higgs boson is excluded at 95% C.L. in three mass ranges 145-216, 226-288, and 310-400 GeV. The expected exclusion in the absence of a signal is 130-440 GeV.

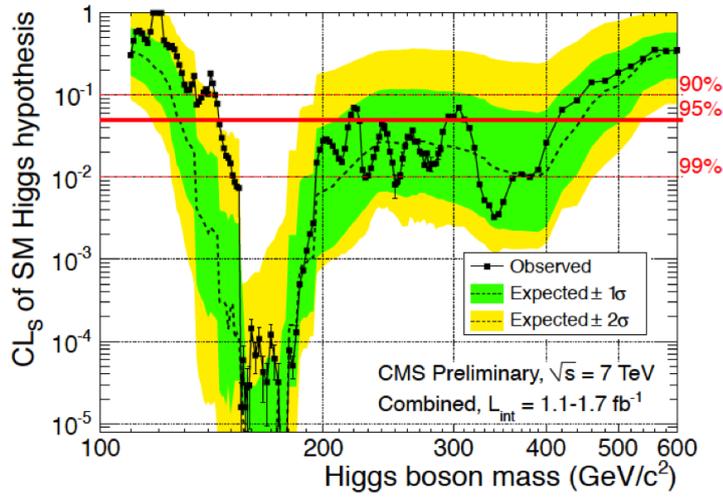


Figure 1.12: CL_s value for the SM Higgs boson hypothesis as a function of Higgs boson mass in the range 110 GeV - 600 GeV. As usual the dashed black line is the median expected for the background-only hypothesis, the solid black line is the observed one. Green and yellow bands indicate respectively 1σ and 2σ deviations from the background hypothesis.

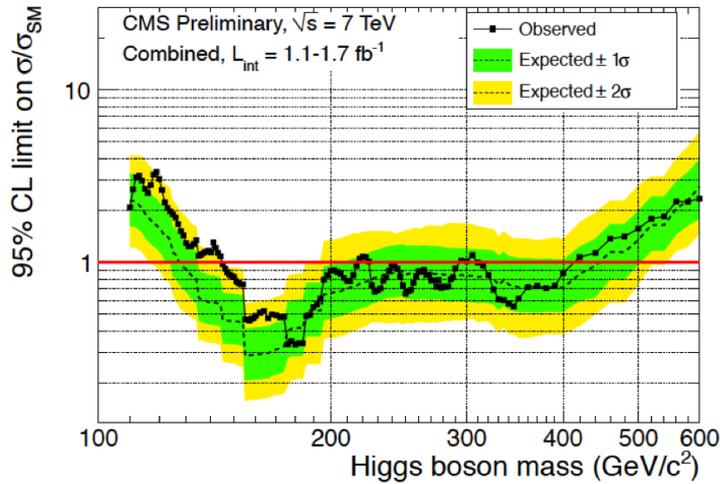


Figure 1.13: The combined 95% C.L. upper limits on the signal strength modifier as a function of Higgs boson mass. The black points and the black solid line indicates the observed limits, the dashed black line indicates the median expected limit in the background only hypothesis. The green (yellow) band indicates 1σ (2σ) departures from the median.

Chapter 2

The Large Hadron Collider and the CMS Detector

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is the world's largest and highest energy particle accelerator ever built. Proposed and realized by the European Organization for Nuclear Research (CERN), it was design to collide protons, as well as lead ions, at an unprecedented energy and rate, in order to address some of the most fundamental questions of physics.

In the following we give a short description of the design of the LHC and an introduction to the experiments. A more detailed description of the Compact Muon Solenoid experiment can be found in section 2.2.

2.1.1 Design and performances

The LHC lies in a circular tunnel, 27 km long and 3.8 km wide, at a depth varying from 50 to 175 m underground. This tunnel was constructed between 1983 and 1988, at the boundary between Switzerland and France, and was formerly used to host the Large Electron Positron collider (LEP) and its experiments.

The LHC is mainly designed to collide two beams of protons, each one with a nominal energy of 7 TeV, for a total centre-of-mass energy of 14 TeV. However, it accelerates lead ions beams as well, at a energy of 2.76 TeV/nucleon, in order to fulfill the heavy ion physics program.

The tunnel contains two adjacent and parallel beam pipes, where proton (or ion) beams travel in opposite directions and intersect in four points, where the main experimental halls are built and detectors are placed (see figure 2.1). Some 1232 dipole magnets keep the beams on their circular path, while additional 392 quadrupole magnets are used to keep the beams focused, in order to maximize the chances of interaction in the four intersection points,

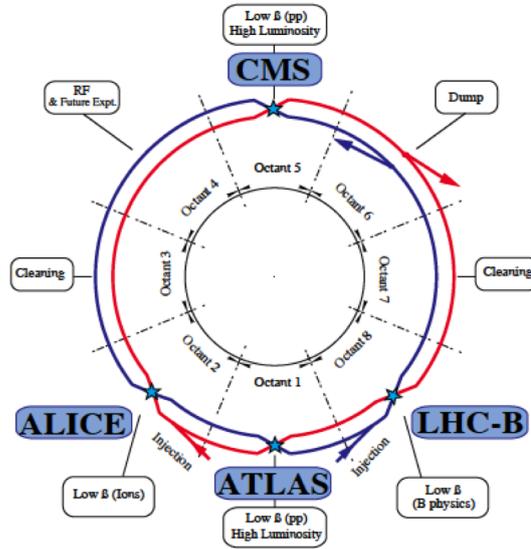


Figure 2.1: A schematic picture of the LHC layout.

where the two beams cross. In total, over 1600 superconducting magnets are installed. Approximately 96 tonnes of liquid helium is needed to keep the superconducting magnets at their operational temperature of 1.9 K. The field in the magnets increase from 0.53 T to 8.3 T while the protons are accelerated from 450 GeV to 7 TeV.

Before being injected into the main accelerator, the protons are prepared by a series of systems that successively increase their energy. The first system is the linear particle accelerator (LINAC 2) generating 50 MeV protons, which feeds the Proton Synchrotron Booster (PSB). There the protons are accelerated to 1.4 GeV and injected into the Proton Synchrotron (PS), where they are accelerated to 26 GeV. Finally the Super Proton Synchrotron (SPS) is used to further increase their energy up to 450 GeV before they are at last injected into the main ring. Here the proton bunches are accumulated, accelerated (over a period of 20 minutes) to their peak energy, and finally circulated while collisions occur at the four intersection points (IP).

Two of the main experiments, ATLAS and CMS, are designed for a high luminosity regime, in order to catch the rare events of their physics programs. That's why the beam intensity, together with the beam energy, is a crucial parameter for the LHC. The need for high luminosity excludes the use of anti-proton (like Tevatron collider): the LHC is therefore designed as a proton-proton collider with separate magnet fields and vacuum chambers in the main arcs and with common sections only at the IP.

Assuming a Gaussian beam shape, the luminosity at LHC can be written in

terms of machine parameters as:

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma}{4\pi \epsilon_n \beta^*} F$$

where:

- N_b is the number of particles per bunch;
- n_b is the number of bunches per beam;
- f_{rev} is the revolution frequency;
- γ is the relativistic Lorentz factor;
- ϵ_n is the normalized transverse beam emittance;
- β^* is the optical Beta function at the collision points;
- F is the geometric luminosity reduction factor due to the crossing angle at the IP.

The designed peak luminosity for ATLAS and CMS is $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, but before reaching this value the LHC will have lower luminosity runs.

The main parameters of the LHC are summarized in table 2.1 and refer to the interaction points for ATLAS and CMS.

The first beam was circulated in the LHC tunnel in the morning of 10 September 2008, but 9 days after an accident happened causing a quench in about 100 magnets, dispersion of liquid helium and the rise of the temperature to about 100 K. The incident was initiated by a faulty electrical connection. Due to this accident, the original timeline was slightly changed: most of 2009 was spent to repair the damaged magnets and in November 2009 the first low-energy beams after the accident were circulated. On 30 March 2010, LHC set a record for high-energy collisions, by colliding proton beams at a centre-of-mass energy of 7 TeV. CERN has declared that the LHC will run through to the end of 2012, with a short technical stop at the end of 2011. The energy for 2011-2012 will be 3.5 TeV per beam. In 2013 the LHC will go into a longer shutdown to prepare for higher-energy (7 TeV per beam) running starting in 2014.

2.1.2 The LHC experiments

There are six experiments studying collisions at the LHC, every one with its own detector. The two larger experiments, ATLAS and CMS, are multi-purpose detectors, designed to study a wide range of physics channels. Two medium-size detectors, ALICE and LHCb, are intended for more specific phenomena and two much smaller size experiments, TOTEM and LHCf, are dedicate to study forward particles coming out of the interactions. In

Table 2.1: The main LHC design parameters.

	Injection	Collision
Beam data		
Proton energy	450 GeV	7000 GeV
Relativistic Lorentz factor (γ)	479.6	7461
Number of particles per bunch (N_b)	1.15×10^{11}	
Number of bunches per beam (n_b)	2808	
Circulating beam current	0.582 A	
Stored energy per beam	23.3 MJ	362 MJ
Transverse normalized emittance (ϵ_n)	$3.5 \mu\text{m rad}$	$3.75 \mu\text{m rad}$
Interaction data		
Inelastic cross section	60 mb	
Total cross section	100 mb	
Events per bunch crossing	-	19.02
Luminosity data		
Geometric luminosity reduction factor (F)	-	0.836
RMS bunch length	11.24 cm	7.55 cm
RMS beam size	$375.2 \mu\text{m}$	$16.7 \mu\text{m}$
Optical β function	18 m	0.55 m
Total luminosity lifetime	-	14.9 h
Peak luminosity	-	$1 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$

the following, these six experiments are briefly presented focusing on the physics channels they study. A more detailed description of CMS detector can be found in the following section.

ALICE

ALICE stands for A Large Ion Collider Experiment. It is designed to study the ultra-high energy regime of ion-ion collisions as well as proton-proton collisions. The aim is to explore the state of the matter that is called *quark gluon plasma* (QGP), which is believed to exist in extreme conditions of the matter, at very high temperature and density.

Scientists believe that the universe, a few microseconds after the Big Bang, was made of QGP, a phase of the matter in which quarks and gluons are no more confined in protons and neutrons, but rather they are together in a kind "primordial soup". These extreme conditions can be reproduced in a ion-ion collision (like Pb-Pb at the LHC), for few microseconds, provided the ions collide head-on at ultra-relativistic energy. The QGP can be investigated by looking at the particles coming out of the collisions, searching for "probes",

i.e. for specific final states which are sensible to the formation of QGP. ALICE detector is located at St Genis-Pouilly, France. It is designed with a central barrel with a forward muon spectrometer and it's 26 m long, 16 m high, 16 m wide (see figure 2.2).

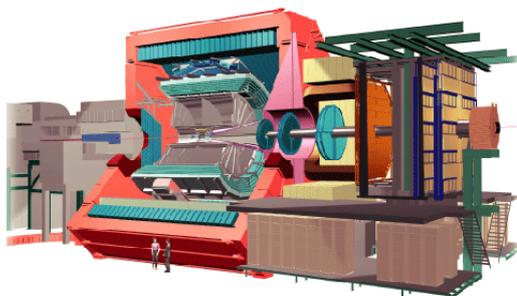


Figure 2.2: 3D view of the ALICE detector.

ATLAS & CMS

ATLAS and CMS (Compact Muon Solenoid) are multi-purpose detectors, designed to investigate a wide range of physics phenomena. Standard Model particles and processes will be measured with unprecedented precision and new physics will be investigated by these detectors. The new physics channels can be summarized as follows:

- **Higgs boson**

The electroweak symmetry breaking is invoked in the Standard Model (SM), as well as in many other theories beyond the Standard Model like Supersymmetry, in order to explain the origin of the mass of elementary particles. It predicts the existence of the Higgs boson, which has, nevertheless, never been observed in any experiment. ATLAS and CMS, with the 10-fold increase in luminosity and the 7-fold increase in the center-of-mass energy will probably give the definitive answer about the existence of the Higgs boson, and if it actually exists, they will investigate its properties (mass, width etc.).

- **Supersymmetry and Dark Matter**

Supersymmetry (SUSY) theories are able to solve many problems of SM and could provide a way to unify the electromagnetic, weak and strong forces. SUSY theories predict the existence of an higher broken symmetry which sets the correspondence between any SM particle with a super-partner (sparticle), which has same charge but different spin. The lightest neutral sparticle is one of the most promising candidate to form the dark matter, of which most of the universe is made but

could not be detected so far. ATLAS and CMS will look for signatures of particles, believed to be much heavier of the SM partners.

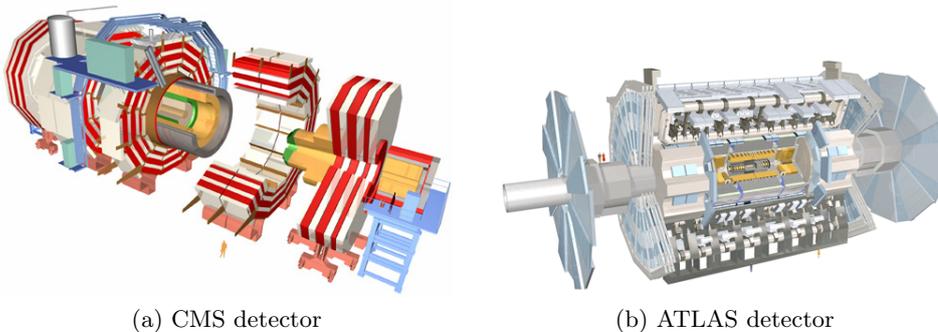
- **Extra dimensions**

Many theories, like string theory, predict additional dimensions with respect to the four of the well-known space-time. These dimensions could manifest themselves as new particles or little black-holes.

- **Matter - Antimatter asymmetry**

At the origin of the universe, matter and antimatter should have been created in equal amounts, but the universe today is dominated by matter. Subtle asymmetries between particles and anti-particles have been observed in B mesons, and this is also one of the main physics channel under investigation at the LHC.

Both CMS and ATLAS detectors are endowed with tracker system, calorimeters and muon spectrometer in order to measure momentum and energy of particles. They are both designed with a central barrel and end-caps on both sides of the barrel (see figure 2.3). The main difference is in the magnet system, which has a toroidal design for ATLAS and a solenoidal one for CMS. ATLAS is located in Meyrin, Switzerland, it's 46 m long, 25 m high and 25 m wide and weights about 7000 tonnes. CMS is located in Cessy, France, it's 21 m long, 15 m wide and 15 m high and weights approximately 12500 tonnes.



(a) CMS detector

(b) ATLAS detector

Figure 2.3: 3D view of CMS (a) and ATLAS (b) detectors.

LHCb

LHCb stands for LHC beauty. The LHCb experiment is designed to study the decays of B mesons, in order to investigate the asymmetry between matter and antimatter known as CP violation. LHCb physicists measure important CP violation properties looking for new subatomic mechanisms for the matter-antimatter imbalance. The design of LHCb detector (shown

in figure 2.4) allows the study of forward particles and comprises a forward spectrometer with planar subdetectors. It's 21m long, 10m high and 13m wide, and has a total weight of about 5600 tonnes. It's located in Ferney-Voltaire, France.

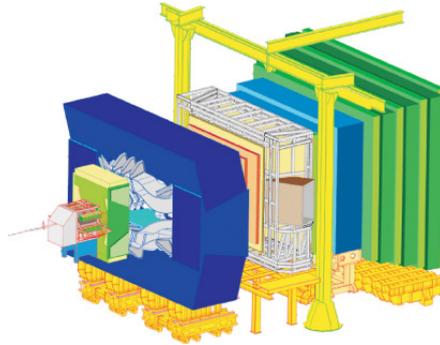


Figure 2.4: 3D view of the LHCb detector.

LHCf

The LHCf ("f" means forward) experiment uses forward particles created inside the LHC as a source to simulate ultra-high energy cosmic rays in laboratory conditions. Cosmic rays collide with nuclei at the upper atmosphere level, producing cascades of particles reaching the ground. Studying how collisions inside the LHC cause similar cascades will help scientists to interpret and calibrate large-scale cosmic-ray experiments at ground level. Two identical detectors are placed on both sides of ATLAS experiment, each one measuring $30\text{ cm} \times 80\text{ cm} \times 10\text{ cm}$.

TOTEM

TOTEM stands for TOTAl Elastic and diffractive cross section Measurement. Its detector will measure the size of the proton and will monitor accurately the LHC's luminosity, by detecting particles very close to the beam line. Eight vacuum chambers called "Roman pots", containing GEM detectors and cathode strip chambers, will be placed in pairs in four different points near the CMS detector.

2.2 The CMS detector

2.2.1 Coordinate conventions

The CMS experiment uses a cylindrical coordinate system, whose origin is at the nominal collision point inside the detector. The y axis points verti-

cally upward, the x axis points radially towards the centre of the LHC and the z axis is along the beam direction. The azimuthal angle ϕ is measured in the $x - y$ plane from the x axis, while the polar angle θ is measured from the z axis. Instead of θ , the *pseudorapidity* η is often used, which is defined as:

$$\eta = -\ln \tan \frac{\theta}{2}$$

The transverse momentum p_T is defined from x, y components of the momentum and the transverse energy is defined as $E_T = E \sin \theta$ and the lack of energy in the transverse plane, the missing transverse energy or simply MET, is denoted with E_T^{miss} . The ΔR parameter is defined as:

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

2.2.2 Overall design

One of the four main experiments at LHC is CMS, which stands for Compact Muon Solenoid. In order to deal with LHC physics goals, CMS detector was designed to satisfy many requirements that can be summarized as:

- good muon identification, charge distinction capability up to $p \sim 1$ TeV, good momentum resolution and dimuon mass resolution;
- good diphoton, dielectron mass resolution up to $|\eta| < 2.5$;
- good dijet mass resolution up to $|\eta| < 5$;
- efficient b and τ tagging.

The design of the detector was dominated by the aspects related to magnetic field configuration: large bending power is mandatory to have good resolution in measuring charge particle of high momentum. This reason led to superconducting magnets technology in order to achieve the 4 T solenoidal field used in the center of the detector. 4 "muon stations" are added in the 1.5 m thick iron return yoke, made of layers of aluminium drift tube (DTs) in the barrel region, cathode strip chambers (CSCs) in the endcap region complemented by resistive plate chambers (RPCs) in the region $|\eta| < 1.6$ (see figure 2.5). In the inner cavity of the coil, the inner tracker and both the hadronic calorimeter (HCAL) and electromagnetic calorimeter (ECAL) are placed. The inner tracker consists of 3 layers of silicon pixel detectors, close to the interaction point, and 10 layers of silicon microstrip detectors for a total tracking diameter of 2.6 m. The ECAL is made of lead tungstate ($PbWO_4$) crystals and the scintillation light is detected by avalanche photodiodes (APDs) for the barrel and vacuum phototriodes (VPTs) for the

endcaps. In front of the ECAL a pre-shower system is placed for π^0 rejection. The barrel HCAL is a sampling calorimeter, made of brass and scintillator material in which light is channeled to photodetectors (hybrid photodiodes, HPDs) via wavelength-shifting (WLS) fibres. In order to reach a pseudorapidity coverage of $|\eta| < 5$, there is another calorimeter made of iron and quartz fibres in the endcap region, coupled to photomultipliers to read the Cherenkov light.

The overall dimensions of CMS detector are quite impressive: about 21 m

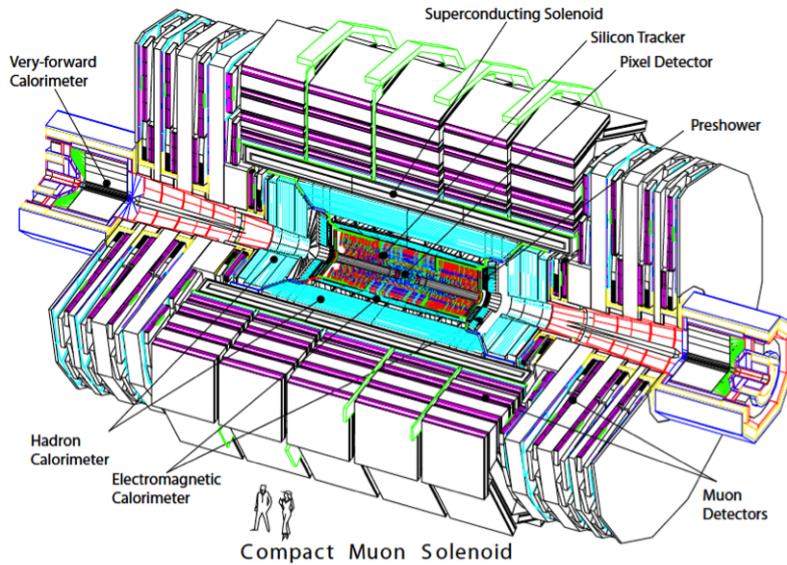


Figure 2.5: 3D view of CMS detector.

long, about 15 m thick, 12500 tons heavy. In 2000 the assembly work started in the surface hall, in the 2008 the first cosmic rays data were recorded and in March of 2010 the first 7 TeV collisions took place. While writing this thesis, CMS and the other LHC experiments are taking data at $\sqrt{s} = 7$ TeV with an unprecedented peak luminosity of $4.67 \cdot 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$, and they are able to collect $\sim 60 \text{ pb}^{-1}$ every day.

In the following, every component of the detector is briefly presented focusing on the performances and on their impact on physics analysis.

2.2.3 Magnet

As already said before, CMS aims to achieve a good momentum resolution, say $\Delta p/p \simeq 10\%$, even for momenta up to 1 TeV. The larger the bending power the better the momentum resolution, so even with a not enormous solenoid it's possible to meet this goal obtaining an high axial

magnetic field of 4 T with the superconducting technology. This makes it possible to avoid stringent demands on spatial resolution of muon chambers and on tracker alignment. Moreover, the axial direction of the field allows to start measure the momentum at $r = 0$ (unlike a toroidal configuration), resulting in a more compact design of the whole spectrometer. In table 2.2 the main parameters of the CMS magnet are listed.

Table 2.2: The main parameters of the CMS magnet.

Magnetic field	4 T
Inner bore	5.9 m
Length	12.9 m
Number of turns	2168
Current	19.5 kA
Stored energy	2.7 GJ
Total weight	~ 12000 tons

Superconducting magnets were already used in high energy physics successfully (LEP and HERA experiments) but new challenges came out for CMS to reach better performances.

CMS solenoid employs high purity aluminium conductors, with an overall cross section of $64 \times 22 \text{ mm}^2$. To keep the necessary low temperatures (about 4 K) an indirect termosyphon cooling is used together with epoxy resin impregnation.

2.2.4 Muon system

As only muons can penetrate many meters of iron without interacting, muon chambers are placed at the very edge of the experiment. Here muons are measured again just after the coil and for the third time in the return flux. In the muon chambers, momentum resolution is dominated by multiple scattering in the material before the chambers for low momentum muons ($p_T \leq 200 \text{ GeV}$), while detectors spatial resolution dominates for high momentum muons. As it was said before, muons are measured three times: the reason for this redundancy is that the combination of muon system and inner tracker measurements sensibly improves the resolution in the high momentum region, while for lower muon's momentua the only inner tracker is sufficient to get a good resolution and muon system does not improve it anymore.

Three different gas detector are used for the muon system: drift tubes (DTs), cathode strip chambers (CSCs) and resistive plate chamber (RPCs). Thanks to the fast response they provide, they are also used within the first level trigger system.

In figure 2.6 a quarter of the muon system is shown. The muon system is divided in barrel ($|\eta| < 1.2$) and endcap ($1.0 < |\eta| < 2.4$).

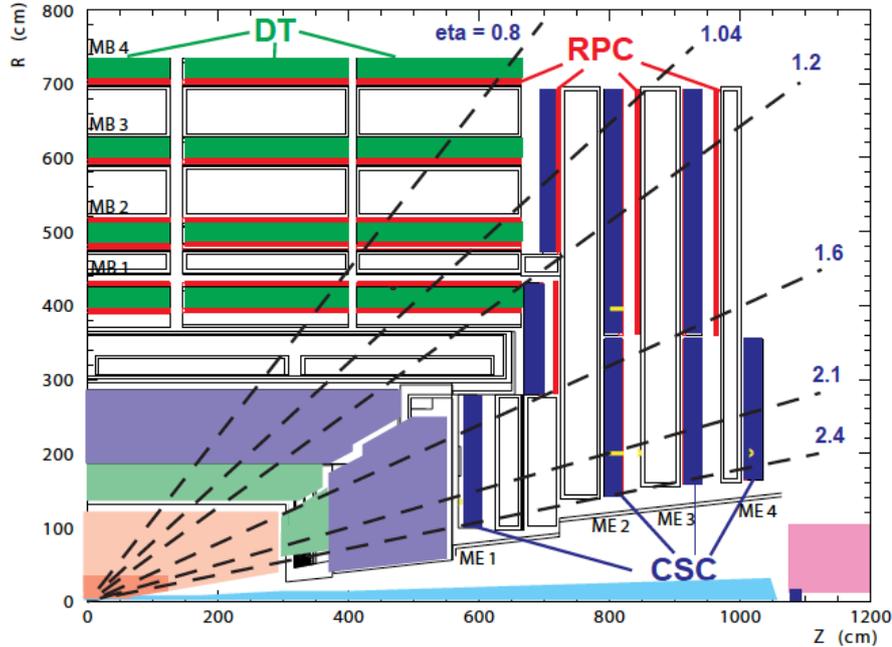


Figure 2.6: Layout of the muon system with the position of DTs, CSCs and RPCs.

Barrel detector

The barrel detector consists of 250 chambers, placed in 4 concentric stations (MB1, MB2, MB3 and MB4 starting from the innermost one) placed at $r = 4.0$ m, 4.9 m, 5.9 m, 7.0 m respectively. Each station is divided in 5 wheels along the z -axis and 12 sectors in the azimuthal angle. MB1 and MB2 are made by "sandwiching" one DT chamber in between two RPCs, while in the two outermost stations, MB3 and MB4, one DT is coupled with 1, 2 or 4 RPC layer, depending on the position.

A single DT chamber layout is different for MB1-MB3 stations and MB4. For the formers, it's made of 12 layers of drift tubes, grouped 4 by 4 forming 3 Super Layers (SL): 2 of them measures the $r - \phi$ coordinate while the remaining one measures the z coordinate. A honeycomb structure separates the first SL from the last two, thus giving a larger lever arm for the measurement in the bending plane. For the last station (MB4), DT chambers consist of only 2 SLs measuring $r - \phi$ coordinate.

The Barrel detector hosts 480 RPCs. Each RPC is a double-gap bakelite chamber operating in avalanche mode, with gap width of 2 mm. The strips,

running along the beam direction, are segmented, in order to match trigger requirements, in 2 parts for MB1, 3 and 4 (130 cm length) in 3 parts for the innermost and outermost layer in MB2 (85 cm length).

Endcap detector

Each Endcap detector is composed of 4 stations labeled ME1 to ME4 starting from the closest station with respect to the interaction point. They are mounted in disks perpendicular to the beam direction, surrounding CMS magnet, and every disk is divided in 2 or 3 concentric rings.

There are 468 CSCs in total, each one arranged in a trapezoidal shape and made of 6 gas gaps (7 layers) with planes of cathode strips in the radial directions and anode wires almost perpendicular to the strips. Most CSCs are overlapped in ϕ in order to avoid gaps in acceptance. There are 36 chambers for every ring, except the innermost ring of ME2,3 and 4 which has 18 chambers. The ionization of a charge particle passing through the planes cause charge to form on the anode wire and image charge on the cathode strips, thus allowing to get (r, z, ϕ) hits in each layer.

As it is shown in figure 2.6, in the endcap region, like in the barrel, there are layers of double-gap RPCs.

2.2.5 Electromagnetic calorimeter

The Electromagnetic Calorimeter (ECAL) is made of lead tungstate ($PbWO_4$) crystals, readout by silicon avalanche photodiodes (APDs) in the barrel and vacuum phototriodes (VPTs) in the endcap. Lead tungstate crystals provide short radiation length ($X_0 = 0.89$ cm) and Moliere length (2.2 cm), are fast because most of the light is produced in about 25 ns and are radiation hard. In order to cope with the non excellent light yield, photodetectors with intrinsic high gain are used. APDs and VPTs can also be use in strong magnetic field.

In figure 2.7 a transverse section of ECAL is shown.

Electromagnetic barrel

The barrel section of ECAL (EB) is made of 61200 crystals organized in 36 "supermodules", covering a pseudo-rapidity region of $0 < |\eta| < 1.479$. Each crystals has a front area of 22×22 mm² and a length of 230 mm (about $26 X_0$). They are quasi-projective, that is they have only a small axis tilt with respect to the nominal vertex position and cover 0.0174 in $\Delta\phi$ and $\Delta\eta$.

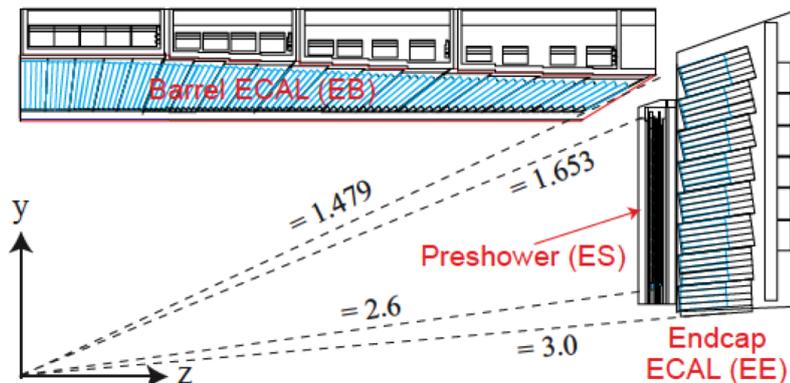


Figure 2.7: Scheme of the transverse section of the ECAL.

Electromagnetic endcap

The endcap section (EE) covers the pseudo-rapidity region $1.479 < |\eta| < 3.0$. The endcap crystals (7324 for each endcap) have a $28.6 \times 28.6 \text{ mm}^2$ cross section and a length of 220 mm (about $25 X_0$). They are arranged in a $x - y$ grid, unlike the barrel crystals which lie in a $\eta - \phi$ grid.

2.2.6 Hadron Calorimeter

The Hadron Calorimeter (HCAL) completes the calorimetry system with the measurements of jets and missing transverse energy. Its design is strongly influenced by the presence of the magnetic field because most of the HCAL is placed inside the coil, surrounding the ECAL. The main goals in the realization were to obtain good hermeticity for the measurements of the missing transverse energy (E_T^{miss}) and to minimize the non-Gaussian tails in the energy resolution, leading to the choice of maximize the budget material (i.e. interaction lengths). Brass (70% copper and 30% zinc) was chosen as absorber because of its non magnetic behavior and for its quite short interaction length ($\lambda_I \sim 151 \text{ mm}$). The active part is made of scintillator tiles coupled with wavelength shifting (WLS) fibres and clear fibres carrying the light to the readout system. The HCAL can be divided in the barrel HB ($|\eta| < 1.4$) and endcap HE ($1.3 < |\eta| < 3.0$), outer HO ($|\eta| < 1.26$) and forward HF ($3.0 < |\eta| < 5.0$).

Hadron barrel

The HB inner radius is 177.7 cm, the outer one is 287.65 cm. It's made of two half barrels, each one divided in 18 20° - ϕ wedges. Each wedge is

made of 17 layers of active scintillator (3.7 mm thick) interspersed with brass plates (about $50\div 60$ mm depending on the radial position), while the innermost and the outermost layers are made of stainless steel, which ensure structural strength. The first active layer is placed directly behind the ECAL, in order to deal with low showering particles coming from the material between ECAL and HCAL. The $r - z$ scheme of the HB is shown in figure 2.8, together with the HE. Each tile has a size of $\Delta\eta \times \Delta\phi = 0.087$

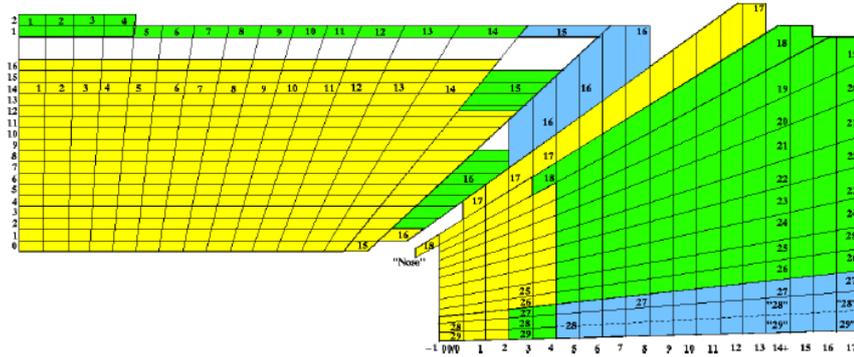


Figure 2.8: $r - z$ scheme of HB and HE.

$\times 0.087$ and is coupled with one WLS fibre. The overall HB consists then in 32 "towers" in η , readout by pixelated hybrid photodiodes (HPDs).

Hadron endcap

As it's clear from figure 2.8, HE partially overlaps HB. The ϕ -design of HE matches the HB, even though for $|\eta| > 1.74$ the ϕ granularity is halved to host the bending radius of the WLS fibres. The η size of the towers slightly increase for $|\eta| > 1.74$ up to $\Delta\eta=0.15$. The thickness of the absorber plates is 78 mm, thus reducing the sampling fraction.

Hadron outer

The HO is placed outside the HB, inside the barrel muon system. It is divided in 5 sections, or "rings", along η , each one covering 2.5 m length in z : the ring 0 has 2 scintillator layers on both side of an iron absorber, while the other rings have a single layer. The scintillator tiles are 10 mm thick and match the ϕ segmentation of DT chambers. The tower geometry of the scintillator in η and ϕ is the same of the HB. The main motivation for the HO is to measure the energy of penetrating hadron showers, in order to get the tails of the jets. Increasing the material budget up to about 11 interaction lengths in the mid-rapidity region improves also the resolution in E_T^{miss} .

Hadron forward

The forward hadron calorimeter HF is located at 11.2 m from the interaction point. Steel is the absorber material, while the active one is radiation hard quartz in fibres with varying length embedded to the steel, which provide fast Cherenkov light yield. The distance between 2 adjacent fibres is 5 mm and they are readout separately by phototubes in the rear of the detector. Each module of HF is made of 18 wedges, which are placed in a non-projective way, with fibres running along beam direction.

2.2.7 Inner tracking system

The inner tracking system of CMS detector has a "onion-like" layout, with 3 different choices depending on the distance from the interaction point. Closest to the interaction point ($r \leq 10$ cm), where particle density is the highest, pixel detector are placed. In the region $20 \text{ cm} < r < 55$ cm the particle flux decreases enough to allow the use of silicon microstrips, while in the outermost region of the tracker ($r > 55$ cm) larger pitch microstrips are placed. In figure 2.9 a 1/4 longitudinal view of the tracker system is shown. The inner tracking system extends to 110 cm in r and 540 cm in z .

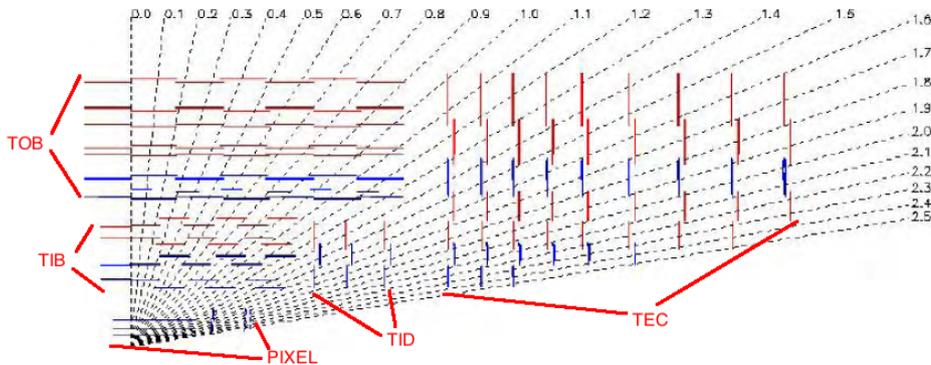


Figure 2.9: 1/4 of the tracker from a $r - z$ point of view.

Strip tracker

The barrel region of the strip tracker is divided in Tracker Inner Barrel (TIB) and Tracker Outer Barrel (TOB). The TIB consists of 4 layers covering the region $|z| < 65$ cm, while the TOB is made of 6 layers covering $|z| < 110$ cm. In the TIB, silicon strips with a thickness of $320 \mu\text{m}$ are used, with a strip pitch varying from $80 \mu\text{m}$ to $120 \mu\text{m}$. The first two layers are arranged in a "stereo" mode, allowing the measurements of both $r - z$ and

$r - \phi$ coordinates, with a stereo angle of 100 mrad. In the TOB, the radiation levels are smaller, thus allowing to use thicker strips ($500 \mu\text{m}$) with larger strip pitch ($120\mu\text{m} - 180 \mu\text{m}$). The first two layers of the TOB are made of "stereo" modules too, with the same stereo angle used for the TIB. In the TIB the single point resolution varies between $23\mu\text{m} - 34 \mu\text{m}$ in $r - \phi$ and is $230 \mu\text{m}$ in z , while in the TOB is between $35\mu\text{m} - 52 \mu\text{m}$ in $r - \phi$ and is $530 \mu\text{m}$ in z .

The endcaps are divided into the TEC (Tracker End Cap) and TID (Tracker Inner Disks) on both sides of the barrel. Each TEC is made of 9 disks in the region $120 \text{ cm} < |z| < 280 \text{ cm}$. The TID comprises 3 disks placed in the gap between the TIB and the TEC (see figure 2.9). Both the TID and the TEC are arranged in rings, with silicon strips that points towards the beam line, with pitch varying from ring to ring. The thickness of the sensors is $320 \mu\text{m}$ for all the modules of the TID and for the 3 innermost rings of the TEC, while the rest of the TEC has $500 \mu\text{m}$ thick sensors. The first 2 rings of the TID and the innermost 2 rings and the fifth ring of the TEC have "stereo" modules.

The strip modules are mounted on carbon-fibre structures with special tubes to control the temperature, which is designed to be about $-20 \text{ }^\circ\text{C}$.

Pixel tracker

The pixel detector consists of 3 barrel layers at $r = 4.4 \text{ cm}$, 7.3 cm , 10.2 cm and 53 cm long in z and 2 endcap disks at $|z| = 34.5 \text{ cm}$ and 46.5 cm , with $6 \text{ cm} < r < 15 \text{ cm}$. High granularity is mandatory to get good vertex resolution, so $100 \times 150 \mu\text{m}^2$ pixel area is set either along z and in $r - \phi$ plane. The whole inner tracker hosts about 66 million pixels. The spatial resolution of pixel detector is approximately $10 \mu\text{m}$ and $20 \mu\text{m}$ in $r - \phi$ and z coordinate respectively. In figure 2.10 a 3D view of the pixel detector layout is displayed.

2.2.8 Trigger and data acquisition

One of the biggest challenge in the design of CMS was the trigger and data acquisition system: some 10^9 interactions/sec happen at the LHC with the design instantaneous luminosity in proton-proton collisions, but only a small part of them can be written to archival media. A rejection factor of approximately 10^6 is needed.

The trigger system of CMS is divided in Level 1 (L1) trigger and High Level Trigger (HLT): the first is mainly hardware-based the second one is software based. The overall trigger architecture comprises detector electronics, processors for L1 and HLT running and the readout network.

Performances of the trigger system from the early collisions can be found in section 2.3.

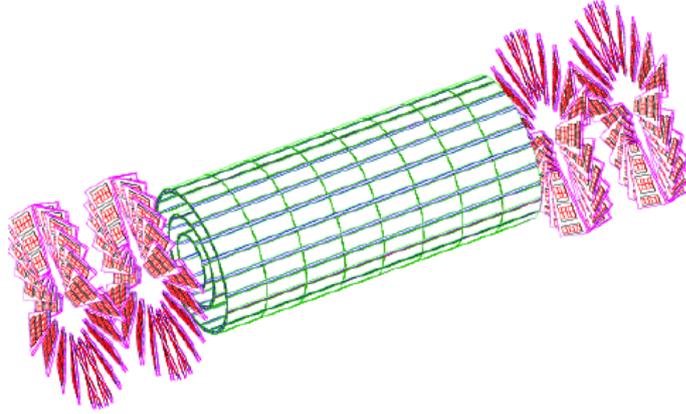


Figure 2.10: The pixel detector of the inner tracker system.

Level 1 trigger

A minimum time is needed for the transit of data from front-end electronics to Level 1 trigger processors and back: the total time allocated for this transit and the decision on whether to keep or discard the event is $3.2 \mu\text{s}$. During this time, the detector data must be held in special buffers (pipelined memories) while trigger data is collected from the front-end electronics. Less than $1 \mu\text{s}$ is used for L1 calculations.

The L1 decision involves only calorimeters and muon informations and uses objects called "trigger primitives", such as muons, photons, jets which are above certain E_T or p_T thresholds. To build this objects, larger granularities and resolutions are used.

The design rate of the L1 trigger system is 100 kHz, set by the average time needed for the full transfer of detectors data to the readout system. Then other processors make subsequent decisions based on full detector informations and more sophisticated algorithms, approaching the complexity of final reconstruction.

High level triggers

After $3.2 \mu\text{s}$, if the event passes the L1 trigger filter, detector data are transferred from the pipelines to the front-end readout buffers. Here further signal processing, zero suppression and data compression are performed. The total size of an event at this point is 1.5 MB for p-p collisions. Each event, contained in several hundreds different buffers, is transferred to a processor, running a HLT software. The HLT further filters the events, reducing the rate from 100 kHz to 100 Hz. Based on the concept of discarding the event as soon as possible, the HLT runs the reconstruction procedure

only of those physics objects that are really needed at that moment. For this reason, many different virtual trigger levels, with increasing complexity of reconstruction, are set.

2.3 CMS performances from early data

The earlier data collected with the CMS detector were extensively used to test the detector response, calibrate it and to study its performances in some crucial measurements. In general, they have been found to be in good agreement with design expectations and well reproduced by simulation. In this section some of these results are summarized, focusing on physics object reconstruction performances of interest for the $Z + b$ analysis.

2.3.1 Muon identification and reconstruction

Muon reconstruction in CMS is performed with different algorithms:

- *Global muon reconstruction*: starting from a segment in the muon chambers, a matching tracker track is found and a global track is fitted combining the hits.
- *Tracker muon reconstruction*: each track in the tracker with $p_T > 0.5$ GeV is extrapolated to the muon system and, if at least one matching segment is found, the muon is reconstructed as tracker muon.
- *Standalone muon*: only the track in the muon chambers is reconstructed (about 1% of muons, thanks to the high tracker efficiency).

The performance of muon identification in CMS has been studied on a sample of muons corresponding to an integrated luminosity of up to 84 nb^{-1} (see [5]). The study reported here has been done with three different muon selections:

- *Soft muons*: tracker muons with one matched segment in the outermost muon station.
- *Global muons*: muons reconstructed as global muons.
- *Tight muons*: global muons with $p_T > 3$ GeV and additional requirements on the quality of the track and the impact parameter.

Reconstruction efficiency

In figure 2.11 the muon reconstruction efficiency for low p_T muons is shown as a function of p_T for the three different muon selection. The efficiency is estimated with the "tag-and-probe" method on the J/ψ resonance,

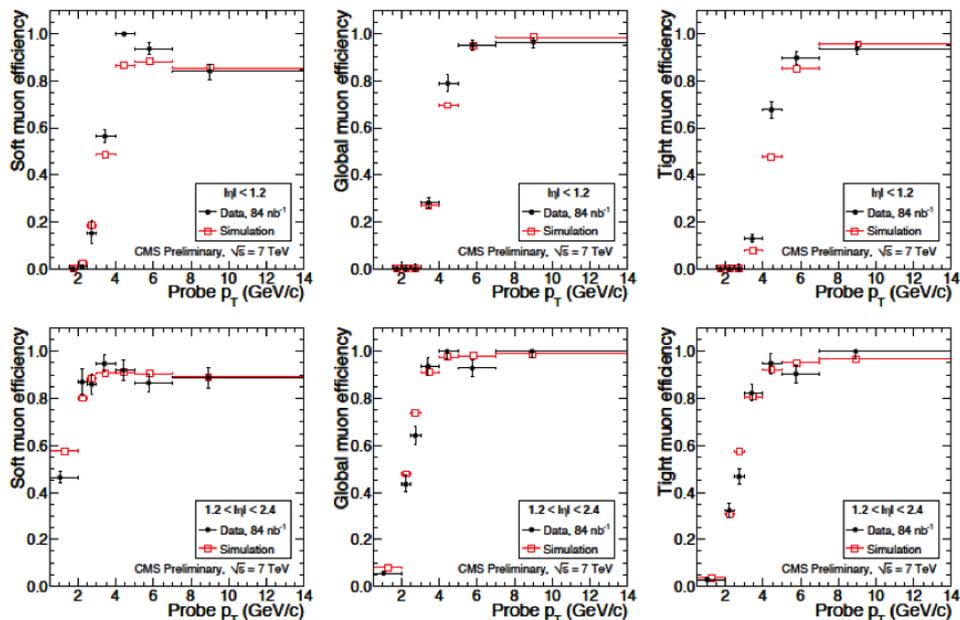


Figure 2.11: Results for the muon reconstruction efficiency in data compared to simulation [5]. The plots show the efficiency as a function of muon p_T for soft muons (left), global muons (middle) and tight muons (right) in the barrel (top) and endcaps (bottom).

which is described elsewhere [5]. The results in simulation/data agree rather well, within 5-10 %. As expected the soft muon selection has an higher efficiency at low momentum with respect to the the others.

L1 and HLT trigger efficiency

HLT, described above, follows an "outside-in" logic. First a L1 object is used to reconstruct a standalone muon (L2 muon), then seeds around the L2 muon are used to reconstruct a tracker muon and if it matches with the L2 muon a global fit is performed (L3 muon). On L2 or L3 muon the p_T threshold is applied.

In figure 2.12 we report the results for the region $|\eta| < 2.1$ and L2 and L3 thresholds at $p_T > 3$ GeV, using again the "tag-and-probe" method on the J/ψ . The shapes of the curves in figure 2.12 agree quite well between data/simulation, except for the plateau value, that is lower than expected in data and this is due to time synchronization of the muon detector at start-up.

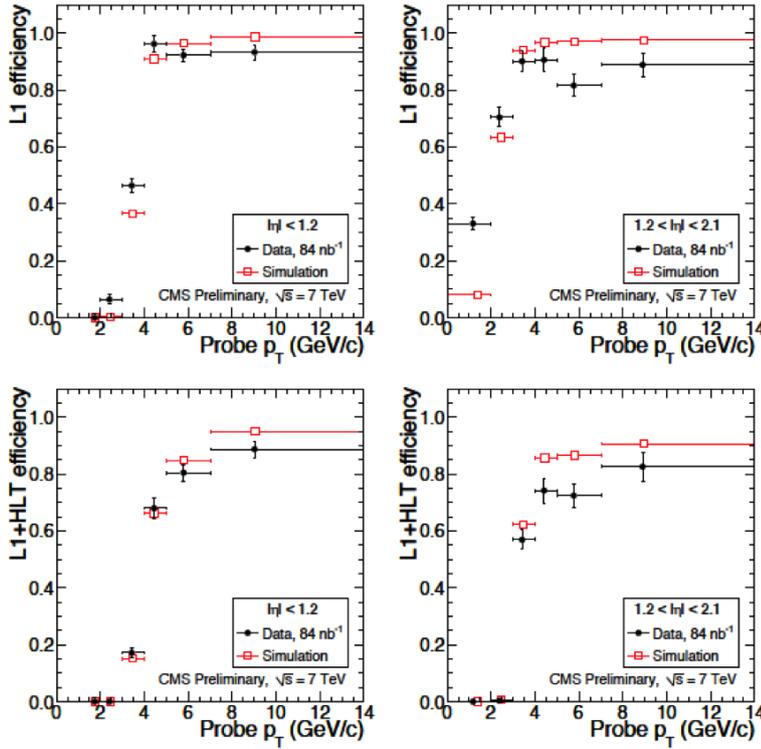


Figure 2.12: Muon trigger efficiency for data compared to simulation: the absolute trigger efficiency as a function of p_T , for L1 (top) and the combined efficiency of L1 and HLT with a threshold at 3 GeV (bottom) in the barrel (left) and endcap (right) [5].

2.3.2 Electron identification and reconstruction

Electron reconstruction uses two different algorithms at the track-seeding stage: a "tracker-driven" seeding, more suitable for low p_T electrons and the "ECAL-driven" seeding, optimized for p_T regions relevant in W and Z decays. The latter starts from the reconstruction of "superclusters", group of clusters of energy deposits in the ECAL, which are first matched with track seeds (2 or 3 hits) in the inner tracker. An electron track is first built from track seeds, while trajectories are reconstructed using a model of energy loss fitted with a Gaussian Sum Filter (GSF electrons). As the final preselection step, a matching in $\eta - \phi$ is done between GSF tracks and ECAL superclusters. For the "tracker-driven" seeding, the preselection is based on a multivariate analysis.

Studies about reconstruction and identification of electrons were performed using the first LHC collisions at $\sqrt{s} = 7$ TeV, with an approximated integrated luminosity of 200 pb^{-1} [6].

L1 and HLT trigger efficiency

The L1 EG (e/γ) trigger is based on electron and photon candidates using energy deposits called trigger primitives as inputs. The transverse energy of each ECAL trigger tower is computed in 250 MeV steps, then adjacent trigger towers are summed to form pairs using a 3×3 sliding window around the trigger primitive with maximum energy. The total transverse energy of a pair is used to compute L1 trigger energy values. An isolation criterium, based on the amount of energy deposited in the towers around the central one, is used to separate candidates into isolated and non-isolated. Only the four most energetic isolated and non-isolated L1 candidates are sent to the global trigger which generates the final decision. The L1 trigger requires a transverse energy above a configurable cut: L1 SingleEG5, the one used in this analysis, is requiring $E_T > 5$ GeV. The efficiency is computed with respect to reconstructed electrons in minimum bias data sample enriched with electrons from conversions.

Figure 2.13 shows L1 EG5 trigger efficiency as a function of the transverse energy of the electron in the offline reconstruction.

At HLT level, electron and photon selection proceeds requiring a super-

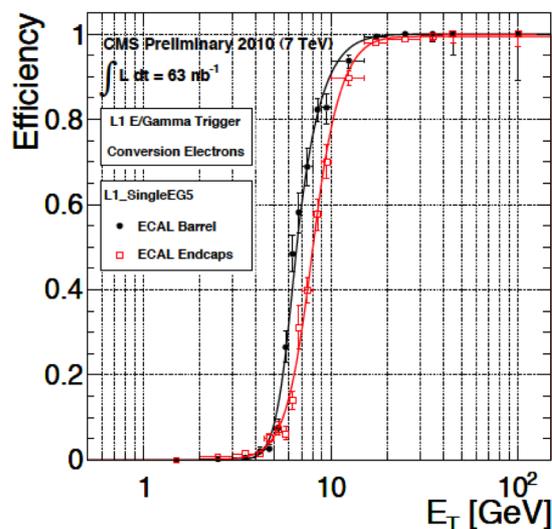


Figure 2.13: L1 EG5 trigger efficiency for electron candidates from minimum bias data sample as a function of the electron transverse energy [6].

cluster with E_T above a given threshold matching an electromagnetic L1 candidate. The HLT runs the standard ECAL superclustering algorithm with almost identical settings to the offline reconstruction. The electron paths, with respect to the photon paths, additionally require a hit in the

pixel layers of the CMS detector compatible with an electron trajectory, with matching requirements currently looser than those of the offline reconstruction. The HLT trigger path used in this analysis is HLT Ele15 LW, requiring a matching with a L1 candidate with $E_T > 5$ GeV and the online supercluster with $E_T > 15$ GeV.

The HLT Ele15LW efficiency for an offline reconstructed electron which passes HLT Photon15, as a function of the supercluster E_T , is shown in figure 2.14. The dataset is the same used for the L1 trigger efficiency analysis described above.

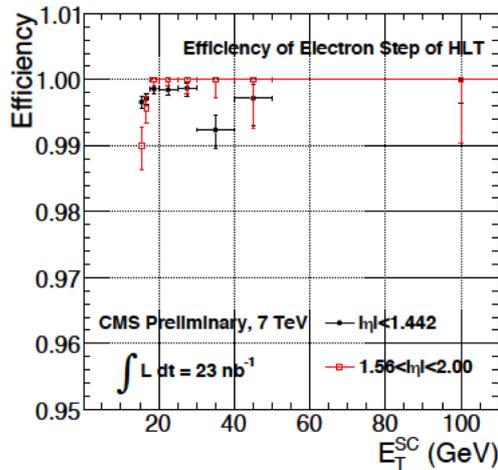


Figure 2.14: HLT Ele15LW trigger efficiency as a function of E_T [6].

Reconstruction efficiency

The reconstruction efficiency for electrons is computed using the "tag-and-probe" method for Z decays, described in [6]. Selected electrons pass standard cuts applied for the $Z \rightarrow ee$ selection (see section 4.2.1). Approximately 70 $Z \rightarrow ee$ events were used: an efficiency of $99.3\% \pm 1.4\%$ is obtained for electrons in the ECAL barrel and $96.8\% \pm 3.4\%$ in the end-cap, in good agreement with the expected efficiency from the Monte Carlo simulation of 98.5% and 96.1% respectively. In figure 2.15 the comparison between data and MC estimate of the reconstruction efficiency is shown using a complementary method based on a maximum likelihood (ML) fit on W events, described in [6].

Selection efficiency

Electron selection is based on three types of variables: identification (ID), isolation and conversion rejection variables. A simple cut based selec-

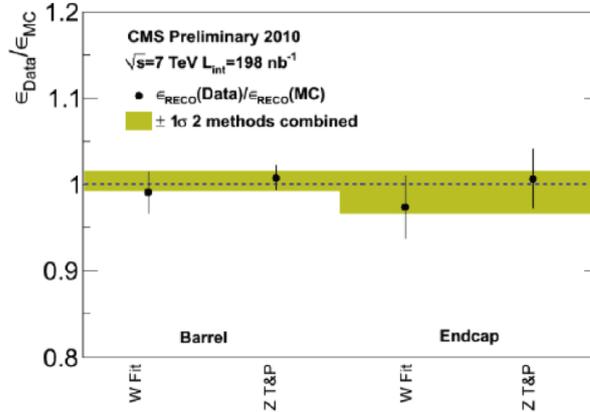


Figure 2.15: Electrons reconstruction efficiency estimated with two different methods and compared with Monte Carlo predictions [6].

tion method has been developed, with Monte Carlo studies, which is based on the combined cuts on these variables. Several references are available for prompt electrons with $p_T > 20 \text{ GeV}$, having nominal values of efficiency: here working points (WP) of 95% and 80% are used for the analysis of the selection efficiency. Three methods are used to estimate the efficiency: the Z "tag-and-probe" method and the ML fit on W events used in the reconstruction efficiency analysis described above and a W "tag-and-probe" method, which is also described in [6].

The efficiencies computed with the $Z \rightarrow ee$ events, together with MC predictions, are listed in table 2.3. In figure 2.16 data/MC comparison for selection efficiencies are shown for the different methods and working points.

Table 2.3: Electron selection efficiency values obtained with the "tag-and-probe" method on Z decays.

	ECAL barrel		
Selection	Efficiency data	Error (stat.+syst.)	Efficiency MC
WP 95 %	92.5 %	3.2 %	95.4 %
WP 80 %	77.5 %	4.7 %	85.1 %
	ECAL endcap		
Selection	Efficiency data	Error (stat.+syst.)	Efficiency MC
WP 95 %	86.4 %	6.7 %	92.9 %
WP 80 %	75.1 %	8.6 %	76.2 %

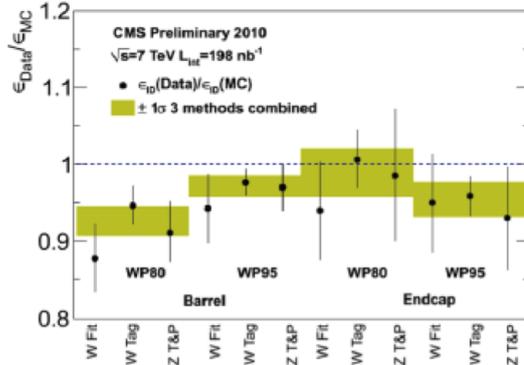


Figure 2.16: Electron selection efficiency estimated with three different methods and two different working points [6].

2.3.3 Jet performances

Four type of jets are reconstructed in CMS detector, based on the anti- k_T clustering algorithm (see Appendix D), but here only three of them are considered, which are listed in order of increasing complexity:

- *Calorimeter jets* are reconstructed using energy deposits in the electromagnetic (ECAL) and hadronic (HCAL) calorimeter.
- *The Jet-Plus-Tracks (JPT)* algorithm associate tracks coming from charged particles to the previously reconstructed calorimeter jets, based on the $\eta - \phi$ separation from the jet axis. Then the algorithm corrects energy and direction of the jets using the informations of the associated tracks.
- *The Particle Flow (PF)* algorithm takes advantage of the list of particles that are reconstructed in the event and use them to reconstruct the jet: it's expected to sensibly improve p_T and spatial resolution.

Jet performances have been studied by the CMS Collaboration from the data based on up to 73 nb $^{-1}$ of proton-proton collisions at $\sqrt{s} = 7$ TeV (see [7] for more details).

The selection of jets in this analysis is based on standard jet identification criteria, allowing rejection of fakes.

p_T resolution

The asymmetry method, described in [7], is used to estimate the p_T resolution from data: in figure 2.17 the results are shown for the three type of jets described above, with data/QCD dijet simulation comparison. Data

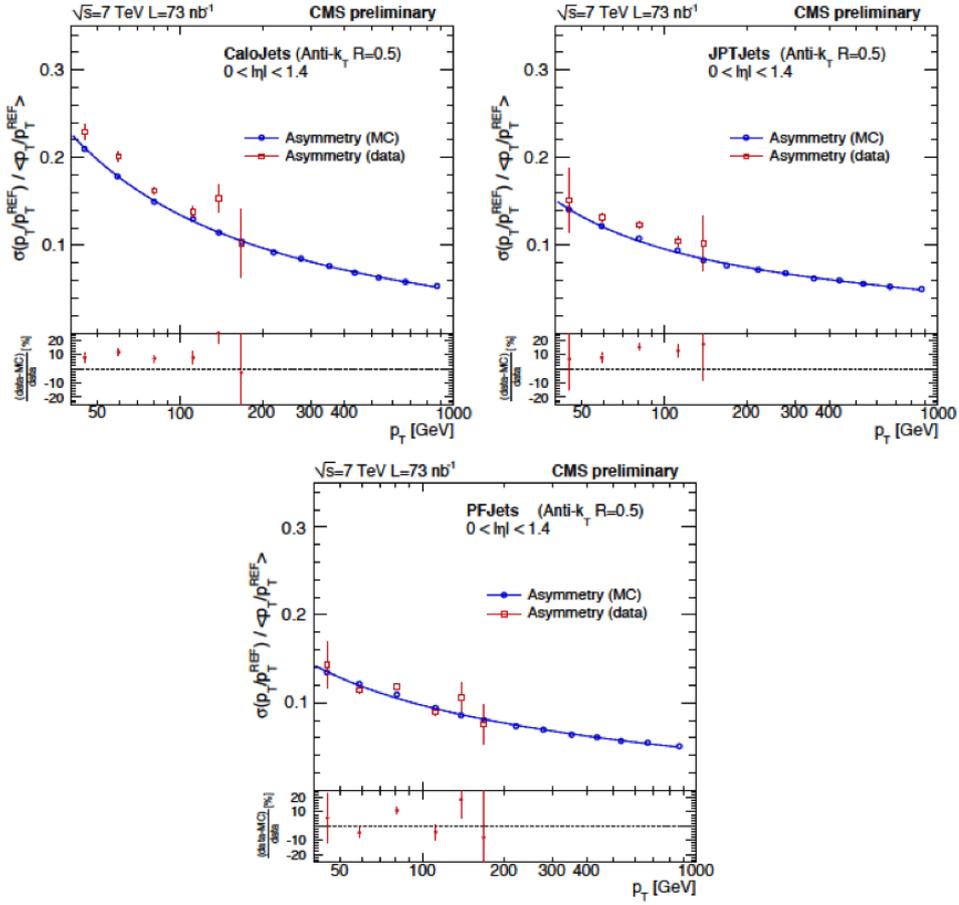


Figure 2.17: Calorimeter (top left), JPT (top right) and PF (bottom) resolution on the p_T of the jet [7].

and simulation are in reasonable agreement, typically at 10% level. As expected, the resolutions for jets including also track informations (JPT and PF) are improved with respect to calorimeter jets.

Position resolution

In order to validate the simulation-based results on jet position resolution of calorimeter jets, dijet data were selected to study $\Delta\eta$ and $\Delta\phi$ between matched calorimeter jets and PF jets, under the assumption of a smaller resolutions of PF jets. Looking at figure 2.18 we see again a good agreement between data and Monte Carlo, indicating that simulation in CMS gives a reliable jet position resolution.

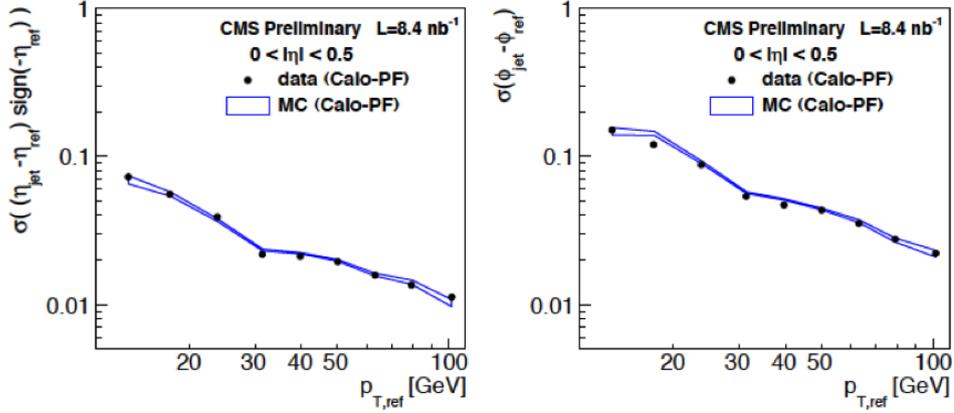


Figure 2.18: $\sigma(\Delta\eta)$ (left) and $\sigma(\Delta\phi)$ (right) in data and Monte Carlo of calorimeter jets with respect to PF jets [7].

2.3.4 B-tagging performances

B-tagging performances have been studied using data collected in p-p collisions at $\sqrt{s} = 7$ TeV, for an integrated luminosity of 0.50 to 0.89 fb^{-1} and they have been compared to the simulation [9]. In the following, different methods to estimate the b-tagging efficiency and the mis-identification efficiency of light-flavor jets (mistag rate) are compared, with focus on the track counting and simple secondary vertex algorithms, described in Appendix B.

B-tagging efficiency

B-tagging efficiency is defined as:

$$\epsilon_b^{tag} = \frac{N_{b-tag}}{N_b} \quad (2.1)$$

where N_b is the number of selected b-jets and N_{b-tag} is the number of selected b-jets fulfilling the b-tagging algorithm selections.

Semileptonic decays of B-hadrons can lead to a jet containing a muon: thanks to the high muon reconstruction efficiency in CMS, this feature can be used to enhanced the purity of the sample used for efficiency studies. Thus, a sample contained a jet with a muon associated to it (a "muon-jet") is used for the analysis presented here.

Muons coming from b quark semileptonic decays have larger momentum transverse to the jet axis, the so-called p_T^{rel} , with respect to light flavors decays, due to the larger b-quark mass. Thus p_T^{rel} can be used as a discriminant variable in the efficiency estimate. Two methods have been compared:

- The "PtRel" method relies on the fit to the p_T^{rel} distribution in the data using simulated spectra from signal (b) and background (charm+light flavors).
- The "System8" method is based on the solution of an 8-equation system involving different tagging criteria applied to the data.

More details of these methods can be found in [9].

In figure 2.19 fits to the muon p_T^{rel} distributions are shown for two different algorithms.

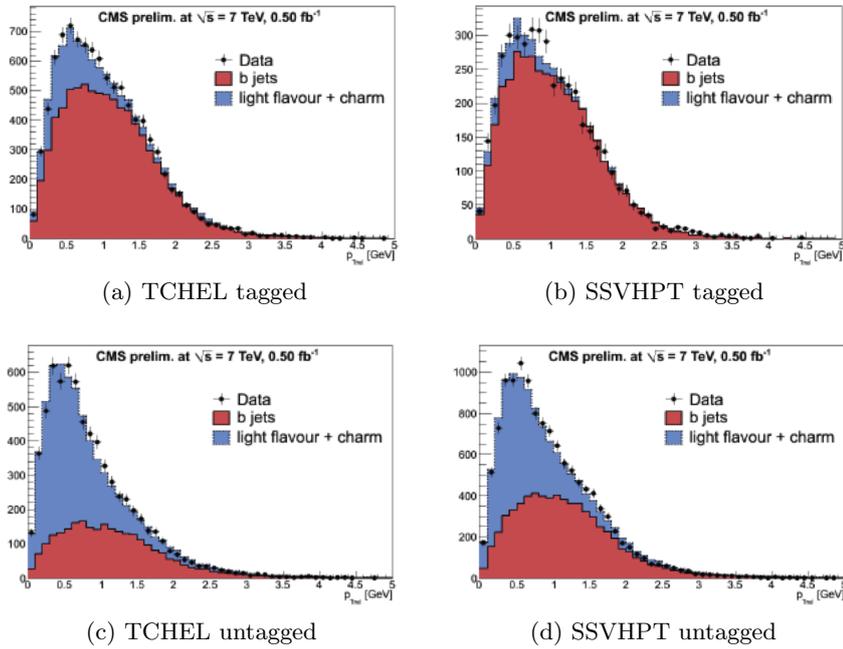


Figure 2.19: Fits of the muon p_T^{rel} distributions for muon jets that pass (a, b) or fail (c, d) the b-tagging criteria for Track Counting High Efficiency Loose (TCHEL) and Simple Secondary Vertex High Purity Tight (SSVHPT) algorithms (the muon-jet p_T is between 50 and 80 GeV) [9].

In table 2.4 results are shown for the two different methods applied to the two taggers. They are in good agreement within the statistical uncertainties.

Mistag rate

The mistag rate is defined as:

$$\epsilon_b^{mistag} = \frac{N_{l-tag}}{N_l} \quad (2.2)$$

Table 2.4: B-tagging efficiency estimated with the PtRel and Sistem8 method, for muon-jets p_T between 50 GeV and 80 GeV [9].

b-tagger	ϵ_b^{tag} (PtRel)	ϵ_b^{tag} (System8)
TCHEL	0.76 ± 0.01	0.77 ± 0.01
TCHEM	0.63 ± 0.01	0.63 ± 0.02
SSVHEM	0.62 ± 0.01	0.60 ± 0.01
TCHPM	0.48 ± 0.01	0.49 ± 0.01
SSVHPT	0.38 ± 0.01	0.37 ± 0.01
TCHPT	0.36 ± 0.01	0.37 ± 0.01

where N_l is the number of selected light flavor jets and N_{l-tag} is the number of selected light flavor jets fulfilling the b-tagging algorithm selections.

The measurement of the mis-identification rate from light flavor jets relies on the definition of negative discriminator values for each b-tagging algorithm. Impact parameters can be signed as positive (negative) if the associated tracks are produced downstream (upstream) with respect to the primary interaction vertex. The reconstructed decay length between the primary and secondary vertices can be signed in a similar way.

The mistag rate is evaluated from tracks with a negative impact parameter or from secondary vertices with a negative decay length. When applied to jets of any flavor but using only negative discriminator values, the corresponding tagging efficiency is denoted negative tag rate.

In figure 2.20 spectra of TCHE and SSVHP discriminator from a an inclusive jet sample are shown for positive and negative tags. The data/MC comparison shows a good agreement, most of the time within about 20 %.

The negative tagged jets are enriched in light flavors, so the mistag rate can be evaluated from data using Monte Carlo simulations to get a correction factor. The mistag rate as it is measured from data reads:

$$\epsilon_{DATA}^{mistag} = \epsilon_{DATA}^- R_{light} \quad (2.3)$$

where ϵ_{DATA}^- is the negative tag rate in the data and R_{light} is a correction factor, which is calculated from:

$$R_{light} = \frac{\epsilon_{MC}^{mistag}}{\epsilon_{MC}^-} \quad (2.4)$$

where ϵ_{MC}^{mistag} and ϵ_{MC}^- have the same definition as the ones from the data. In table 2.5 results of mistag rate estimates for different taggers and working points are listed, together with data/MC scaling factors.

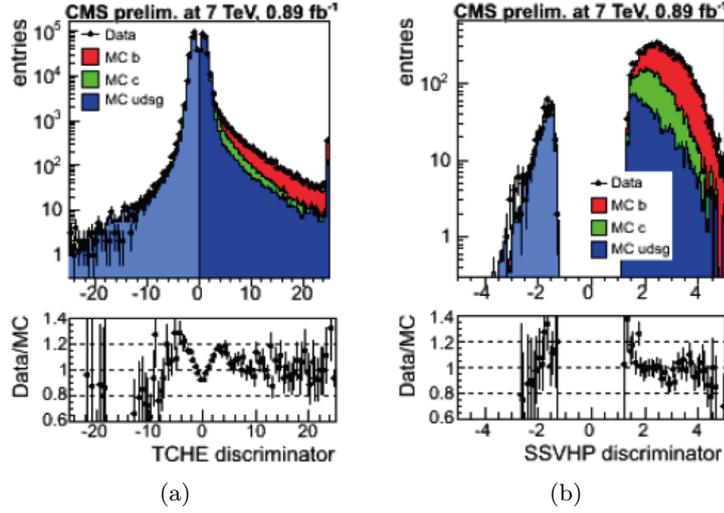


Figure 2.20: Signed b-tag discriminator for the TCHE algorithm (a) and the SSVHP algorithm (b). Both data and MC samples are required to satisfy a jet p_T threshold of 30 GeV [9].

Table 2.5: Mistag rate and data/MC scale factor for different b-tagger, with statistical and systematic uncertainties respectively quoted. Jets p_T is between 50 and 80 GeV [9].

b-tagger	ϵ_{DATA}^{mistag}	$\frac{\epsilon_{DATA}^{mistag}}{\epsilon_{MC}^{mistag}}$
TCHEL	$0.128 \pm 0.001 \pm 0.026$	$0.98 \pm 0.01 \pm 0.11$
TCHEM	$0.0175 \pm 0.0003 \pm 0.0038$	$1.21 \pm 0.02 \pm 0.17$
TCHPM	$0.0177 \pm 0.0002 \pm 0.0036$	$1.27 \pm 0.02 \pm 0.15$
SSVHEM	$0.0144 \pm 0.0003 \pm 0.0029$	$0.91 \pm 0.02 \pm 0.10$
SSVHPT	$0.0012 \pm 0.0001 \pm 0.0002$	$0.93 \pm 0.09 \pm 0.12$
TCHPT	$0.0017 \pm 0.0001 \pm 0.0004$	$1.21 \pm 0.10 \pm 0.18$

Chapter 3

Production of Z in association with one or more bottom-quark jets

In proton-proton collisions a Z boson can be produced in association with one or more b -tagged jets, in a process that in the following will be called $Z+b$. This process is a background for SM Higgs search as well as for extensions of the SM Higgs sector (like the minimal supersymmetric extension or MSSM and the minimal non-supersymmetric two Higgs doublets model or M2HDM). In the following an overview of the theoretical motivations for the measurement of the $Z+b$ cross section is presented, focusing on the channels of new physics. In section 3.2 the most recent calculations of $Z+b$ cross section are presented.

3.1 Theoretical motivations for $Z+b$ measurement

The $Z+b$ cross sections measurement is important either to confirm SM predictions either because its signature is a background for new physics channels, especially for Higgs physics.

These channels of new physics are briefly summarized in the following paragraphs, focusing on the golden channel which is the subject of this thesis.

3.1.1 SM Higgs search

There are two SM Higgs channels which are mainly affected by $Z+b$ background: the ZH associated production and the "golden channel" (see section 1.2.3). As presented in section 1.2.2, the SM Higgs boson can be produced in association with a Z boson (Higgs-strahlung) in proton proton

collisions. If at the same time H decays to a $b\bar{b}$ pair, as it is the most probable case in the low mass range (see figure 1.6), the resulting semi-inclusive signature should be 2 leptons and 2 b-tagged jets, being the Z boson identified via its decay into a pair of leptons¹. This is the same signature of a $Z+b$ event, as it will be described in section 4.2, in the case when there are two b-tagged jets. Hence, $Z+b$ is a background for the SM Higgs search in the channel $Z(\rightarrow l^+l^-)H(\rightarrow b\bar{b})$. As far as the "golden channel" is concerned, $Z+b$ can lead to more-than-two leptons events. The bottom quark, infact, can decay semileptonically in 3 different ways:

- Direct decay :
 $b \rightarrow l$, BR \sim 10.7%
- Cascade decay :
 $b \rightarrow c \rightarrow l$, BR \sim 8%
- "Wrong sign" cascade decay :
 $b \rightarrow \bar{c} \rightarrow l$, BR \sim 1.6%

resulting in an inclusive BR into leptons of about 20%. Thus a $Zb\bar{b}$ final state can lead to a 3 or 4 leptons signature, being the Z identified by its decay into leptons.

As we saw in section 1.2.3, the SM Higgs boson can decay into a pair of Z bosons, and one of the most promising signature for the discovery of the Higgs is the inclusive search for a 4 leptons final state (sometimes referred to as the "golden channel"). That's why $Z+b$ is one of the major background in the golden channel together with the continuum non-resonant $ZZ^{(*)}$ production. This topic will be further investigated in chapter 4.

3.1.2 MSSM Higgs search

Higgs sector in the MSSM

Without any claim of completeness here we give some hints on the construction of the Higgs sector in the minimal supersymmetric extension of the Standard Model (MSSM).

In constructing the MSSM, both $Y = -1$ and $Y = +1$ Higgs complex doublets (and their superpartners) must be considered in order to get an anomaly-free theory. Two types of models can be defined corresponding to different scalar-fermion couplings:

- in the *Type I* model, all fermions couple to one doublet in a generic basis and the Yukawa couplings are rescaled to $\tan\beta \equiv v_2/v_1$, where $v_{1,2}$ are the vacuum expectation values of the two doublets;

¹From now on, with leptons we mean electrons and muons, and $l = e, \mu$

- in the *Type II* model, the up-fermions couple to one doublet with strength $\propto \cot\beta$ while the down-fermions couple to the other doublet with strength $\propto \tan\beta$.

In a Type II two Higgs doublet model the doublets are written as $\Phi_d = (\Phi_d^0, \Phi_d^-)$ ($Y = -1$) and $\Phi_u = (\Phi_u^+, \Phi_u^0)$ ($Y = +1$), where Φ_d couples only to down-type fermions, Φ_u only to up-type fermions. Once the Higgs potential is minimized, they get their respective vacuum expectation values (v.e.v.):

$$\langle \Phi_d \rangle = \begin{pmatrix} v_d \\ 0 \end{pmatrix}; \quad \langle \Phi_u \rangle = \begin{pmatrix} 0 \\ v_u \end{pmatrix} \quad (3.1)$$

where the v.e.v. satisfy: $v^2 = v_d^2 + v_u^2 = (246 \text{ GeV})^2$ and $\tan\beta \equiv v_u/v_d$. The spontaneous symmetry breaking mechanism results in three Goldstone bosons (giving longitudinal polarization to W^\pm and Z) and five physical Higgs particle: a charged pair H^\pm , one CP-odd neutral scalar A and two CP-even neutral scalars h and H ($m_h \leq m_H$). Introducing the mixing angle α the physical fields are:

$$\begin{cases} h = -(\sqrt{2} \text{Re}\Phi_d^0 - v_d) \sin\alpha + (\sqrt{2} \text{Re}\Phi_u^0 - v_u) \cos\alpha \\ H = (\sqrt{2} \text{Re}\Phi_d^0 - v_d) \cos\alpha + (\sqrt{2} \text{Re}\Phi_u^0 - v_u) \sin\alpha \end{cases} \quad (3.2)$$

Physical quantities in MSSM Higgs sector depend on two parameters, which may be taken as $\tan\beta$ and the mixing angle α or $\tan\beta$ and m_A . The MSSM Higgs couplings to SM gauge bosons and fermions typically depend on these two parameters. Focusing on Type-II Yukawa couplings of h with fermions, which are the most relevant for $Zb\bar{b}$ analysis as a background in MSSM Higgs search, the invariant Yukawa couplings are, using the 3rd family notation:

$$\mathcal{L}_{Yuk} = -Y_t [\bar{t}_R \Phi_u^0 t_L - \bar{t}_R \Phi_u^+ b_L] - Y_b [\bar{b}_R \Phi_d^0 b_L - \bar{b}_R \Phi_d^- t_L] + h.c. \quad (3.3)$$

from which a relation is obtained between Yukawa couplings and quark masses:

$$Y_b = \frac{\sqrt{2}m_b}{v_d} = \frac{\sqrt{2}m_b}{v \cos\beta}; \quad Y_t = \frac{\sqrt{2}m_t}{v_u} = \frac{\sqrt{2}m_t}{v \sin\beta} \quad (3.4)$$

The $hf\bar{f}$ couplings in units of m_f/v can be written as:

$$hb\bar{b}(h\mu^+\mu^-): \quad -\frac{\sin\alpha}{\cos\beta} = \sin(\beta - \alpha) - \tan\beta \cos(\beta - \alpha) \quad (3.5)$$

It can be demonstrated that, unlike the SM case, there is a tree level upper bound on m_h : $m_h \leq M_Z |\cos 2\beta| \leq M_Z$.

In the limit $m_A \gg M_Z$, the expressions of the masses of the Higgs particles simplify:

$$m_h^2 \simeq M_Z^2 \cos^2 2\beta; \quad m_H^2 \simeq m_A^2 + M_Z^2 \sin^2 2\beta; \quad m_{H^\pm}^2 = m_A^2 + M_W^2 \quad (3.6)$$

hence $m_H \simeq m_{H^\pm} \simeq m_A$. In this limit one can also get:

$$\cos^2(\beta - \alpha) \simeq \frac{M_Z^4 \sin^4 \beta}{4m_A^4} \quad (3.7)$$

which shows that $\cos(\beta - \alpha) \sim \mathcal{O}\left(\frac{M_Z^2}{m_A^2}\right)$. This is called the *decoupling limit* because it can be shown that when m_A is large, an effective low energy theory exists below the m_A scale in which the only effective particle in the Higgs sector is h . In particular, for $\cos(\beta - \alpha) = 0$ the tree-level couplings of h are exactly those of the SM Higgs boson.

There is a large region in MSSM parameter space where the decoupling limit applies, because $\cos(\beta - \alpha)$ approaches 0 quite rapidly when m_A is larger than about 200 GeV. As a result, in this region the search for the lightest CP-even Higgs boson of MSSM is equivalent to the search for the SM Higgs boson.

$\phi b\bar{b}$ production in the MSSM

For large values of $\tan\beta$, the production of a neutral Higgs boson in MSSM ($\phi = h/H/A$) at the LHC is dominated by the Higgs-radiation of bottom quarks, leading to the associated production $\phi b\bar{b}$ (see figure 3.1). As in the case of the calculation of $Z+b$ cross section (section 3.2), the

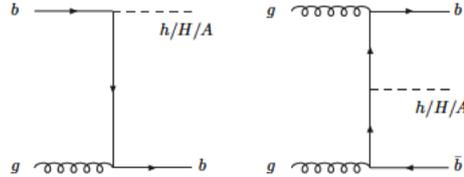


Figure 3.1: Feynman diagrams at LO for the production of $\phi = h/H/A$ in association with bottom quarks.

calculation of these processes can be done in a 4-flavor scheme or 5-flavor scheme: the two different approaches will converge at higher perturbative orders.

Among the channels for the discovery of the ϕ , the exclusive $pp \rightarrow \phi(\rightarrow \mu^+\mu^-)b\bar{b}$ channel is the most important in the $Z+b$ measurement, because the Z boson is reconstructed also with the decay in $\mu^+\mu^-$. Despite the small branching ratio, $BR(\phi \rightarrow \mu^+\mu^-) \simeq 10^{-4}$ for the relevant region in the parameter space, this channel provides a good reduction of Drell-Yan $\mu\mu$ background thanks to the b-tagging requirement.

3.1.3 M2HDM Higgs search

A minimal two Higgs doublet model

The minimal choice for the Higgs sector is adopted in the SM where a single $SU(2)_L$ scalar doublet is introduced, giving one physical particle, the Higgs boson and three Goldstone bosons which explain the masses of W^\pm and Z . Such a minimal solutions has many advantages and experimental evidences. It leaves the photon massless, it ensures that the one-loop corrections to the relation between M_W and M_Z introduce only logarithmic dependence on Higgs mass and exclude flavor changing neutral currents (FCNC) at tree level. Thus any extension of the SM scalar sector must introduce new symmetries to explain all this well-known phenomenology. The minimal realization of a two Higgs doublet model (M2HDM) starts from the definition of the scalar potential:

$$V(\phi_1, \phi_2) = -m_1^2 \phi_1^\dagger \phi_1 - m_2^2 \phi_2^\dagger \phi_2 + \frac{\lambda_S}{2} (\phi_1^\dagger \phi_1 + \phi_2^\dagger \phi_2)^2 + \frac{\lambda_{AS}}{2} (\phi_1^\dagger \phi_2 - \phi_2^\dagger \phi_1)^2 \quad (3.8)$$

where $\phi_{1,2}$ are the two doublets, $m_{1,2}$ and $\lambda_{S,AS}$ are 4 real parameters, against the 14 real parameters of the most general 2HDM: this is the reason of the appellative "minimal".

The lagrangian is invariant under CP , leading to CP -odd and a CP -even Higgs bosons (like in the MSSM) and it's invariant under a *custodial symmetry*, i.e. a symmetry that protects the tree level value of the ρ parameter:

$$\rho = \frac{M_W^2}{M_Z^2 \cos^2 \theta_w} = 1 \quad (3.9)$$

The stability of the ρ parameter depends, in general, on the scalar mass spectrum of the 2HDM: as it's clear from figure 3.2, M_W and M_Z receive corrections from the whole scalar sector. In the SM case, where only one

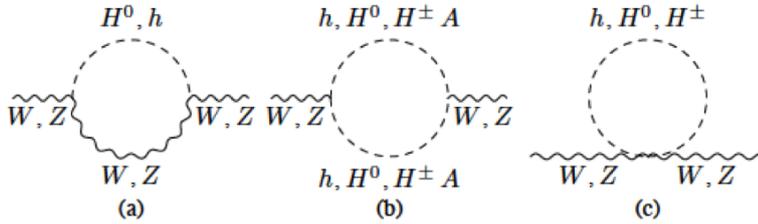


Figure 3.2: Relevant contributions to W and Z masses involving scalar loops.

doublet is present, the scalar potential has a $SO(4)$ global symmetry spontaneously broken into a custodial $SO(3)$. This global symmetry is apparent

at the level of the massless Goldstone bosons:

$$m_{G^\pm}^2 = m_{G^0}^2 = 0, \text{ with } G^0 \text{ CP - odd} \quad (3.10)$$

It can be demonstrated that the custodial symmetry can be extended to the case of two doublets if:

$$m_{H^\pm}^2 = m_A^2, \text{ with } A \text{ CP - odd} \quad (3.11)$$

This is the degeneracy foreseen in MSSM, for example. However, it has been demonstrated that the potential in equation (3.8) holds another $SO(3)$ mass degeneracy:

$$m_{H^\pm}^2 = m_{H^0}^2, \text{ with } H^0 \text{ CP - even} \quad (3.12)$$

which corresponds to the so-called "twisted" custodial symmetry. It can be shown that the potential (3.8) has also a \mathbb{Z}_2 symmetry which forbids FCNC. As in any two Higgs doublet model, also the M2HDM can be a Type I or a Type II model depending on the choice of the scalar-fermion couplings (see section 3.1.2).

Without entering into details of the couplings of the scalars and the constraints on the Higgs masses, we just state that in this model the mass of A , is no more forced to be degenerate with the charged Higgs, thus leaving the possibility to choose its mass. One possible way to do it, it's to focus on the region of the M2HDM parameter space not covered by the MSSM, i.e. imposing A to be the lightest Higgs ($m_A < 90$ GeV) and h to be the heaviest, while the triplet $T \equiv (H^\pm, H^0)$ lies in the middle. The choice $m_A < m_T < m_h$ is sometimes referred to as the "inverted" mass spectrum (i M2HDM) and leads to larger possibilities for scalar-to-scalar decays and, if they are kinematically allowed, the unusual decay modes $H^\pm \rightarrow W^\pm A$ or $H^0 \rightarrow ZA$ are the dominant ones.

Many more details about the M2HDM and its possible signatures at the LHC can be found in the documents [10] and [11].

ZA production in M2HDM

With the the inverted mass spectrum choice in the scalar sector, new possibilities for the decay modes of the Higgs boson can be considered and new experimental signatures become interesting. In table 3.1 the values of i M2HDM parameters corresponding to the benchmark points (BP) extracted from theoretical (direct and indirect) constraints are listed for the Type I and Type II models. For a Type II 2HDM, like MSSM, the cross section for the process $b\bar{b} \rightarrow H^0$ is enhanced as $\tan\beta$ increases. Thus a process like $b\bar{b} \rightarrow H^0 \rightarrow ZA$ (see figure 3.3) is a promising channel for the discovery of both H^0 and A . The final state in which the Z decays into a pair of leptons and A decays into a $\tau^+\tau^-$ pair is one of the most promising, especially when each τ decays into leptons of different flavor. This $l^+l^-e^\pm\mu^\mp + E_T^{miss}$

Table 3.1: Parameters corresponding to the benchmark points of Type I and Type II models together with some branching ratios relevant for the corresponding region of the $iM2HDM$.

Parameter	BP (Type I)	BP (Type II)
m_h	400 GeV	400 GeV
m_T	180 GeV	350 GeV
m_A	30 GeV	40 GeV
$\tan\beta$	0.2	30
Branching Ratio (%)	BP1	BP2
$A \rightarrow b\bar{b}$	86	90
$A \rightarrow \tau^+\tau^-$	10	10
$H^0 \rightarrow ZA$	~ 100	63
$H^\pm \rightarrow W^\pm A$	~ 100	79
$h \rightarrow H^+H^-$	20	-
$h \rightarrow H^0H^0$	10	-

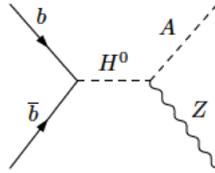


Figure 3.3: Feynman diagram for $b\bar{b} \rightarrow H^0 \rightarrow ZA$ process.

final state, in which l^+l^- come from the decay of the Z and $e^\pm\mu^\mp$ are close to each other due to the boost of A , is extremely clean and does not suffer from uncertainties related to the jet reconstruction.

However, we see from table 3.1 that the dominating decay mode of the light pseudo-scalar A is into a $b\bar{b}$ pair. This channel provides a larger cross section but more uncertainties coming from jets and b-tagging requirements. The case of a very light A results into a final state with two b-quarks very collinear: using a larger cone size they can be tagged as one b-jet. A way to reduce the background coming from Z and light-quark jets is to look for two b-tagged jets, i.e a $l^+l^-b\bar{b}$ final state, which is the same one of the $Z + b$ analysis (in the case of exclusive $Z + 2b$ final state), as it will be explained in the next sections. The probability of disentangling the two jets is shown in figure 3.4 as a function of m_A , for different values of m_{H^0} and for $\Delta R_{bb} = 0.5$, when the b-quarks are generated with $p_T^b > 30$ GeV and $|\eta^b| < 2.5$.

The ZZ background is the most problematic when $|m_Z - m_A|$ is small.

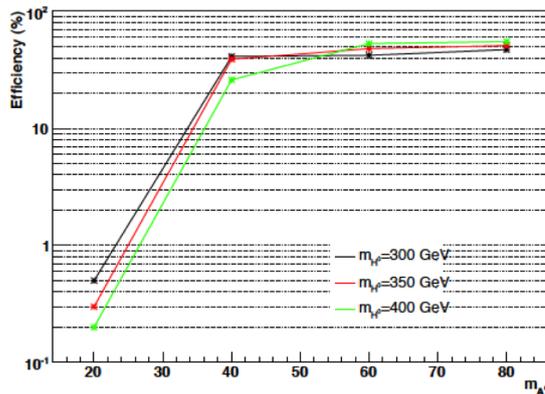


Figure 3.4: Efficiency in disentangling two b-jets in a ZA signal as a function of m_A , with $\Delta R_{bb} > 0.5$, $p_T^b > 30$ GeV and $|\eta^b| < 2.5$ (from [11]).

Possible discriminating variables are the boost of Z , the total invariant mass and the coplanarity $\Delta\phi$ between the 4-vector $l_1 + l_2$ and $j_1 + j_2$.

3.2 $Z+b$ cross section

A great theoretical effort was spent in order to get reliable cross section calculations for the production of a Z in association with one or more b-tagged jets (for an introduction on hadronic cross sections see Appendix A). With no claim of completeness, here we give some hints on them, but we address to [12] through [15] for the details.

In literature it's common to define two different approaches for these calculations: the so called *fixed flavor number scheme* (FFNS) and *variable flavor number scheme* (VFNS). In the former, only massless partons densities are considered for the initial state and that's why FFNS is sometimes referred to as "4 flavor-scheme" (4FS). In the latter a b-quark density is introduced for the initial state, hence VFNS is sometimes called "5 flavor-scheme" (5FS). The two schemes have a different ordering of the perturbative series for the production cross section: in the 4FS the perturbative series is ordered strictly by powers of the strong coupling α_S , whereas in the 5FS the introduction of a b-quark PDF allows to resum terms of the form $\alpha_S^n \ln(m_b^2/M^2)^m$ at all orders in α_S (for fixed order of logarithms m), where M represents the upper integration limit of the b-quark transverse momentum and can be thought to be of the order of M_W or M_Z . At the lowest order, VFNS and FFNS can lead to different results, but from NLO in QCD on they seem to be consistent within their uncertainties, although studies are still in progress. In the following we present the predictions for the inclusive production of Z with one and two b-tagged jets.

3.2.1 $Z + b$ in the fixed flavor number scheme

The FFNS accounts for the production of a Z with a $b\bar{b}$ pair. At leading order (LO) there are two subprocesses that contribute to it, $q\bar{q} \rightarrow Zb\bar{b}$ and $gg \rightarrow Zb\bar{b}$ (where $q = u, d, s, c$). The tree level diagrams contributing to the LO cross section for both subprocesses are shown in figure 3.5 and 3.6.

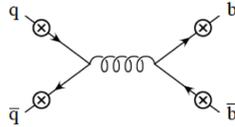


Figure 3.5: Tree level Feynman diagrams for $q\bar{q} \rightarrow Zb\bar{b}$. The circled crosses correspond to all possible insertions of the Z boson, each one representing a different diagram.

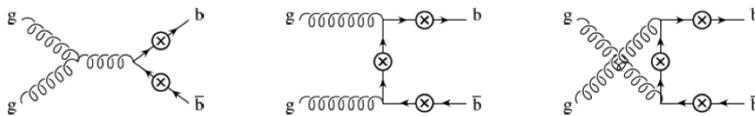


Figure 3.6: Tree level Feynman diagrams for $gg \rightarrow Zb\bar{b}$. The circled crosses correspond to all possible insertions of the Z boson, each one representing a different diagram.

At NLO we have to consider both one-loop virtual corrections to the

LO subprocesses (see figure 3.7) as well as $\mathcal{O}(\alpha_s^3)$ real subprocesses like $q\bar{q} \rightarrow Zb\bar{b} + g$, $gg \rightarrow Zb\bar{b} + g$ and $q(\bar{q})g \rightarrow Zb\bar{b} + q(\bar{q})$ (see figure 3.8).

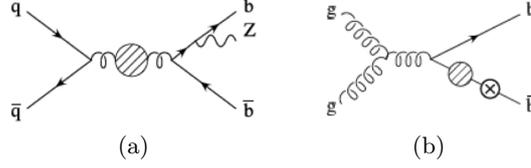


Figure 3.7: One loop virtual corrections to $q\bar{q} \rightarrow Zb\bar{b}$ (a) and $gg \rightarrow Zb\bar{b}$ (b).

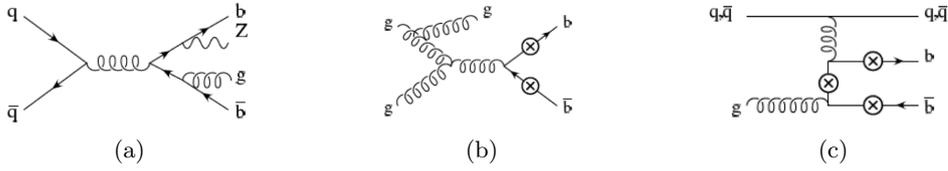


Figure 3.8: Real corrections to $Zb\bar{b}$ cross section

Calculations have been performed by F. Febres Cordero, L. Reina and D. Wackerth (see [12]), including NLO QCD corrections and full bottom-quark mass effects. Numerical results are provided for proton-proton collisions at $\sqrt{s} = 10$ and 14 TeV. A pair of b -tagged jets is required, with $p_T^b > 15$ and 25 GeV and $|\eta^b| < 2.5$. Results are reported in table 3.2: in the inclusive cross sections both 2 and 3 jets events (two of which are b -tagged) are included while in the exclusive one only events with a pair of b -tagged jets are considered. The uncertainties are due to the factorization scale.

3.2.2 $Z + b$ in the variable flavor number scheme

The production of a Z boson together with one or more b -tagged jets can be calculated within the VFNS. The reference papers for these calculations are [14] and [15].

The leading order process for the Zb final state in the VFNS is $gb \rightarrow Zb$, shown in figure 3.9. Other sources of events with only one b -tag are the case



Figure 3.9: LO diagrams for the production of Z in association with a heavy quark Q , where $Q = b$ considering $Z+b$ production.

Table 3.2: Cross sections for inclusive and exclusive $Zb\bar{b}$ production at $\sqrt{s} = 10$ and 14 TeV (from [12]).

	$p_T^b > 15$ GeV	$p_T^b > 25$ GeV
	$\sqrt{s} = 10$ TeV	
σ_{LO} (pb)	55.1^{+16}_{-12}	$24.6^{+7.6}_{-5.4}$
$\sigma_{NLO,inc}$ (pb)	82.5^{+12}_{-11}	$36.0^{+3.9}_{-4.6}$
$\sigma_{NLO,exc}$ (pb)	$52.1^{+0.1}_{-1.7}$	$24.6^{+0.3}_{-1.2}$
	$\sqrt{s} = 14$ TeV	
σ_{LO} (pb)	101^{+26}_{-20}	$46.8^{+13.1}_{-9.6}$
$\sigma_{NLO,inc}$ (pb)	145^{+20}_{-17}	$66.6^{+8.8}_{-8.3}$
$\sigma_{NLO,exc}$ (pb)	$88.4^{+0.1}_{-3.0}$	$43.7^{+0.1}_{-1.6}$

when a $b\bar{b}$ pair is produced but only one b-quark is within the acceptance of the detector, and also the case in which a $b\bar{b}$ pair is in the same b-tagged jet, but these contributions are rather small with respect to the first. NLO corrections can be applied, such that additional partons are produced in the final state in processes like $bg \rightarrow Zbg$, $bq \rightarrow Zbq$ or $qg \rightarrow Zb\bar{b}q$ (figure 3.10). In table 3.3 numerical values of the cross sections as calculated

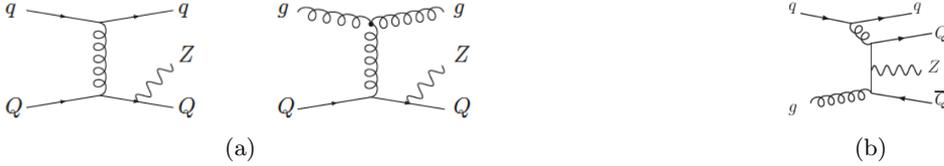


Figure 3.10: NLO QCD processes for Zb production.

by J. Campbell, R. K. Ellis, F. Maltoni, and S. Willenbrock [14] for the LHC (pp at $\sqrt{s} = 14$ TeV) are listed together with uncertainties coming from factorization scale, renormalization scale and PDF respectively. The b-tagged jets are required to have $p_T^b > 15$ GeV and $|\eta^b| < 2.5$ and $m_b = 0$ approximation is used throughout.

The VFNS leads to different results from FFNS in the cross sections for the inclusive production of $Z + 2$ b-jets, but when NLO corrections are applied the two approaches give compatible results, as it's possible to see from the comparison between table 3.2 and 3.3.

Table 3.3: Cross sections for inclusive production of Zb and Zbb in VFNS, in p-p collisions at $\sqrt{s} = 14$ TeV.

	σ_{LO} (pb)	σ_{NLO} (pb)
Z+b (incl.)	826	1060
Z+2b (incl.)	111	$153^{+20+2+9}_{-20-2-9}$

Chapter 4

Z+b events selection with CMS detector

An analysis is being carried out within the CMS collaboration to select Z+b events in order to get a measurement of Z+b cross section. In this chapter we present this analysis, focusing on the event selection carried out for the data collected in 2010: Monte Carlo and data samples are described, Z reconstruction and b-tagging techniques are presented and finally event yields are listed and control plots of some important spectra are shown.

4.1 Monte Carlo and data samples

All Monte Carlo (MC) samples used in this analysis are generated with MADGRAPH [16] and interfaced with PYTHIA 6.4 [17] for the parton shower and the hadronization. They are:

- Dedicated **Z+b** and **Z+c** samples

These samples correspond to a FFNS (4FS) approach and account for the process $pp \rightarrow ZQ\bar{Q}$, $Z \rightarrow l^+l^-$, where Q can be a b or c quark, generated with finite mass ($m_b \simeq 4.2$ GeV, $m_c \simeq 1.3$ GeV). Some kinematical cuts are imposed at generator level:

$$p_T^Q > 15 \text{ GeV}$$

$$M_{l^+l^-} > 40 \text{ GeV}, |\eta^l| < 2.5$$

- **Z+j** sample

This sample corresponds to an inclusive Z + jets sample, where at matrix element (ME) level the generated process is $pp \rightarrow Z + n$ partons, $Z \rightarrow l^+l^-$, with $n = 0, 1, 2, 3, 4$. Here also heavy quarks PDFs are considered and therefore this sample should be regarded as a VFNS

(5FS) sample. Heavy quarks are generated massless at ME level. The following kinematic cuts are applied to jets and di-lepton candidates:

$$p_T^j > 10 \text{ GeV}, |\eta^j| < 5$$

$$M_{l+l-} > 50 \text{ GeV}$$

Additional hard partons could be added during PYTHIA processing and for this reason a k_T MLM matching procedure is required to avoid double counting.

- **tt+j** sample

This sample accounts for the production of a $t\bar{t}$ pair in association with jets.

For a cross section measurement, efficiency and acceptance of signal events have to be defined using simulated events. There are two ways to do this:

- Use the dedicated $Z+b$ sample as a signal. For this approach a merging procedure is required with the $Z+j$ sample: $Z+j$ events with $p_T^{b,c} > 15 \text{ GeV}$ are vetoed allowing only $Z + \text{jets}$ events, with jets coming from light partons or gluons, to pass the filter and to be used in the analysis. This definition of the signal is FFNS-like.
- Use the inclusive $Z+j$ sample to define the signal, whenever a b -quark is generated in the final state and simply don't use $Z+b$ and $Z+c$ samples. This definition of the signal is VFNS-like.

In the FFNS approach, $Z+c$ sample and the vetoed $Z+j$ sample described before are used as background to account for the chance to mistag the jets, i.e. to b -tag a jet that is not initiated from a b quark. In both FFNS and VFNS approaches, $tt+j$ sample is included in the background.

In table 4.1 the relevant parameters of all the MC samples are listed. All samples are rescaled to their *equivalent luminosity*, which is defined as:

$$\mathcal{L}_{eq} = \frac{N_{evts}}{\sigma}$$

The datasets used for 2010 data analysis are listed in table 4.2. In order to simplify the analysis, within a dataset only events selected by unrescaled triggers were considered. In fact, in 2010 LHC runnings the instantaneous luminosity changed very often together with the triggers and the thresholds of unrescaled trigger was raised many times.

The total integrated luminosity corresponding to these datasets is:

$$\int \mathcal{L} dt = 35.9 \pm 1.4 \text{ pb}^{-1}$$

Table 4.1: MC samples used for Z+b analysis.

Dataset	Sample	N_{evts}	NLO σ (pb)	\mathcal{L}_{eq} (pb^{-1})
ZbbToLL	Z+b	26840	26.3	1020
ZccToLL	Z+c	40000	41.1	973
DYJetsToLL	Z+j	5172930	3048 (NNLO)	1697
TTJets-tauola	tt+j	200000	165 (NNLO)	1212

Table 4.2: Datasets, run ranges and corresponding triggers used for 2010 Z+b events selection.

Dataset	Run range	Trigger
EG	135821-137028	HLT_Photon10_L1R
EG	138564-140401	HLT_Photon15_Cleaned_L1R
EG	141956-144114	HLT_Ele15_SW_CaloEleId_L1R
Electron	146428-147116	HLT_Ele17_SW_CaloEleId_L1R
Electron	147196-148102	HLT_Ele17_SW_TightEleId_L1R
Electron	148822-149063	HLT_Ele22_SW_TighterCaloIdIsol_L1R.v1
Electron	149181-149442	HLT_Ele22_SW_TighterCaloIdIsol_L1R.v2
Muon	135821-149442	HLT_Mu9 OR HLT_Mu11 OR HLT_Mu15

4.2 Reconstruction and selection of Z+b candidates

For the selection of Z+b events, standard selection cuts on physics objects like leptons and jets are applied to ensure good reconstruction and rejection of fakes. Also acceptance sequential cuts are implemented in order to define different event categories, the last being the Z+b candidates. In the following every selection step is described and the yields of the different event categories are listed including MC expectations. Plots of control variables are also shown to complete the data/Monte Carlo comparison.

4.2.1 Di-lepton selection

The Z boson in the event is reconstructed via the pair of leptons coming from its decay. The di-lepton is built with a pair of electrons or muons with same flavor and opposite charge.

The leptons are requested to satisfy additional standard cuts, which are

modeled on the reconstruction of W and Z and are widely applied in CMS in the analyses involving vector bosons. In the following they will be referred to as *tight leptons*. These include acceptance cuts:

$$p_T^e > 25 \text{ GeV}, |\eta^e| < 2.5$$

and:

$$p_T^\mu > 20 \text{ GeV}, |\eta^\mu| < 2.1$$

Isolation cuts are based on the sum of p_T and E_T of tracks lying in a cone of dimension $\Delta R = 0.3$ centered around the lepton track. Tight muons satisfy:

$$\frac{\sum_{TRK} p_T + \sum_{ECAL} p_T + \sum_{HCAL} p_T}{p_T} < 0.15$$

where TRK stands for tracker, ECAL for the electromagnetic calorimeter, HCAL for the hadronic calorimeter. For the electrons:

$$\frac{\sum_{TRK} p_T}{p_T} < 0.09, \frac{\sum_{ECAL} p_T}{p_T} < 0.08, \frac{\sum_{HCAL} p_T}{p_T} < 0.1$$

Muons are required to be reconstructed in the muon chambers and matched with a track in the tracker ("global muons"). Additional standard requests on the number of hits in the detector are implemented in order to ensure a good track quality for the muon candidate.

For electrons, an identification cut is applied, based on the matching between track and cluster in the ECAL. A rejection for the electrons coming from photon conversion is implemented too, and a "cleaning" procedure is applied between jet candidates and electrons to reject the jets which are seen as electrons in the detector.

Both electrons and muons have to be matched with a trigger object in $\Delta R < 0.3$ and $\Delta p_T/p_T < 0.5$.

Finally, the Z candidate is built requiring that the di-lepton invariant mass lies in the mass window:

$$60 \text{ GeV} < M_{l+l-} < 120 \text{ GeV}$$

4.2.2 Jet selection

The jet candidates are reconstructed with the *anti- k_T particle flow jet clustering algorithm*, using a cone of $\Delta R = 0.5$ (see Appendix C for more details). Jet energy corrections are applied, which aim to correct for detector effects and to approach the energy of the parton which originated the jet (see [8] for more details).

A standard jet identification cut is applied, based on the deposits in the ECAL and HCAL of the charged and neutral fractions of the jet. The definition of *good jet*, which will be used throughout, is based on the following kinematic cuts:

$$p_T^j > 25 \text{ GeV}, |\eta^j| < 2.1$$

4.2.3 B-tagging

B-tagging is a crucial tool for Z+b measurement. The b-tagging algorithm used in 2010 Z+b selection is the *simple secondary vertex* (SSV), which is described in Appendix B. Two different "taggers" are used, defining two different selection steps:

- High efficiency discriminator, working point "medium" (HEM): it corresponds to the SSVHE discriminator value of:

$$\log \left(1 + \frac{D}{\sigma_D} \right) \geq 1.74$$

and leads to a mis-tagging rate of about 1%.¹

- High purity discriminator, working point "tight" (HPT): it corresponds to the SSVHP discriminator value of:

$$\log \left(1 + \frac{D}{\sigma_D} \right) \geq 2$$

and leads to a mis-tagging rate of about 0.1%.

4.3 Results

The following event categories are defined, each one with additional requests with respect to the previous one:

- **Zll**: tight Z candidate
- **Good jet**: at least 1 good jet
- **HE tag**: at least 1 HE b-tagged good jet
- **HP tag**: at least 1 HP b-tagged good jet

In table 4.3 the event yields for each of the selection steps described above are shown with their statistical uncertainties.

¹This value are estimated from Monte Carlo studies using QCD samples.

Table 4.3: The event yields for 2010 data in the muon and in the electron channel.

Selection	Data	Z+b	\sum MC
muon channel			
Zll	11951 ± 109	179 ± 3	12187 ± 23
Good jet	2194 ± 47	111 ± 2	2267 ± 11
HE tag	101 ± 10	40.9 ± 1.2	94 ± 3
HP tag	48 ± 7	35.8 ± 1.1	53.8 ± 2.3
electron channel			
Zll	8477 ± 92	118 ± 2	8874 ± 19
Good jet	1464 ± 38	71.5 ± 1.6	1519 ± 9
HE tag	54 ± 7	32.6 ± 1.1	67.3 ± 3.0
HP tag	31 ± 6	22.5 ± 0.9	34.1 ± 1.8

In figure 4.1 the yields are shown as histograms, detailing the different contributions of the samples with different colors. The disagreement in the first steps of the selection (before the cut on the mass of the Z candidate) is due to the lack of MC events at low dilepton invariant masses with respect to the data. This figure also shows how the fraction of the signal (red) increases after every additional requirement.

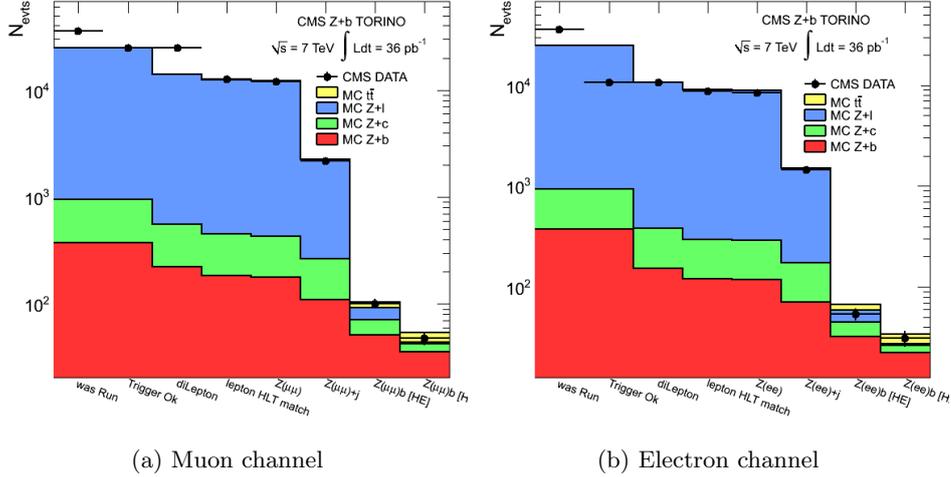


Figure 4.1: Event yields at different selection steps.

From a first look at the yields, we can see that there is a reasonable agreement between data/MC.

In figure 4.2 through 4.5 some interesting spectra are shown, for the FFNS-

like samples at the "Good jet" selection step, for kinematic variables of Z candidates, leading jet, leading muon and leading electron.

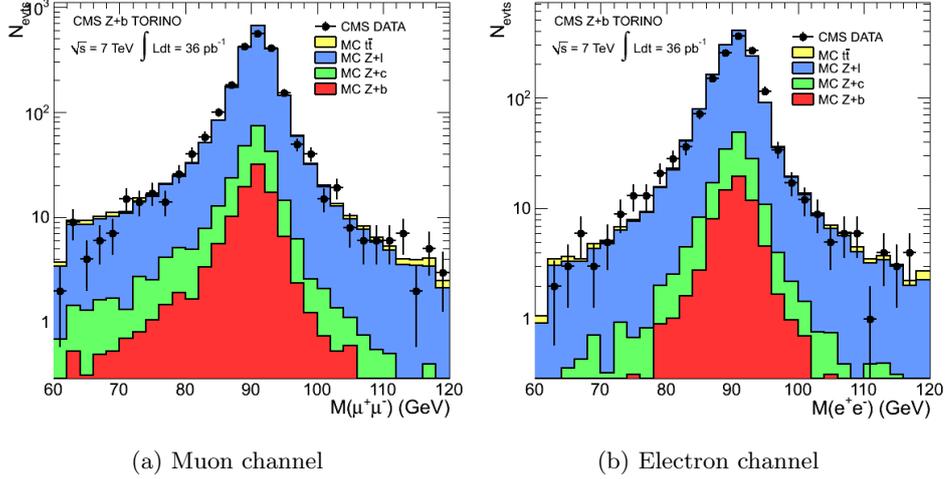


Figure 4.2: Invariant mass of Z candidates at the "Good jet" selection step.

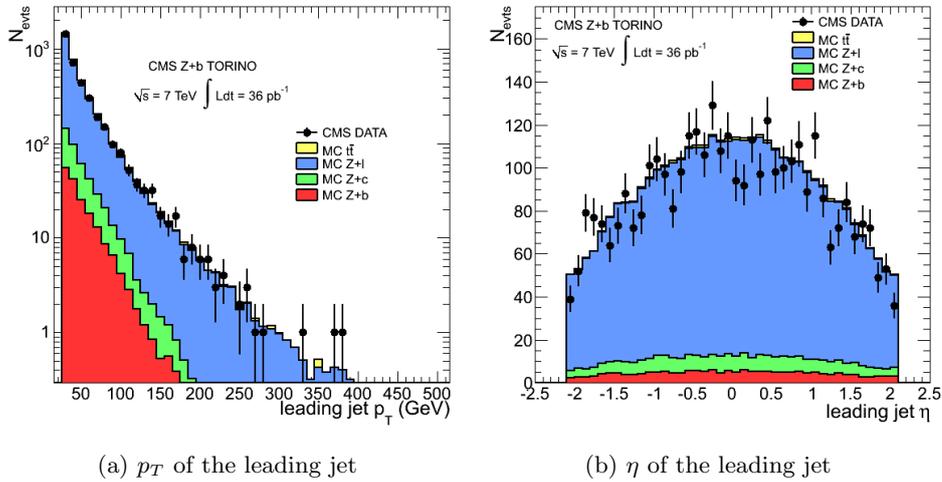


Figure 4.3: Kinematic distributions for the leading jet in the event at the "Good jet" selection step.

In these plots the acceptance cuts applied are clearly visible, for example $|\eta| < 2.1$ for jets or $p_T > 20$ GeV for muons coming from Z decays. In figure 4.5 (b) it's also possible to see the "crack-points", i.e. the η values in the separation between barrel and endcap detectors.

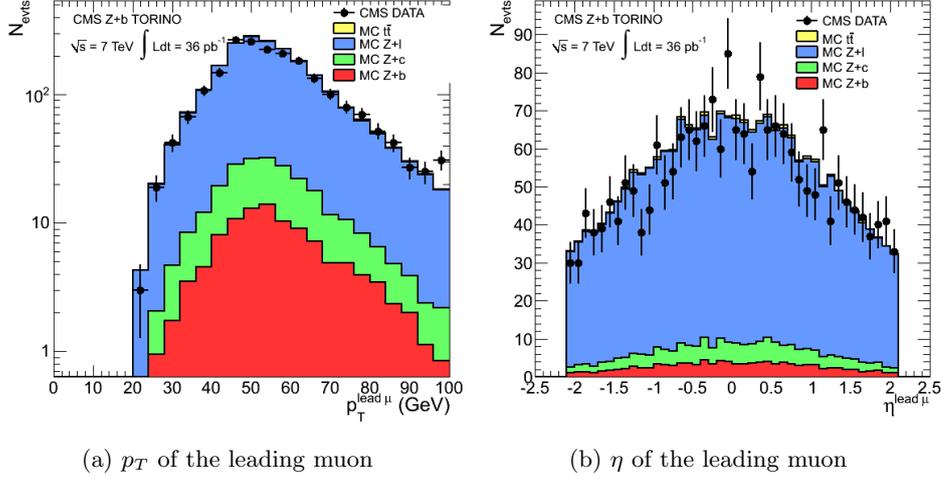


Figure 4.4: Kinematic distributions for the leading muon in the event at the "Good jet" selection step.

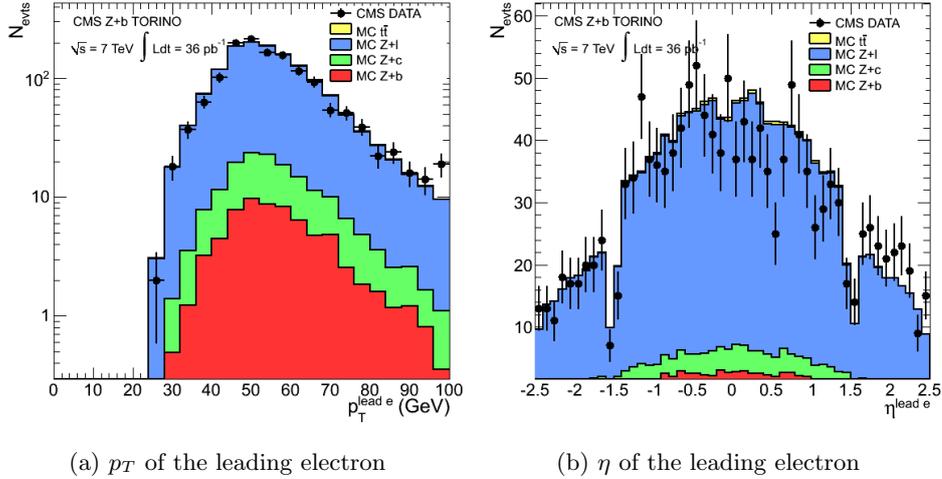


Figure 4.5: Kinematic distributions for the leading electron in the event at the "Good jet" selection step.

In general the spectra are well reproduced by MC and also the statistics is large enough to conclude that at $Z + \text{jet}$ level the agreement is good. In the "HE tag" selection step there is still little statistics, but nevertheless the agreement is reasonable, within 20 % almost everywhere (see figures 4.6 and 4.7).

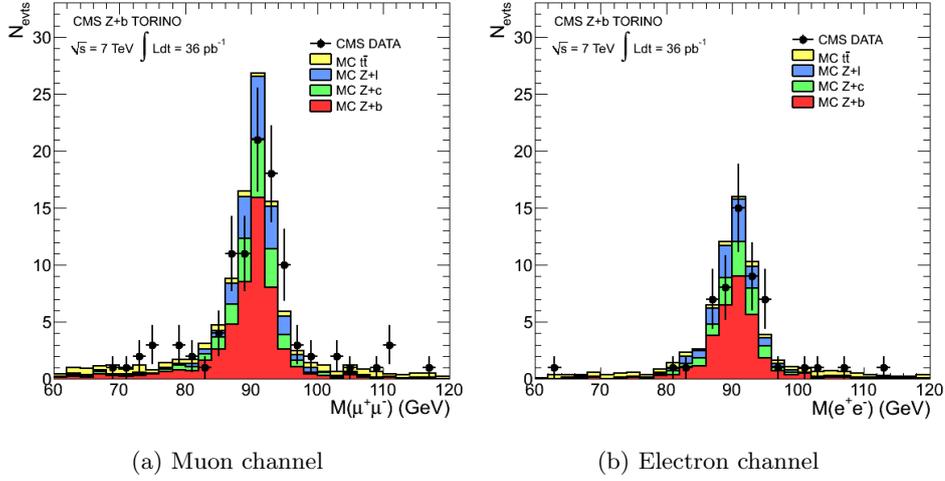


Figure 4.6: Invariant mass of Z candidates at "HE tag" selection step.

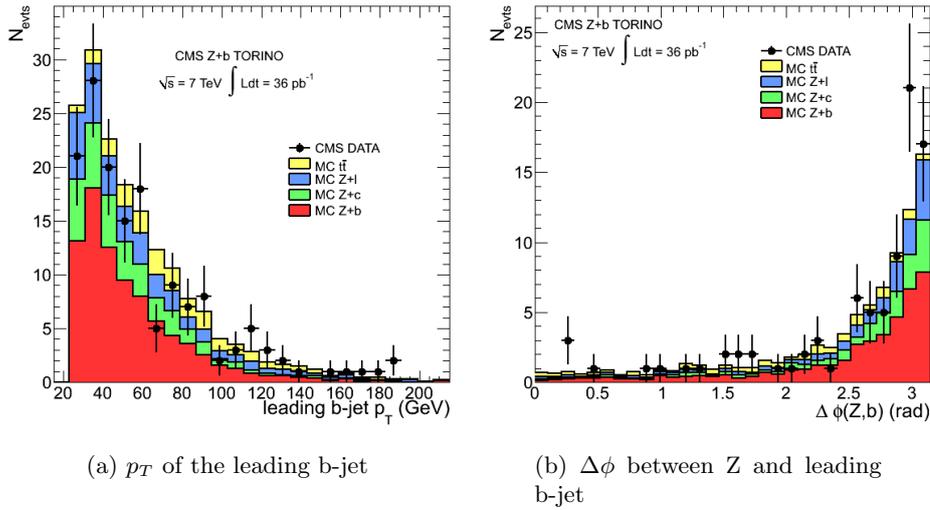


Figure 4.7: Kinematic distributions for the leading b-jet in the event at the "HE tag" selection step.

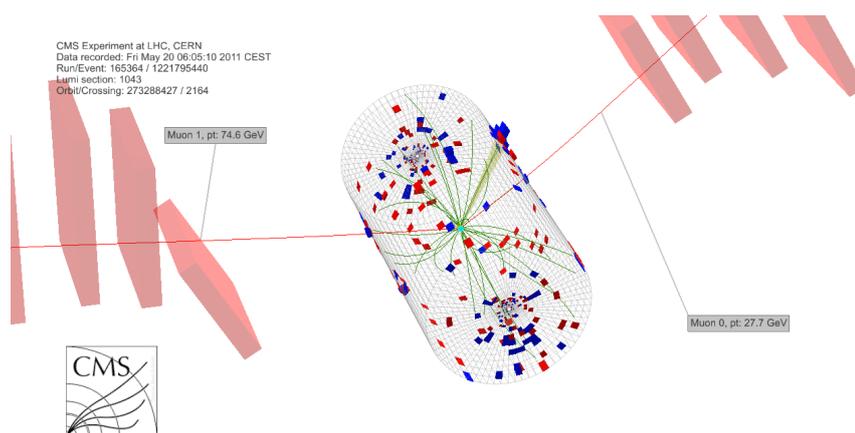
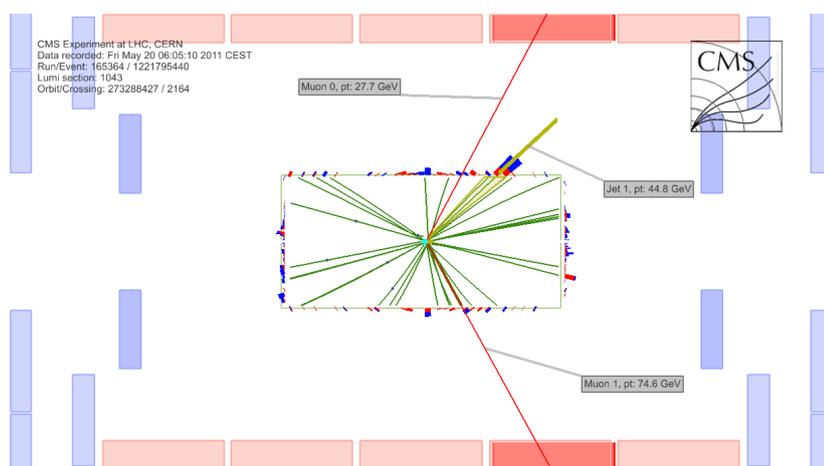
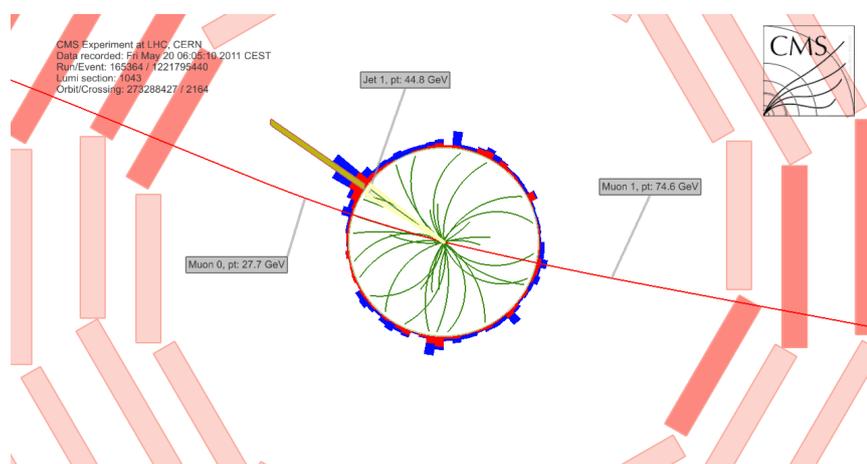
We don't go further in this analysis, that was presented here for 2010 data as a introduction to the tools used for the Z+b selection.

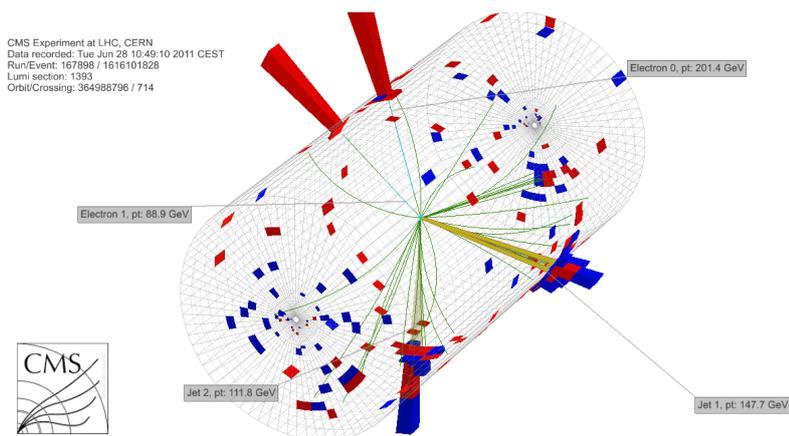
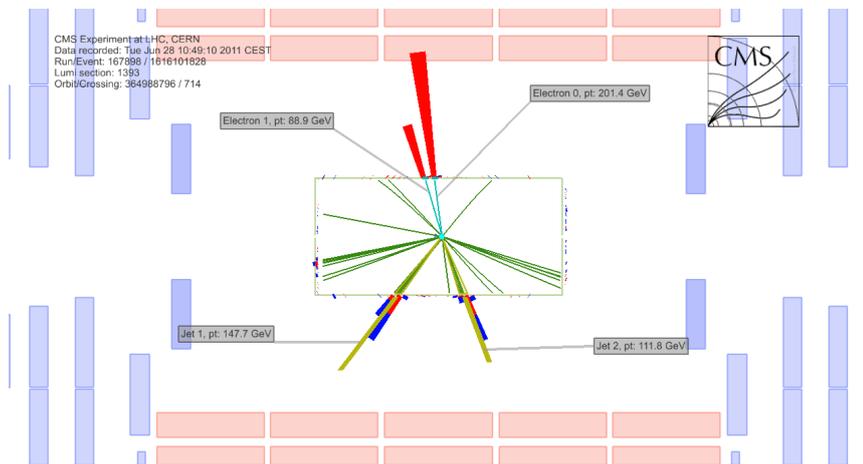
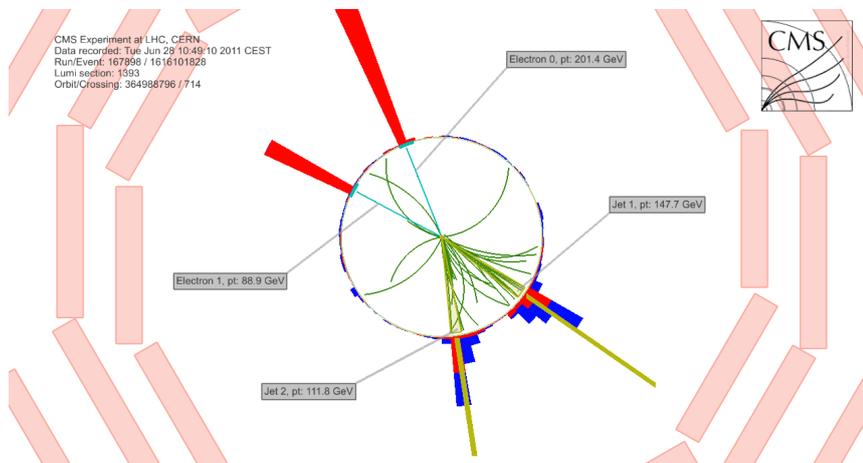
4.4 Event display

Finally, in figures 4.8 and 4.9 we show a $Z(\mu^+\mu^-) + b$ and a $Z(e^+e^-) + b$ candidate events selected in 2010 data.

The following color conventions are used:

- muons are in red;
- electrons are in azure;
- jets are shown as yellow cones around the axis;
- ECAL towers are in red;
- HCAL towers are in blue;
- general tracks are in green;
- MET (missing transverse energy) are shown as red arrows.

Figure 4.8: A $Z(\mu^+\mu^-) + b$ candidate event.

Figure 4.9: A $Z(e^+e^-) + b$ candidate event.

Chapter 5

Z+b events with more than two leptons

As it was anticipated in section 3.1.1 the Z+b process is a background for the search of SM Higgs in the golden channel since events with additional leptons coming from semileptonic decays of b -quark can lead to the reconstruction of a fake Z^* . In this chapter an analysis of "multi-lepton" events is presented, based on an integrated luminosity of 1.14 fb^{-1} , in p-p collisions at $\sqrt{s} = 7 \text{ TeV}$ (in 2011 analysis, simulated events are reweighted as described in Appendix D). Initially the selection of muons and electrons is discussed, then the yields of multi-lepton events are listed for different b-tagging algorithms, control plots at the Z+b multi-lepton selection step and plots of the additional leptons kinematic variables are shown and the results of data/Monte Carlo comparison are presented.

5.1 Selection of additional leptons

The selection of additional leptons has to be tuned carefully in order to achieve good efficiency together with a reliable rejection of fakes. In considering both efficiency and purity, many standard strategies have been developed in CMS software which allow to select leptons coming from decays of W and Z , (see section 4.2). However, those cuts are in general not efficient enough for this kind of analysis, especially as far as tight isolation and p_T cuts are concerned. Hence, it has been decided to use the same selection criteria with relaxed cuts to increase the number of leptons selected.

The benchmark for this selection is the simulated signal sample in which muons and electrons coming from semileptonic decays of b -quarks can be studied. In the following they will be referred to as "b-muons" and "b-electrons" for simplicity. The simulated sample is extracted from the inclusive Z+jets sample described in section 4.1, requiring the presence of a b -quark in the final state.

5.1.1 Muon selection

Muons are required to be reconstructed both with the "global muon" and "tracker muon" algorithms (see section 2.3.1). Besides standard identification and reconstruction, additional cuts on the acceptance and quality of the track are implemented. They are:

- $p_T > 5$ GeV and $|\eta| < 2.1$;
- Normalized χ^2 of track fit < 15 ;
- Number of tracker hits > 10 ;
- Number of pixel hits > 0 ;
- Number of muon station hits > 0 ;
- Number of muon chambers with matched segments > 1 ;
- Distance in z between the vertex of the Z candidate and muon track impact parameter < 1 mm.

The last cut is intended to associate the candidate muon track to the Z vertex, in order to reject tracks coming from pile-up events, i.e. multiple p-p interactions in the same bunch crossing.

Compared to the selection of muons from Z , the p_T cut has been considerably lowered and the isolation cut has been removed. The combined relative isolation, as it was defined in section 4.2.1 for b-muons is shown in figure 5.1. From this plot, it's clear how the isolation cut used for the "tight muons" selection applied for the $Z(\mu^+\mu^-)$ reconstruction would remove most of the additional muon candidates.

In order to test the efficiency of these selection cuts, a comparison between generated and reconstructed b-muons has been performed. From this study we extract the *selection efficiency* defined as:

$$\epsilon_{sel,\mu} = \frac{N_{reco,\mu}}{N_{gen,\mu}} \quad (5.1)$$

where $N_{reco,\mu}$ is the number of reconstructed b-muons and $N_{gen,\mu}$ is the number of generated b-muons.

For this study both "prompt" b-muons and muons coming from $b \rightarrow c \rightarrow \mu$ cascade decay are considered. In figures 5.2 p_T and η spectra of generated and reconstructed b-muons are shown, extracted from the simulated signal sample.

The selection efficiency as a function of the p_T is shown in figure 5.3. The error on the efficiency is estimated using the usual formula for the hypothesis of binomial distribution:

$$\delta\epsilon_{sel,\mu} = \sqrt{\frac{\epsilon_{sel,\mu}(1 - \epsilon_{sel,\mu})}{N_{gen,\mu}}}$$

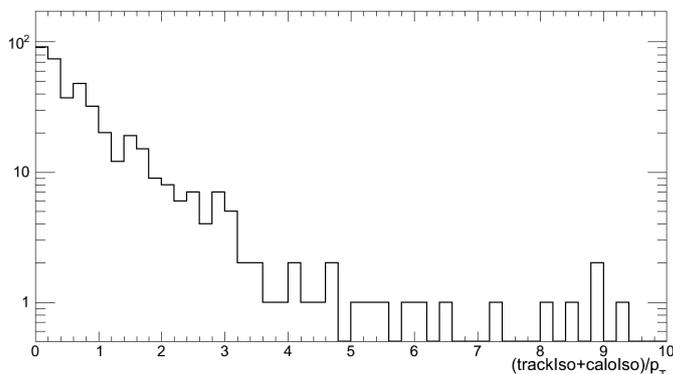


Figure 5.1: Combined relative isolation of selected muons coming from semileptonic decays of the b quark.

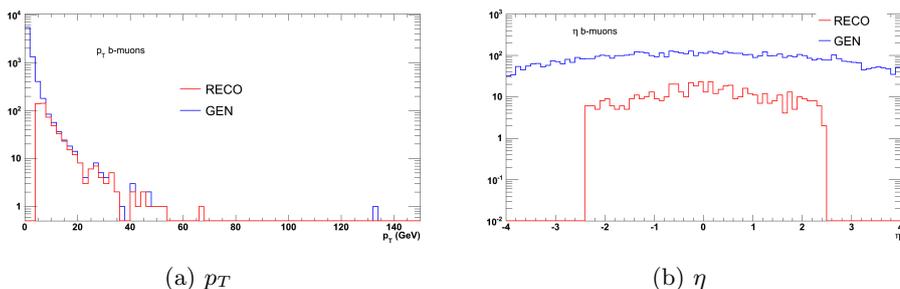


Figure 5.2: p_T (a) and η (b) of generated and reconstructed b-muons.

Even though the statistics is not fully satisfactory, it's possible to make some considerations. As expected the efficiency is null for $p_T < 5$ GeV, reaching a plateau of nearly 80%-90% around 10 GeV. The efficiency over the whole p_T range is:

$$\epsilon_{sel,\mu} = 0.070 \pm 0.003$$

where the lack of efficiency is mainly due to the p_T and the η cut on the muons.

5.1.2 Electron selection

As it was done for muons in the previous section, here the selection of additional electrons is presented. The guideline is again the selection applied for the reconstruction of $Z(e^+e^-)$ in the $Z+b$ analysis, in which the p_T and isolation cuts are respectively lowered and removed. In details, the applied cuts are:

- $p_T > 5$ GeV and $|\eta| < 2.5$;

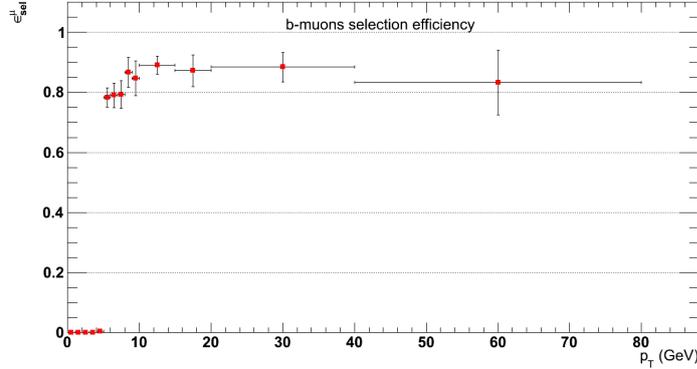


Figure 5.3: Selection efficiency of b-muons as a function of p_T .

- standard rejection of converted photons;
- standard matching between track and ECAL cluster ($\Delta\phi < 0.8$ and $\Delta\eta < 0.007$ for the barrel and $\Delta\phi < 0.7$ and $\Delta\eta < 0.01$ for the endcap);
- standard shower shape cuts;
- rejection of electrons from the gap between barrel and endcap;
- distance in z between the vertex of the Z candidate and electron track impact parameter < 1 mm.

The combined isolation of b-electrons is shown in figure 5.4. The same considerations made for the isolation cut about b-muons are still valid for b-electrons. In figure 5.5 the p_T and η distributions of generated and reconstructed b-electrons as they are extracted from the MC signal sample are shown.

With similar definitions with respect to the previous section, the selection efficiency for b-electrons is studied again as a function of p_T and it is shown in figure 5.6. For the electrons less statistics is available, and the trend of the efficiency is not so clear, but it seems to reach the usual plateau, like the muon efficiency (figure 5.3), but at a lower value ranging from 20%-40%. The value of the selection efficiency over the full p_T range is:

$$\epsilon_{sel,e} = 0.008 \pm 0.001$$

For the electron efficiency considerations similar to the ones for muon the efficiency are valid as well.

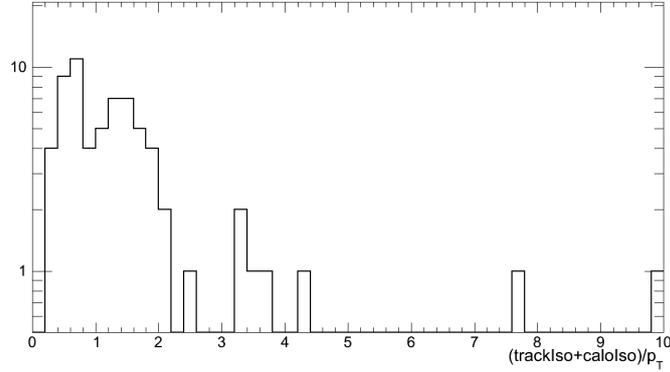


Figure 5.4: Combined relative isolation of selected electrons coming from semileptonic decays of the b quark.

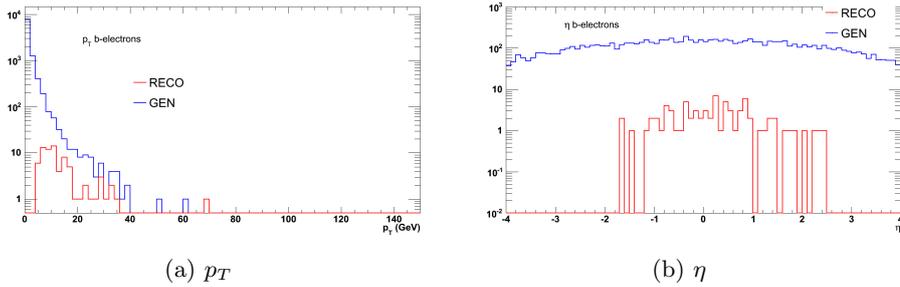


Figure 5.5: p_T (a) and η (b) of generated and reconstructed b-electrons.

5.2 Event Yields

The selection of events with additional leptons has been carried out with four different "taggers" for b-jets (see Appendix B and section 2.3.4), corresponding to two different b-tagging algorithms, the track counting (TC) algorithm and the simple secondary vertex (SSV) algorithm. The taggers correspond to different discriminators and working points (WP): they are SSVHEM, SSVHPT, TCHEL and TCHEM, where "HE" stands for "high efficiency" discriminator, "HP" for "high purity" discriminator, "L" for "loose" WP, "M" for "medium" WP and "T" for "tight" WP.

For this analysis, the η cut defining the "good jet" selection described in section 4.3 has been extended from $|\eta| < 2.1$ to $|\eta| < 2.4$, which is the limit value for the use of b-tag tools, since they rely on the tracker measurements. The main reason for the use of this relaxed cut and looser b-taggers, such as the TCHEL, is to increase the statistics in multi-lepton events and to cross check the comparison between selected and expected events in a wider

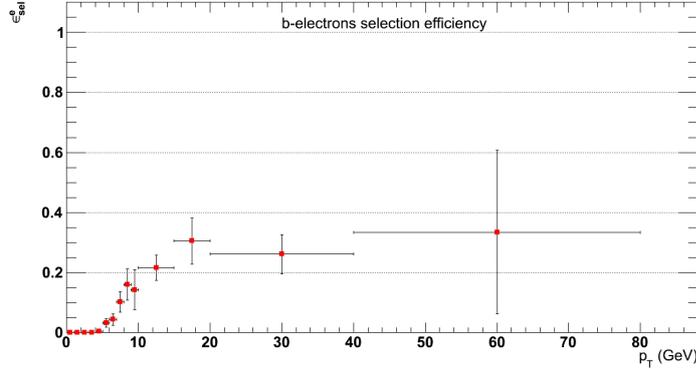


Figure 5.6: Selection efficiency of b-electrons as a function of p_T .

region of the phase space and for different b-tag tools.

Events with additional leptons are divided into 5 different exclusive samples, namely:

- μ channel, which corresponds to the $Zb + \mu$ selection
- e channel, which corresponds to the $Zb + e$ selection
- $\mu\mu$ channel, which corresponds to the $Zb + \mu\mu$ selection
- $e\mu$ channel, which corresponds to the $Zb + e\mu$ selection
- ee channel, which corresponds to the $Zb + ee$ selection

In all the five channels considered, the Z is allowed to decay either in e^+e^- and in $\mu^+\mu^-$. In the channels with two additional leptons no requirement on the charge of the additional leptons is made.

In tables 5.1 through 5.4 the yields of multi-lepton events, with binomial errors (see equation (5.1.1)), are shown corresponding to different values of the η cut on the b-jets. The results are shown for the four different taggers described above, and for all the five channels. The yields quoted in these tables refer to an inclusive $Z+b$ selection, i.e. events with a reconstructed Z boson with at least one b-tagged jet. The yields are also shown in the histograms in figures 5.7 through 5.10. From Monte Carlo truth information about 90%-95% of the selected additional muons come from semileptonic decays of a b quark. For the electrons this percentage is close to 100%.

The TCHEL selection has the highest yield, as expected from the studies of b-tagging efficiencies (see section 2.3.4), while the SSVHPT has the lowest. The events with only one additional lepton range from about 370 to 1000, depending on the b-tagging discriminator used and the channel considered,

while the events with two additional leptons range from about 20-45. The yields don't increase very much with larger acceptance on the b-jet. However, in figure 5.11 the yields as a function of the η cut on the b-tagged jet are shown. In the μ channel we observe the major increase, of about 5%, from $|\eta| < 2.1$ to $|\eta| < 2.4$. In the channels with two additional leptons the number of selected events does not actually change by enlarging the η cut on the b-tagged jet.

For the channels with one additional lepton, the data/MC agreement is in general very good, even though for the e channel we observe a small excess, especially when using the TCHEL and TCHEM discriminators. For the $\mu\mu$ channel the agreement is pretty good for all the selections. In the $e\mu$ and ee channel the lack of statistics, caused by the electron efficiency combined with lower expected rate, does not allow to make definitive conclusions. Nevertheless, for the ee channel the yields seem to be in agreement with the predictions, while in the $e\mu$ channel we steadily observe a lack of events. Generally speaking, the channels containing only additional muons are in good agreement with MC predictions, while the channels with additional electrons show small deviations from the expectations. This disagreement is likely due to the simulation of detector response, rather than a bad description of physical processes.

Table 5.1: Multi-lepton event yields with the $|\eta| < 2.1$ cut on the b-jet.

SSVHEM						
channel	$t\bar{t}$	$Z + b$	$Z + c$	$Z + l$	sum MC	data
μ	114 ± 5	290 ± 20	77 ± 10	9 ± 3	490 ± 20	480 ± 20
e	28 ± 2	82 ± 10	18 ± 5	5 ± 2	133 ± 11	133 ± 12
$\mu\mu$	9.7 ± 1.4	17 ± 4	-	-	26 ± 5	26 ± 5
$e\mu$	5.8 ± 1.1	6 ± 3	-	-	12 ± 3	7 ± 3
ee	1.2 ± 0.5	1.2 ± 1.2	-	-	2.3 ± 1.3	2.0 ± 1.4
SSVHPT						
channel	$t\bar{t}$	$Z + b$	$Z + c$	$Z + l$	sum MC	data
μ	89 ± 4	186 ± 15	19 ± 5	-	294 ± 16	281 ± 17
e	21 ± 2	55 ± 8	4 ± 2	-	80 ± 9	84 ± 9
$\mu\mu$	8.4 ± 1.2	9 ± 3	-	-	18 ± 4	18 ± 4
$e\mu$	4.7 ± 1.0	5 ± 2	-	-	9 ± 3	4 ± 2
ee	1.0 ± 0.4	-	-	-	1.0 ± 0.4	1 ± 1
TCHEL						
channel	$t\bar{t}$	$Z + b$	$Z + c$	$Z + l$	sum MC	data
μ	133 ± 5	400 ± 20	154 ± 14	46 ± 7	730 ± 30	739 ± 30
e	33 ± 3	120 ± 12	54 ± 8	57 ± 8	264 ± 16	305 ± 18
$\mu\mu$	10.1 ± 1.4	18 ± 5	-	-	28 ± 5	32 ± 6
$e\mu$	6.6 ± 1.1	11 ± 4	-	-	17 ± 4	10 ± 3
ee	1.4 ± 0.5	1.2 ± 1.2	-	-	2.5 ± 1.3	4 ± 2
TCHEM						
channel	$t\bar{t}$	$Z + b$	$Z + c$	$Z + l$	sum MC	data
μ	118 ± 5	310 ± 20	68 ± 9	7 ± 3	506 ± 20	500 ± 20
e	29 ± 2	90 ± 10	9 ± 3	7 ± 3	135 ± 11	158 ± 13
$\mu\mu$	9.9 ± 1.4	15 ± 4	-	-	25 ± 5	27 ± 5
$e\mu$	6.6 ± 1.1	8 ± 3	-	-	15 ± 3	8 ± 3
ee	1.2 ± 0.5	1.2 ± 1.2	-	-	2.3 ± 1.3	2.0 ± 1.4

Table 5.2: Multi-lepton event yields with the $|\eta| < 2.2$ cut on the b-jet.

SSVHEM						
channel	$t\bar{t}$	$Z + b$	$Z + c$	$Z + l$	sum MC	data
μ	115 ± 5	290 ± 20	77 ± 10	9 ± 3	490 ± 20	480 ± 20
e	28 ± 2	85 ± 10	19 ± 5	6 ± 3	137 ± 11	134 ± 12
$\mu\mu$	10 ± 1	17 ± 5	-	-	26 ± 5	26 ± 5
$e\mu$	5.8 ± 1.1	7 ± 3	-	-	13 ± 3	7 ± 3
ee	1.2 ± 0.5	1.2 ± 1.2	-	-	2.3 ± 1.3	2.0 ± 1.4
SSVHPT						
channel	$t\bar{t}$	$Z + b$	$Z + c$	$Z + l$	sum MC	data
μ	89 ± 4	186 ± 15	19 ± 5	-	294 ± 16	283 ± 16
e	21 ± 2	55 ± 8	4 ± 2	1.2 ± 1.2	81 ± 9	85 ± 9
$\mu\mu$	8.4 ± 1.3	9 ± 3	-	-	18 ± 4	18 ± 4
$e\mu$	4.7 ± 1.0	5 ± 2	-	-	9 ± 3	4 ± 2
ee	1.0 ± 0.4	-	-	-	1.0 ± 0.4	1 ± 1
TCHEL						
channel	$t\bar{t}$	$Z + b$	$Z + c$	$Z + l$	sum MC	data
μ	134 ± 5	410 ± 20	157 ± 14	48 ± 8	750 ± 30	750 ± 30
e	34 ± 3	124 ± 12	57 ± 8	59 ± 8	273 ± 16	307 ± 18
$\mu\mu$	10.1 ± 1.4	18 ± 5	-	-	28 ± 5	32 ± 6
$e\mu$	6.6 ± 1.1	11 ± 4	-	-	17 ± 4	10 ± 3
ee	1.4 ± 0.5	1.2 ± 1.2	-	-	2.5 ± 1.2	4 ± 2
TCHEM						
channel	$t\bar{t}$	$Z + b$	$Z + c$	$Z + l$	sum MC	data
μ	119 ± 5	320 ± 20	69 ± 9	7 ± 3	510 ± 20	502 ± 20
e	29 ± 2	91 ± 10	9 ± 3	8 ± 3	138 ± 11	158 ± 13
$\mu\mu$	9.9 ± 1.4	15 ± 4	-	-	25 ± 5	27 ± 5
$e\mu$	6.6 ± 1.1	8 ± 3	-	-	15 ± 3	8 ± 3
ee	1.2 ± 0.5	1.2 ± 1.2	-	-	2.3 ± 1.3	2.0 ± 1.4

Table 5.3: Multi-lepton event yields with the $|\eta| < 2.3$ cut on the b-jet.

SSVHEM						
channel	$t\bar{t}$	$Z + b$	$Z + c$	$Z + l$	sum MC	data
μ	116 ± 5	295 ± 19	77 ± 10	9 ± 3	500 ± 20	490 ± 20
e	28 ± 2	85 ± 10	19 ± 5	6 ± 3	138 ± 11	134 ± 12
$\mu\mu$	9.7 ± 1.4	17 ± 4	-	-	26 ± 5	26 ± 5
$e\mu$	5.8 ± 1.1	7 ± 3	-	-	13 ± 3	8 ± 3
ee	1.2 ± 0.5	1.2 ± 1.2	-	-	2.3 ± 1.3	2.0 ± 1.4
SSVHPT						
channel	$t\bar{t}$	$Z + b$	$Z + c$	$Z + l$	sum MC	data
μ	90 ± 4	188 ± 15	19 ± 5	-	297 ± 16	287 ± 17
e	21 ± 2	55 ± 8	4 ± 2	1.2 ± 1.2	81 ± 9	85 ± 9
$\mu\mu$	8.4 ± 1.3	9 ± 3	-	-	18 ± 4	18 ± 4
$e\mu$	4.7 ± 1.0	5 ± 2	-	-	9 ± 3	5 ± 2
ee	1.0 ± 0.5	-	-	-	1.0 ± 0.5	1 ± 1
TCHEL						
channel	$t\bar{t}$	$Z + b$	$Z + c$	$Z + l$	sum MC	data
μ	135 ± 5	410 ± 20	159 ± 14	48 ± 8	750 ± 30	760 ± 30
e	34 ± 3	124 ± 12	57 ± 8	60 ± 8	274 ± 16	310 ± 18
$\mu\mu$	10.1 ± 1.4	19 ± 5	-	-	29 ± 5	32 ± 6
$e\mu$	6.6 ± 1.1	11 ± 4	-	-	17 ± 4	11 ± 3
ee	1.4 ± 0.5	1.2 ± 1.2	-	-	2.5 ± 1.3	4 ± 2
TCHEM						
channel	$t\bar{t}$	$Z + b$	$Z + c$	$Z + l$	sum MC	data
μ	120 ± 5	320 ± 20	71 ± 9	7 ± 3	520 ± 20	510 ± 20
e	30 ± 2	91 ± 10	9 ± 3	8 ± 3	138 ± 11	159 ± 13
$\mu\mu$	9.9 ± 1.4	17 ± 4	-	-	26 ± 5	27 ± 5
$e\mu$	6.6 ± 1.1	8 ± 3	-	-	15 ± 3	9 ± 3
ee	1.2 ± 0.5	1.2 ± 1.2	-	-	2.3 ± 1.3	2.0 ± 1.4

Table 5.4: Multi-lepton event yields with the $|\eta| < 2.4$ cut on the b-jet.

SSVHEM						
channel	$t\bar{t}$	$Z + b$	$Z + c$	$Z + l$	sum MC	data
μ	116 ± 5	298 ± 19	77 ± 10	9 ± 3	500 ± 20	490 ± 20
e	28 ± 2	85 ± 10	20 ± 5	6 ± 3	139 ± 12	134 ± 12
$\mu\mu$	9.7 ± 1.4	17 ± 4	-	-	26 ± 5	26 ± 5
$e\mu$	5.8 ± 1.1	7 ± 3	-	-	13 ± 3	8 ± 3
ee	1.2 ± 0.5	1.2 ± 1.2	-	-	2.3 ± 1.3	2.0 ± 1.4
SSVHPT						
channel	$t\bar{t}$	$Z + b$	$Z + c$	$Z + l$	sum MC	data
μ	91 ± 4	190 ± 15	19 ± 5	-	299 ± 16	292 ± 17
e	21 ± 2	55 ± 8	4 ± 2	1.2 ± 1.2	81 ± 9	85 ± 9
$\mu\mu$	8.4 ± 1.3	9 ± 3	-	-	18 ± 4	18 ± 4
$e\mu$	4.7 ± 1.0	5 ± 2	-	-	9 ± 3	5 ± 2
ee	1.0 ± 0.5	-	-	-	1.0 ± 0.5	1 ± 1
TCHEL						
channel	$t\bar{t}$	$Z + b$	$Z + c$	$Z + l$	sum MC	data
μ	135 ± 5	410 ± 20	159 ± 14	51 ± 8	760 ± 30	770 ± 30
e	34 ± 3	124 ± 12	58 ± 8	60 ± 8	275 ± 16	312 ± 18
$\mu\mu$	10.1 ± 1.4	19 ± 5	-	-	29 ± 5	32 ± 6
$e\mu$	6.6 ± 1.1	11 ± 4	-	-	17 ± 4	11 ± 3
ee	1.4 ± 0.5	1.2 ± 1.2	-	-	2.5 ± 1.3	4 ± 2
TCHEM						
channel	$t\bar{t}$	$Z + b$	$Z + c$	$Z + l$	sum MC	data
μ	120 ± 5	320 ± 20	71 ± 10	7 ± 3	520 ± 20	520 ± 20
e	30 ± 2	91 ± 10	9 ± 3	8 ± 3	138 ± 11	160 ± 13
$\mu\mu$	9.9 ± 1.4	17 ± 4	-	-	26 ± 5	27 ± 5
$e\mu$	6.6 ± 1.1	9 ± 3	-	-	16 ± 4	9 ± 3
ee	1.2 ± 0.5	1.2 ± 1.2	-	-	2.3 ± 1.2	2.0 ± 1.4

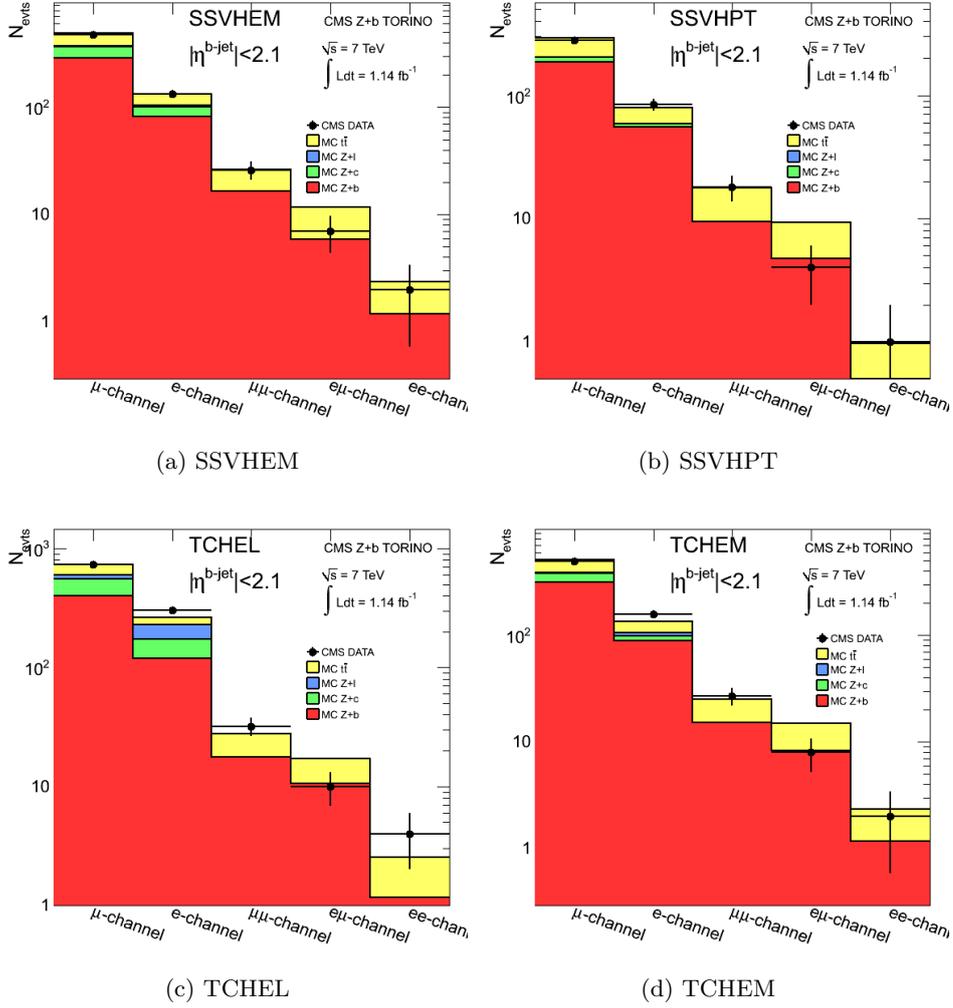


Figure 5.7: Multi-lepton events yields with $|\eta| < 2.1$ cut on b-jet for different choices of the b-tag discriminator.

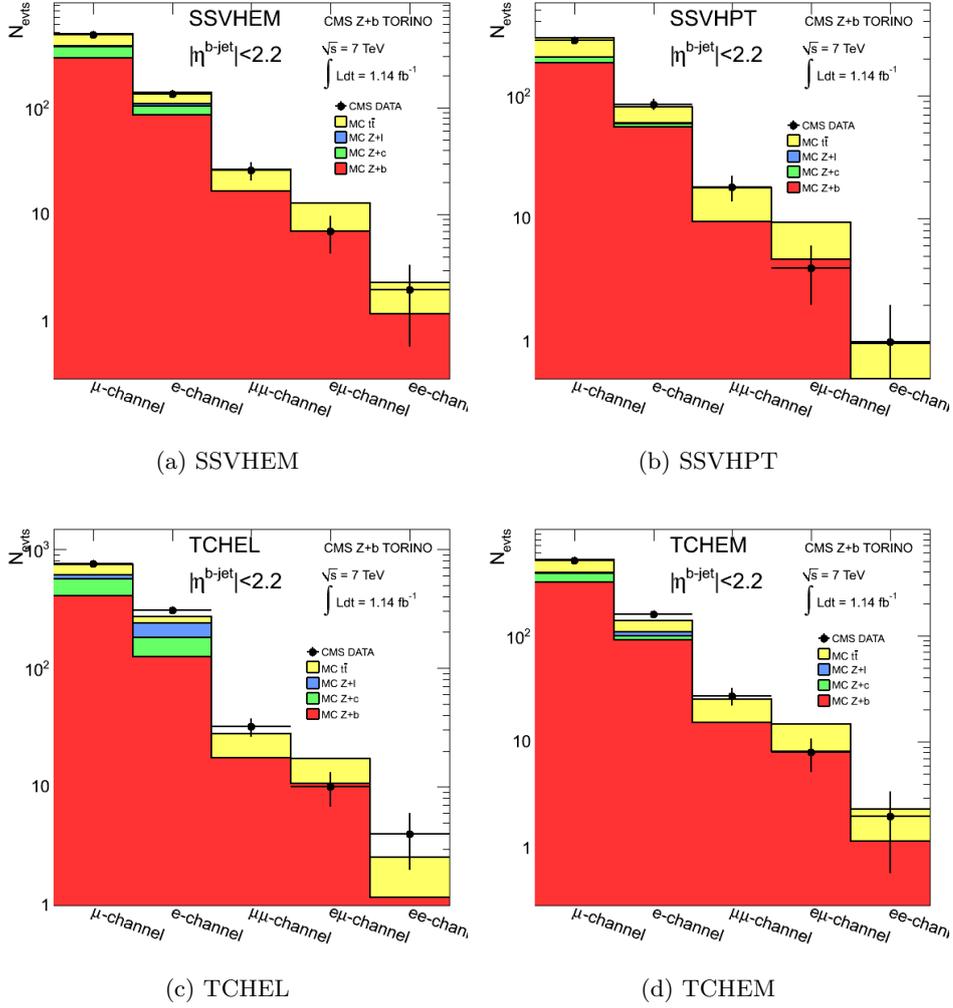


Figure 5.8: Multi-lepton events yields with $|\eta| < 2.2$ cut on b-jet for different choices of the b-tag discriminator.

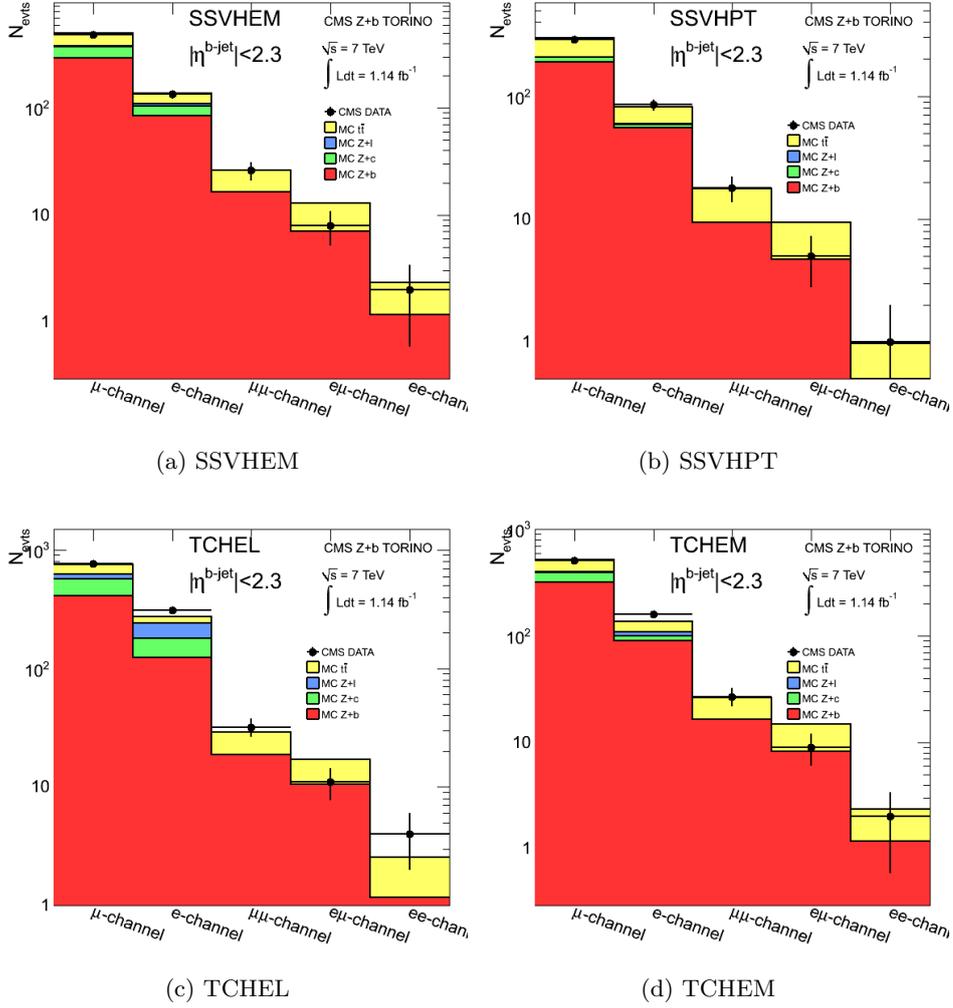


Figure 5.9: Multi-lepton events yields with $|\eta| < 2.3$ cut on b-jet for different choices of the b-tag discriminator.

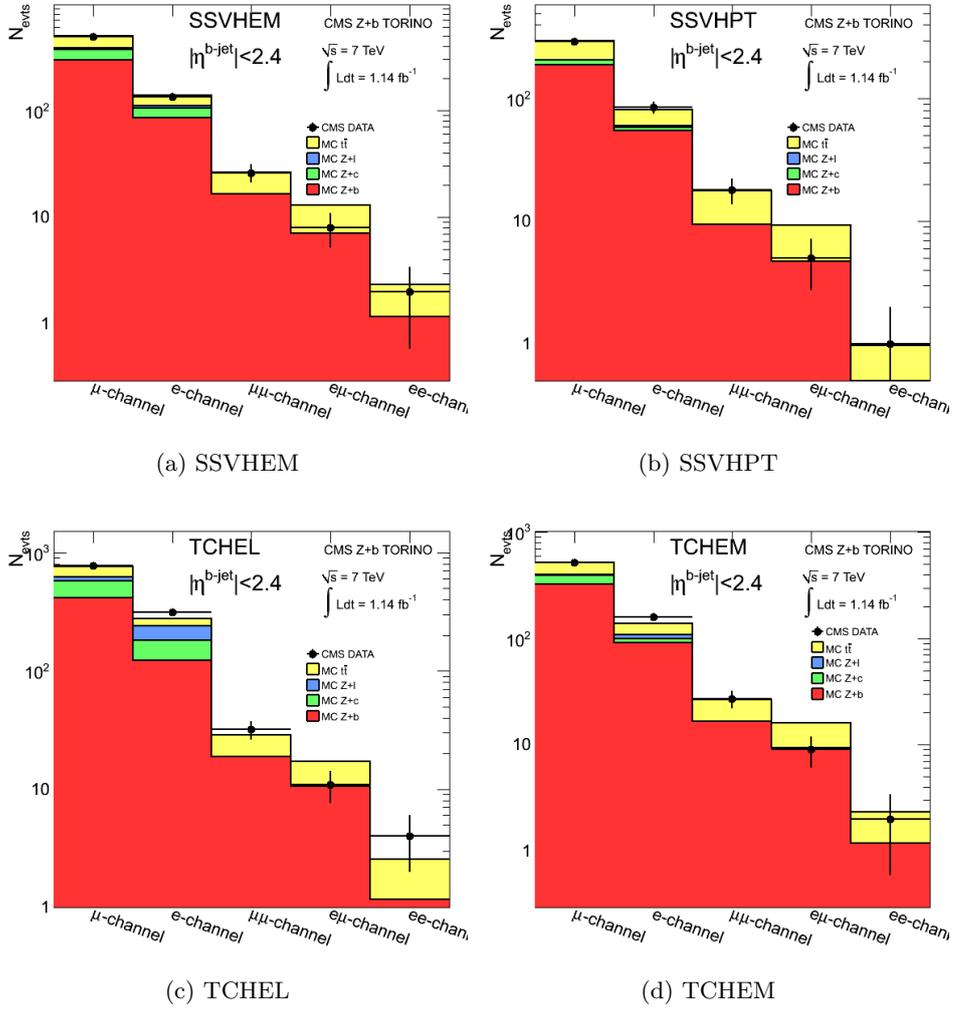


Figure 5.10: Multi-lepton events yields with $|\eta| < 2.4$ cut on b-jet for different choices of the b-tag discriminator.

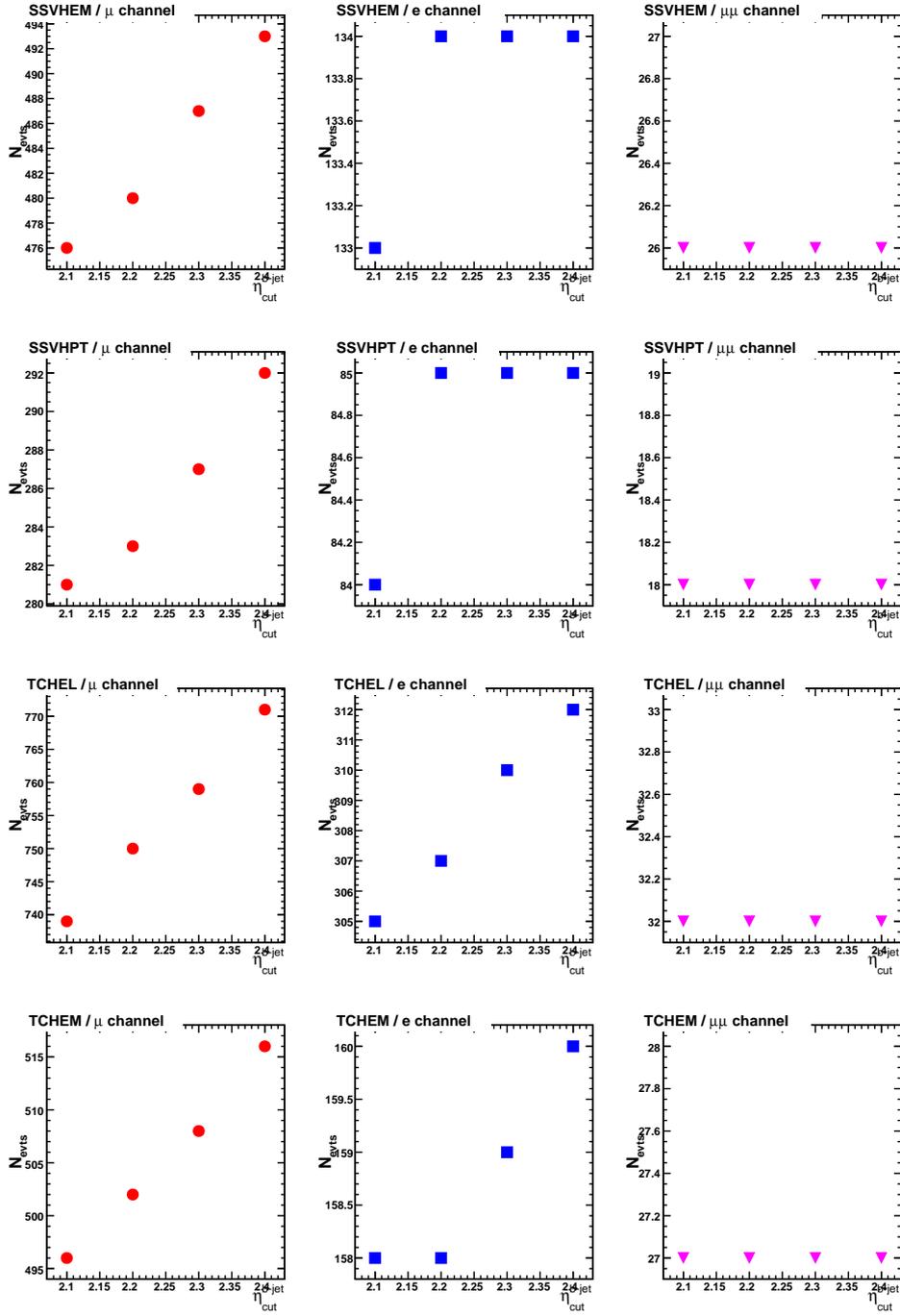


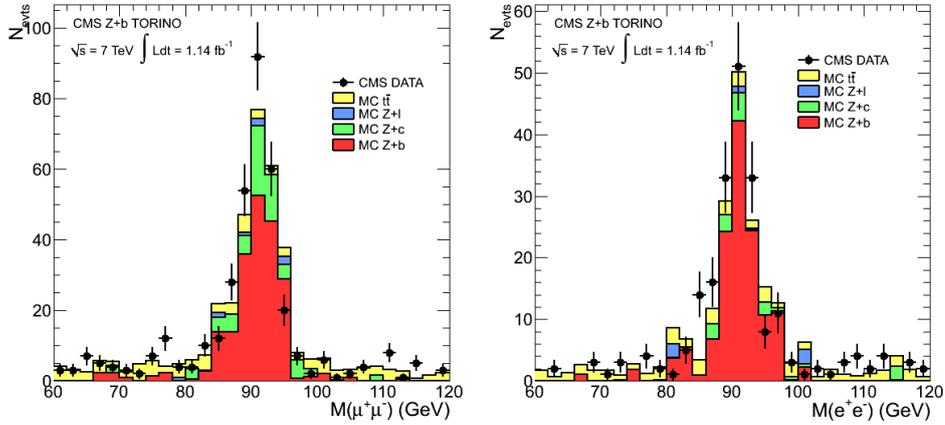
Figure 5.11: Multi-lepton events yields as a function of the η cut on the b-tagged jet, for different choices of the b-tag discriminator.

5.3 Control plots

As it was done for the $Z+b$ selection step in the previous chapter, here we present some control plots concerning the $Z+b$ multi-lepton candidate events. These plots refer to the use of the SSVHEM tagger and the cut on the b -tagged jet at $|\eta| < 2.1$. Even though statistics is not fully satisfactory in the MC samples, still there is, in general, a good agreement between data and simulation.

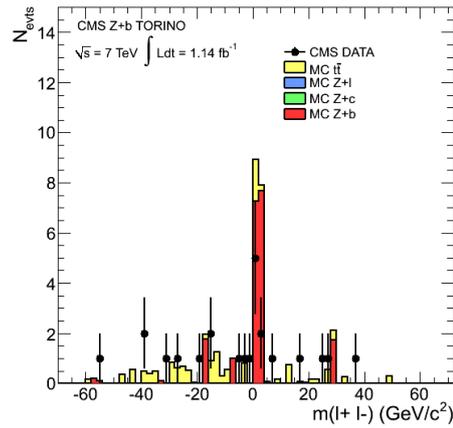
In figure 5.12 (c) the di-lepton mass is shown with the convention of representing positive values for the di-leptons built with combinations of leptons of the same flavor and opposite charge and negative masses for the other combinations. In the positive range of the x axis we can see the Z^* candidates which are reconstructed in $Z+b$ events with our selection. They seem to be concentrated in the lower mass spectrum and no di-lepton with mass above 40 GeV is reconstructed.

In figure 5.13 (a) and (c) the p_T spectra of all muons and electrons show two different contributions: above 20 GeV the leptons coming from Z decays dominate, below there are only the other leptons reconstructed in the event left.



(a) Invariant mass of $Z(\mu^+\mu^-)$ candidates for multi-lepton events.

(b) Invariant mass of $Z(e^+e^-)$ candidates for multi-lepton events.



(c) Invariant mass of additional di-leptons ($m > 0$ for opposite sign and same flavor dileptons, $m < 0$ for the other combinations).

Figure 5.12: Control plots of kinematic variables of dilepton pairs in events with more than two leptons.

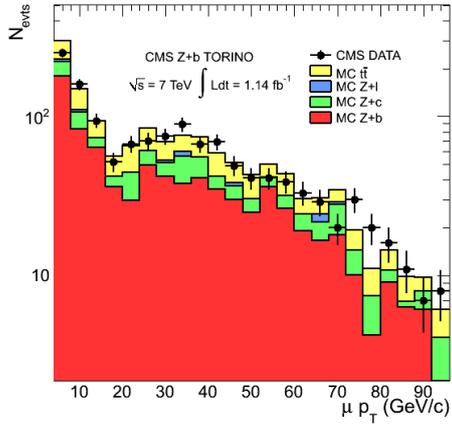
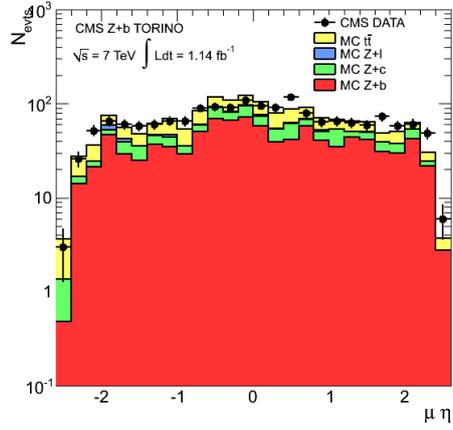
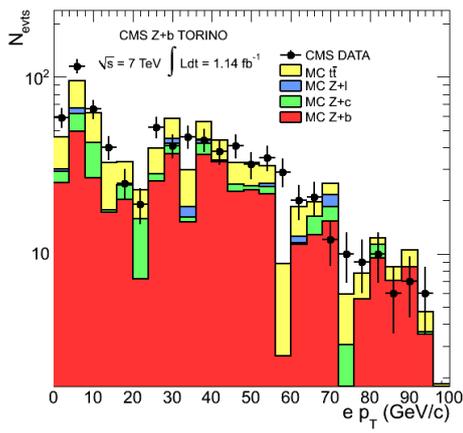
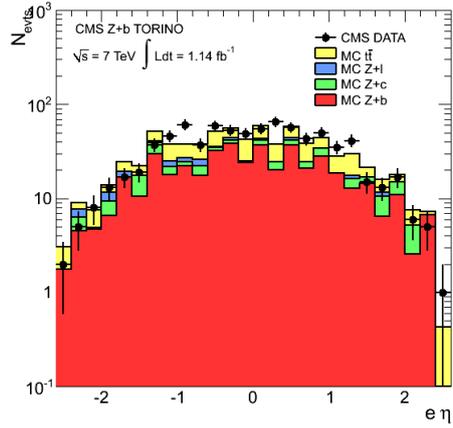
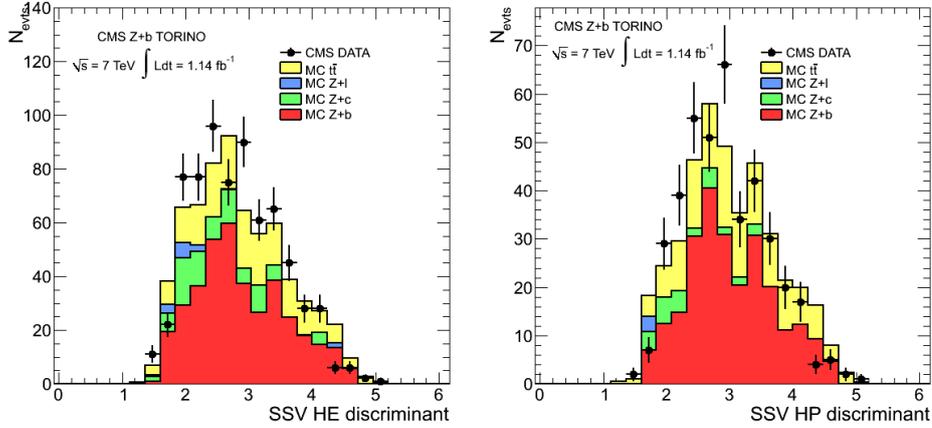
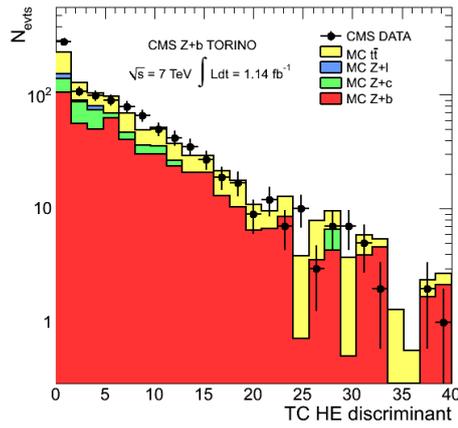
(a) p_T of all muons.(b) η of all muons.(c) p_T of all electrons.(d) η of all electrons.

Figure 5.13: Control plots of kinematic variables of the leptons in events with more than two leptons.

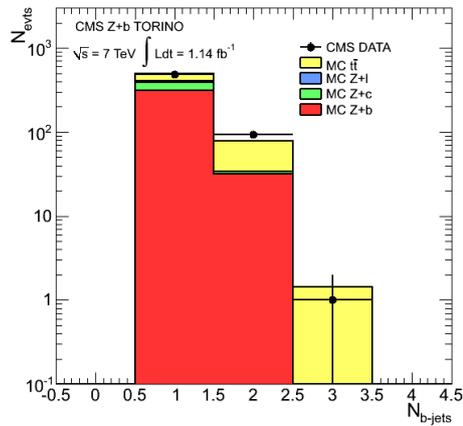


(a) SSVHE discriminator.

(b) SSVHP discriminator.



(c) TCHE discriminator.



(d) Number of b-tagged jets in the event, using the SSVHEM tagger.

Figure 5.14: Control plots of b-tagging related variables in events with more than two leptons.

5.4 Additional leptons kinematics

In this section, some plots of kinematic variables are shown for the additional leptons, referring to the Z+b selection with the SSVHEM tagger and $|\eta^{b-jet}| < 2.1$ (figure 5.15 for muons and figure 5.16 for electrons). As it was said in the previous section, the statistics of MC samples is limited and the plots show MC fluctuations which are purely statistical.

In figures 5.15 (f) and 5.16 (f) the p_T^{rel} of the lepton is defined as the momentum component in the transverse plane with respect to the closest-jet axis, i.e.:

$$p_T^{rel} = \frac{|\vec{p}_{lep} \times \vec{p}_{jet}|}{|\vec{p}_{jet}|}$$

The relative combined isolation (figures 5.15 (c) and 5.16 (c)) is defined as:

$$\text{RelCombIso} = \frac{\sum_{TRK} p_T + \sum_{ECAL} p_T + \sum_{HCAL} p_T}{p_T}$$

In figures 5.15 (e) and 5.16 (e) the minimum ΔR between the additional lepton and the jet axis is taken among all the combinations, where ΔR is defined as usual as $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$.

5.4.1 Muons

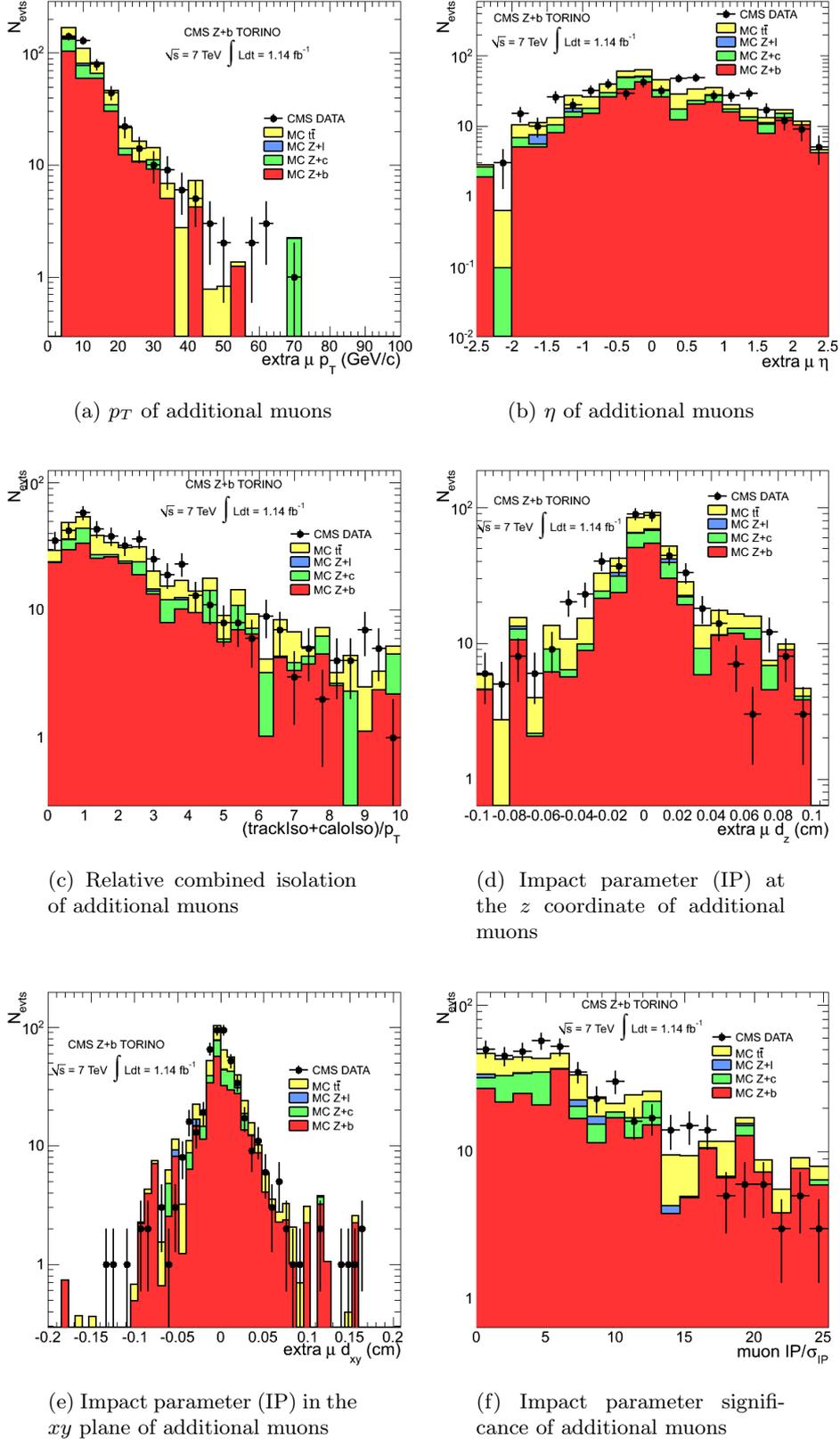


Figure 5.15: Plots of kinematic variables of additional muons.

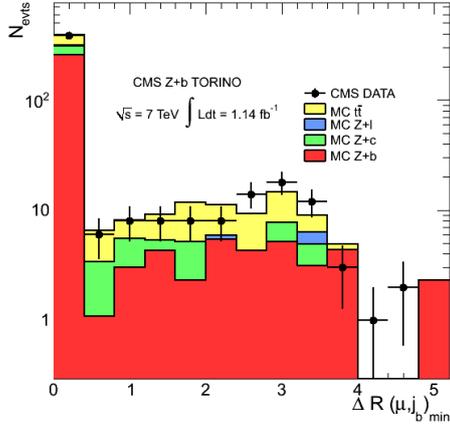
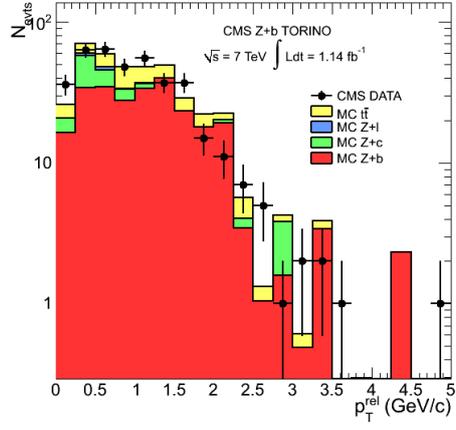
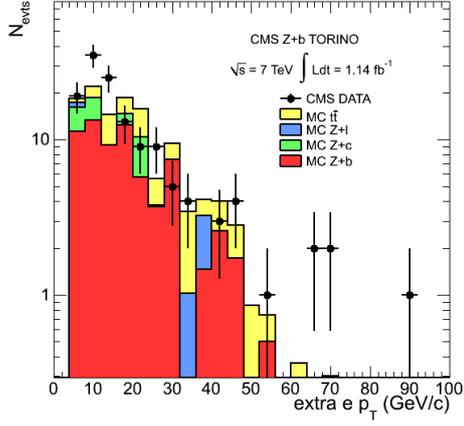
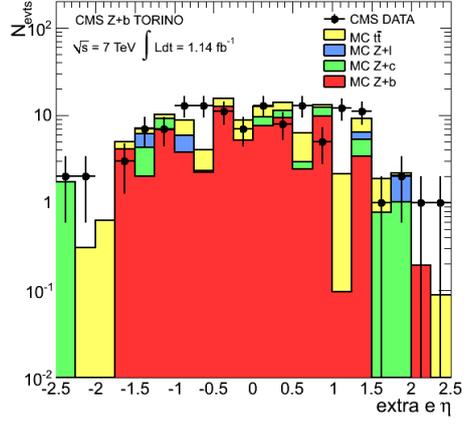
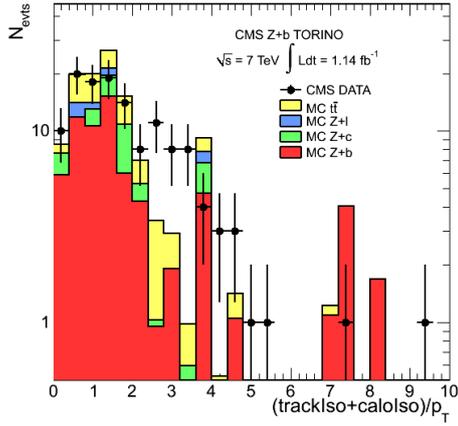
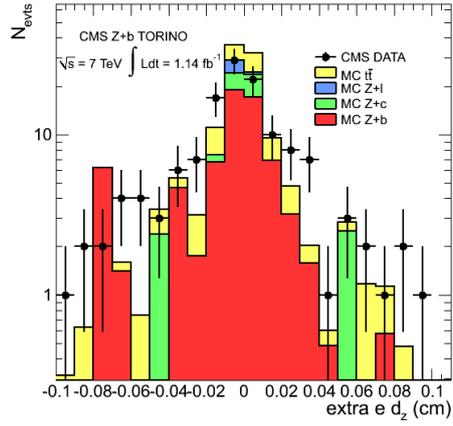
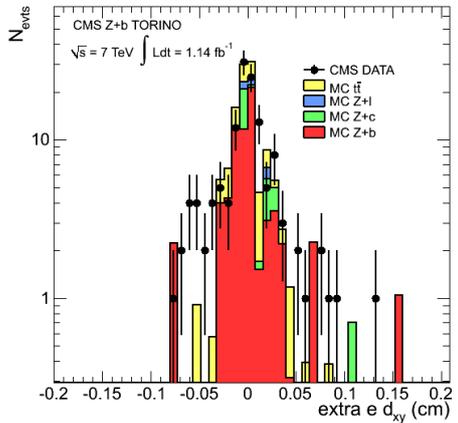
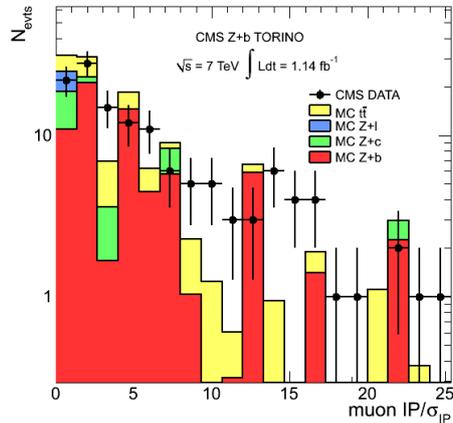
(g) Minimum ΔR between additional muon and b-tagged jets(h) p_T^{rel} of additional muons

Figure 5.15: Plots of kinematic variables of additional muons.

5.4.2 Electrons

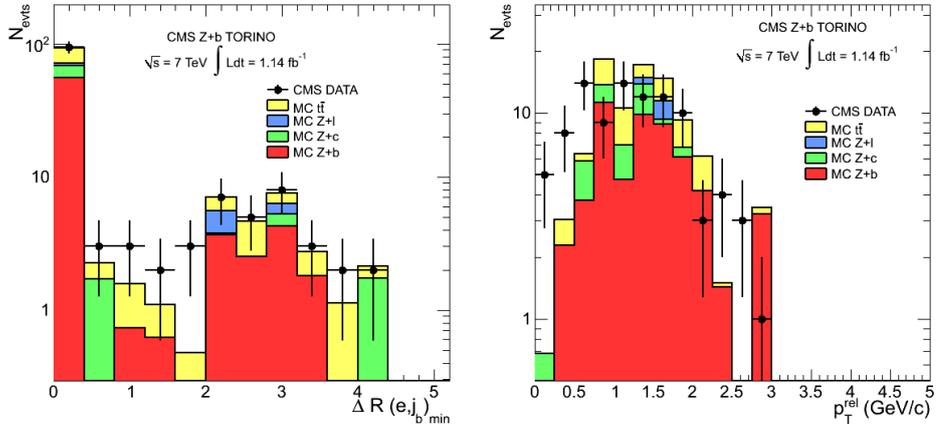
(i) p_T of additional electrons(j) η of additional electrons

(k) Relative combined isolation of additional electrons

(l) Impact parameter (IP) at the z coordinate of additional electrons(m) Impact parameter (IP) in the xy plane of additional electrons

(n) Impact parameter significance of additional electrons

Figure 5.16: Plots of kinematic variables of additional electrons.



(o) Minimum ΔR between additional electron and b-tagged jets

(p) p_T^{rel} of additional electrons

Figure 5.16: Plots of kinematic variables of additional electrons.

5.5 Data/MC comparison

The cross section of the $Z+b$ process can be computed as:

$$\sigma(Z+b) \times BR(Z \rightarrow l^+l^-) = \frac{N(l^+l^-+b) \times (\mathcal{P} - f_{t\bar{t}})}{\mathcal{A}_l \times \mathcal{C}_{parton} \times \epsilon_l \times \epsilon_b \times \mathcal{L}} \quad (5.2)$$

where:

- $N(l^+l^-+b)$ is the number of selected $Z+b$ events;
- \mathcal{P} is the *purity* defined as the number of tagged b-jets over the total number of tagged jets;
- $f_{t\bar{t}}$ is the fraction of $t\bar{t}$ events in the selected events;
- \mathcal{A}_l is a correction factor for the acceptance of the leptons;
- \mathcal{C}_{parton} is a correction factor to correct from the "hadron level" to the "parton level";
- ϵ_l is the lepton selection efficiency;
- ϵ_b is the b-tagging efficiency;
- \mathcal{L} is the integrated luminosity.

In equation (5.2) \mathcal{P} , $f_{t\bar{t}}$, ϵ_l and ϵ_b are extracted from the data in the official CMS analysis. Similar kind of data-driven studies should be carried out to measure $\sigma(Z+b) \times BR(Z \rightarrow l^+l^-) \times BR(b \rightarrow l)$. In this study we just check the agreement in the yields of events with additional leptons between selected events and predictions, defining the ratio:

$$R = \frac{N_{s,meas}}{N_{s,exp}} \quad (5.3)$$

where $N_{s,meas}$ is the number of events with extra leptons, which will be referred to as "signal", extracted from data and $N_{s,exp}$ the yield expected. This definition allows to neglect the correction factors in the denominator of equation (5.2), thus considerably simplifying our analysis. $N_{s,exp}$ is directly estimated from MC, i.e. it is the number of events from the $Z+b$ MC sample which pass the selection in a certain channel (the " $Z+b$ " column in tables 5.1 - 5.4).

A simplified way to get $N_{s,meas}$ is to estimate the number of background events from MC, i.e.:

$$N_{s,meas} = N_{data} - N_{bkg}^{MC} = N_{data} - N_{Zc}^{MC} - N_{Zl}^{MC} - N_{t\bar{t}}^{MC} \quad (5.4)$$

Equation (5.4) should be compared with the numerator of equation (5.2) to test our MC estimate of background. Using the last available values

Table 5.5: Comparison between standard (data-driven) background estimate and the MC background estimate used in the present analysis.

channel	\mathcal{P} (%)	$f_{t\bar{t}}$ (%)	$\mathcal{P} - f_{t\bar{t}}$ (%)	$1 - N_{bkg}^{MC}/N_{data}$ (%)
$Z(\mu\mu) + b$ (SSVHEM)	59.0	11.0	48.0	49.6
$Z(ee) + b$ (SSVHEM)	61.4	11.5	49.9	51.5
$Z(\mu\mu) + b$ (SSVHPT)	79.2	17.8	61.4	65.3
$Z(ee) + b$ (SSVHPT)	78.7	17.6	61.1	67.9

from the Z+b analysis working group in the CMS Collaboration for the purity factor \mathcal{P} and the fraction of $t\bar{t}$ events $f_{t\bar{t}}$, valid for the SSVHEM and SSVHPT taggers, the results of the comparison are those of table 5.5.

The MC-driven method seems to overestimate the correction factor, i.e. it underestimates the background. However, this cross check should be made at the selection level of multi-lepton events. In fact, it is likely that the presence of one or more additional leptons coming from b -quark semileptonic decays enhances the fraction of the signal in the selection. Hence we expect higher values of the purity at the multi-lepton selection level, while we expect $f_{t\bar{t}}$ not to be affected sensibly by this additional selection. So, in principle, as the parameters \mathcal{P} and $f_{t\bar{t}}$ can be different for our selection, also the agreement between data-driven and MC-driven estimate of the background could change. This comparison should be ultimately seen as a semi-qualitative validation of our background estimate, without any claim of rigor.

The values of the ratio R for the μ , e and $\mu\mu$ channels (defined as R_μ, R_e and $R_{\mu\mu}$ respectively) are listed in table 5.6. The $e\mu$ -channel and the ee channel are not considered in this study, because they have too few events as one can check in tables 5.1 - 5.4. In figures 5.17, 5.18 and 5.19 R_μ, R_e and $R_{\mu\mu}$ respectively are shown as a function of the η cut on the b-jet, for the different taggers.

We can see from table 5.6 that $|R_\mu - 1| < 0.04$ and $|R_{\mu\mu} - 1| < 0.2$ for every choice of the tagger. For R_e the considerations made about the yields in the e channel in section 5.2 are still valid so R_e is in general greater than 1, except for the choice of the SSVHEM tagger.

As the number of events increase, in general, enlarging the η cut on the b-tagged jet, we can check the stability of the R ratio with respect to the this cut by defining a "correlated error" for the points corresponding to the relaxed η cut at 2.2, 2.3 and 2.4. By labeling D_i, B_i, S_i the number of selected events in the data, the number of background events and the number of signal events in the Monte Carlo for a certain value i of the η cut on the

b-tagged jet, we can rewrite R as:

$$R_i = \frac{D_i - B_i}{S_i} = \frac{D_0 + \Delta D - (B_0 + \Delta B)}{S_0 + \Delta S}$$

where D_0, B_0 and S_0 are the reference values at $|\eta_{cut}^{b-jet}| < 2.1$ and $\Delta D = D_i - D_0$, $\Delta B = B_i - B_0$, $\Delta S = S_i - S_0$.

The definition of correlated error takes into account only the uncertainty on the additional events, with respect to the reference point, at a certain η cut, i.e. the errors on $\Delta D, \Delta B, \Delta S$. They are, in a poissonian approximation $\sqrt{\Delta D}, \sqrt{\Delta B}, \sqrt{\Delta S}$ respectively, so the correlated error is:

$$\delta_c R_i = \frac{1}{S_i} \sqrt{\Delta D} \oplus \frac{1}{S_i} \sqrt{\Delta B} \oplus \frac{R_i}{S_i} \sqrt{\Delta S} \simeq \frac{\sqrt{2}}{S_i} \sqrt{\Delta D}$$

where the last formula uses the approximation $\Delta B + R_i^2 \Delta S \simeq \Delta B + \Delta S \simeq \Delta D$.

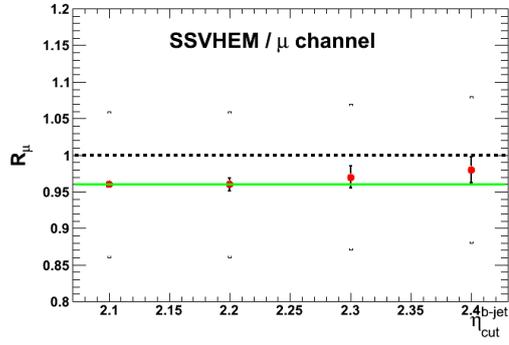
The correlated error reduces to zero anytime no additional event is selected enlarging the η cut, neither in data and MC, i.e. when $\Delta D = \Delta B = \Delta S = 0$. This is never the case for the μ channel, but it is often the case in the $\mu\mu$ channel (see figure 5.11).

In figures 5.17, 5.18 and 5.19 the uncorrelated error is shown as marks without bars, while the correlated one is drawn as error bars. The latter should be directly compared with the value at $|\eta_{cut}^{b-jet}| < 2.1$, which is shown as green horizontal line, in order to check the stability of R with respect to η_{cut}^{b-jet} . The uncorrelated and correlated errors are listed in this order also in table 5.6.

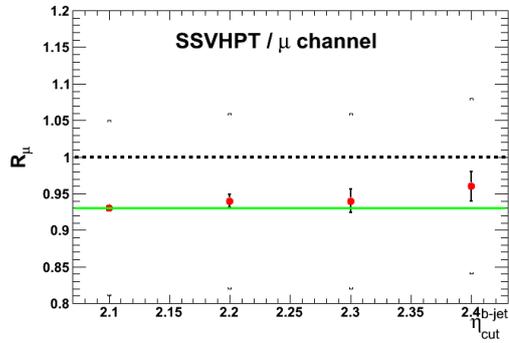
We can deduce that R is almost always stable, except for few points at $|\eta_{cut}^{b-jet}| < 2.4$.

Table 5.6: R_μ, R_e and $R_{\mu\mu}$ for different taggers and different values of the η cut on the b-tagged jet.

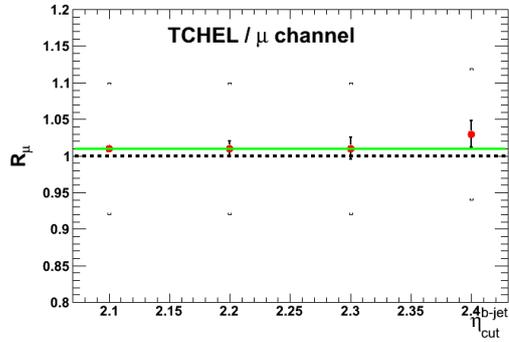
	SSVHEM			
	$ \eta^{b-jet} < 2.1$	$ \eta^{b-jet} < 2.2$	$ \eta^{b-jet} < 2.3$	$ \eta^{b-jet} < 2.4$
R_μ	0.96 ± 0.10	$0.96 \pm 0.10 \pm 0.01$	$0.97 \pm 0.10 \pm 0.02$	$0.98 \pm 0.10 \pm 0.02$
R_e	1.01 ± 0.20	$0.96 \pm 0.19 \pm 0.03$	$0.96 \pm 0.19 \pm 0.03$	$0.94 \pm 0.19 \pm 0.03$
$R_{\mu\mu}$	1.0 ± 0.4	1.0 ± 0.4	1.0 ± 0.4	1.0 ± 0.4
	SSVHPT			
	$ \eta^{b-jet} < 2.1$	$ \eta^{b-jet} < 2.2$	$ \eta^{b-jet} < 2.3$	$ \eta^{b-jet} < 2.4$
R_μ	0.93 ± 0.12	$0.94 \pm 0.12 \pm 0.01$	$0.94 \pm 0.12 \pm 0.02$	$0.96 \pm 0.12 \pm 0.02$
R_e	1.08 ± 0.23	$1.07 \pm 0.23 \pm 0.03$	$1.07 \pm 0.23 \pm 0.03$	$1.07 \pm 0.23 \pm 0.03$
$R_{\mu\mu}$	1.0 ± 0.6	1.0 ± 0.6	1.0 ± 0.6	1.0 ± 0.6
	TCHEL			
	$ \eta^{b-jet} < 2.1$	$ \eta^{b-jet} < 2.2$	$ \eta^{b-jet} < 2.3$	$ \eta^{b-jet} < 2.4$
R_μ	1.01 ± 0.09	$1.01 \pm 0.09 \pm 0.01$	$1.01 \pm 0.09 \pm 0.02$	$1.03 \pm 0.09 \pm 0.02$
R_e	1.34 ± 0.22	$1.28 \pm 0.21 \pm 0.03$	$1.29 \pm 0.21 \pm 0.03$	$1.30 \pm 0.21 \pm 0.04$
$R_{\mu\mu}$	1.2 ± 0.5	1.2 ± 0.5	$1.2 \pm 0.4 \pm 0.07$	$1.2 \pm 0.4 \pm 0.07$
	TCHEM			
	$ \eta^{b-jet} < 2.1$	$ \eta^{b-jet} < 2.2$	$ \eta^{b-jet} < 2.3$	$ \eta^{b-jet} < 2.4$
R_μ	0.97 ± 0.10	$0.97 \pm 0.10 \pm 0.01$	$0.97 \pm 0.10 \pm 0.02$	$0.98 \pm 0.10 \pm 0.02$
R_e	1.26 ± 0.22	$1.22 \pm 0.21 \pm 0.02$	$1.23 \pm 0.21 \pm 0.02$	$1.24 \pm 0.21 \pm 0.03$
$R_{\mu\mu}$	1.1 ± 0.5	1.1 ± 0.5	$1.0 \pm 0.4 \pm 0.07$	$1.0 \pm 0.4 \pm 0.07$



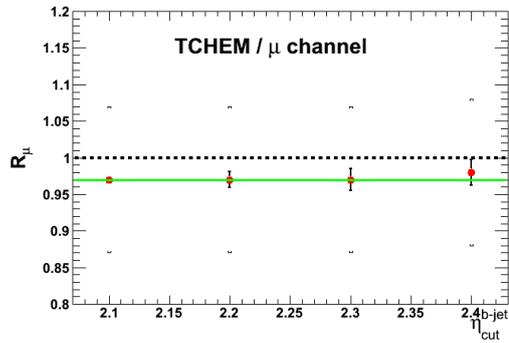
(a) SSVHEM



(b) SSVHPT

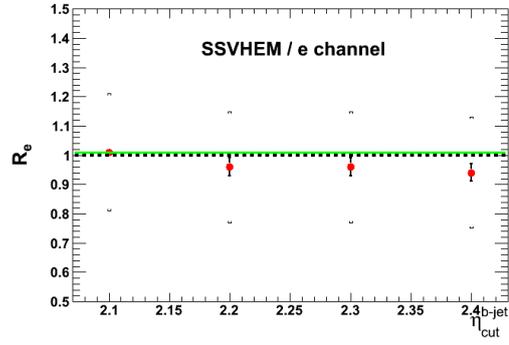


(c) TCHEL

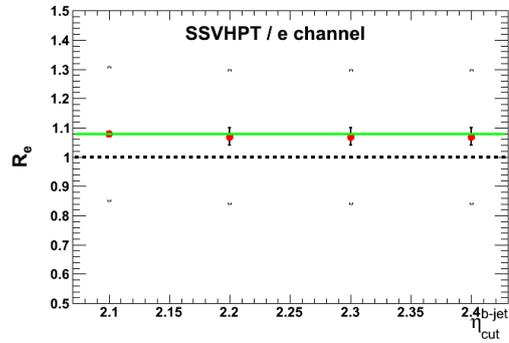


(d) TCHEM

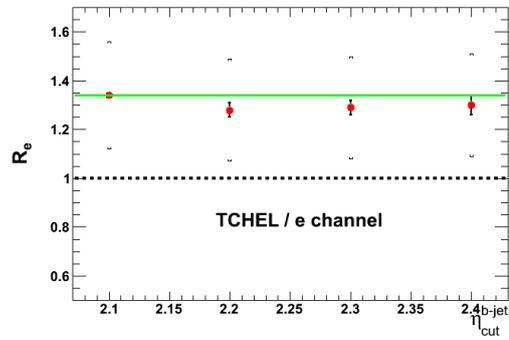
Figure 5.17: R_μ as a function of the η cut on the b-jet for different taggers.



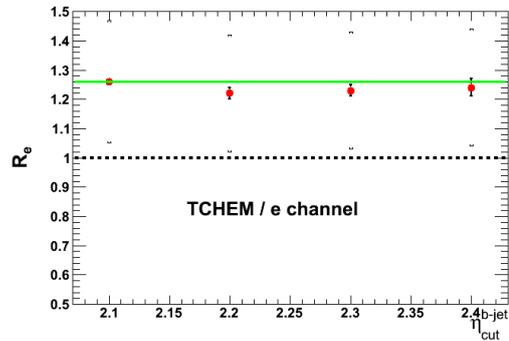
(a) SSVHEM



(b) SSVHPT

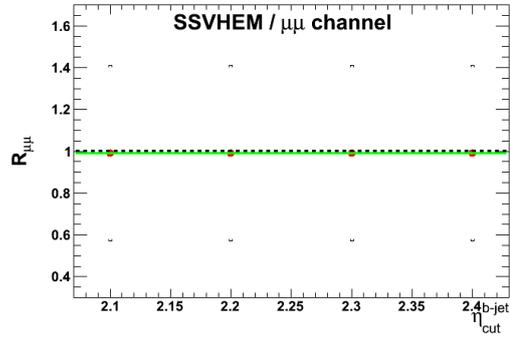


(c) TCHEL

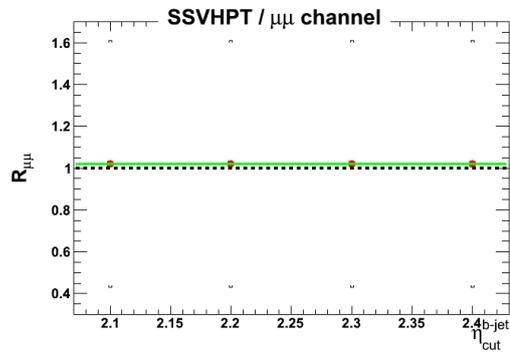


(d) TCHEM

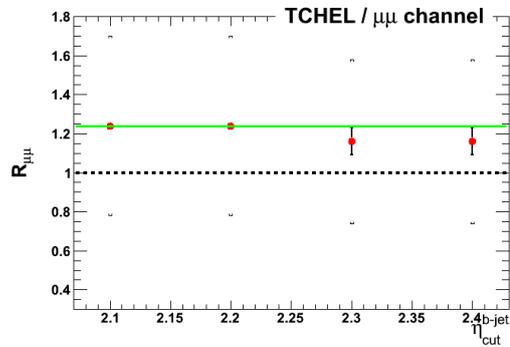
Figure 5.18: R_e as a function of the η cut on the b-jet for different taggers.



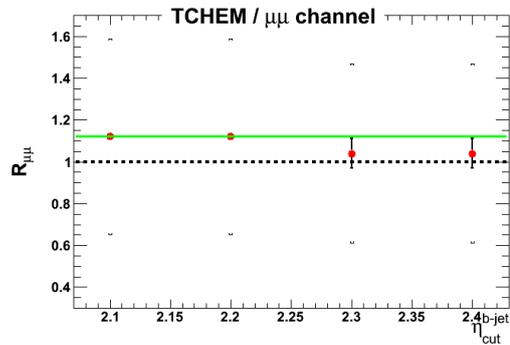
(a) SSVHEM



(b) SSVHPT



(c) TCHEL



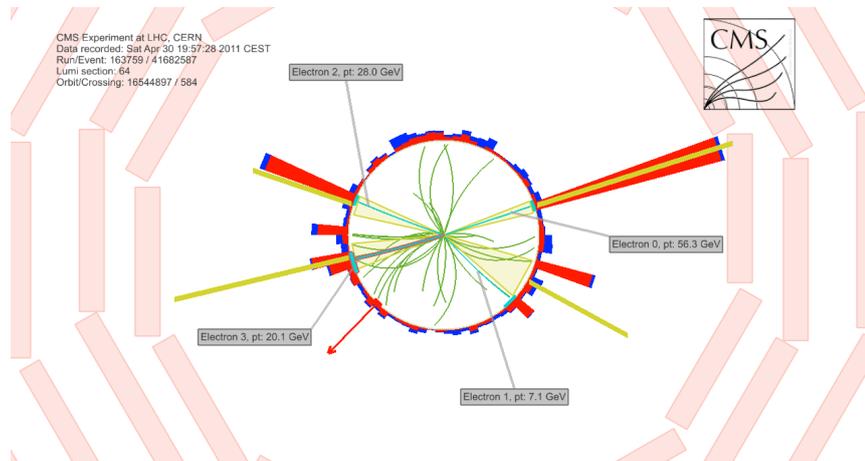
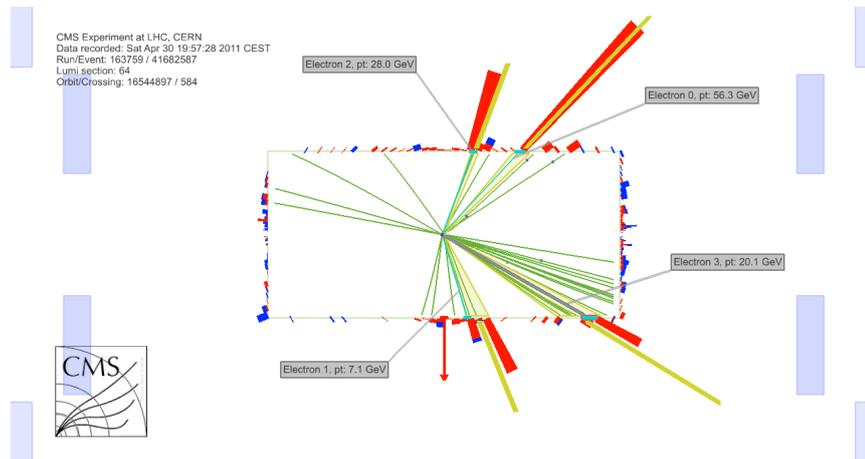
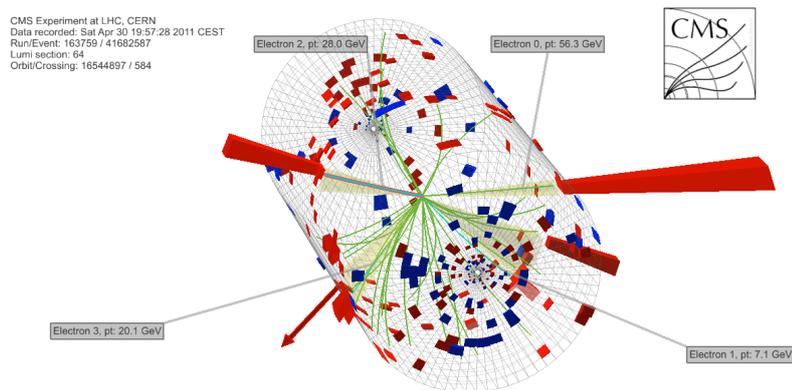
(d) TCHEM

Figure 5.19: $R_{\mu\mu}$ as a function of the η cut on the b-jet for different taggers.

5.6 Event display

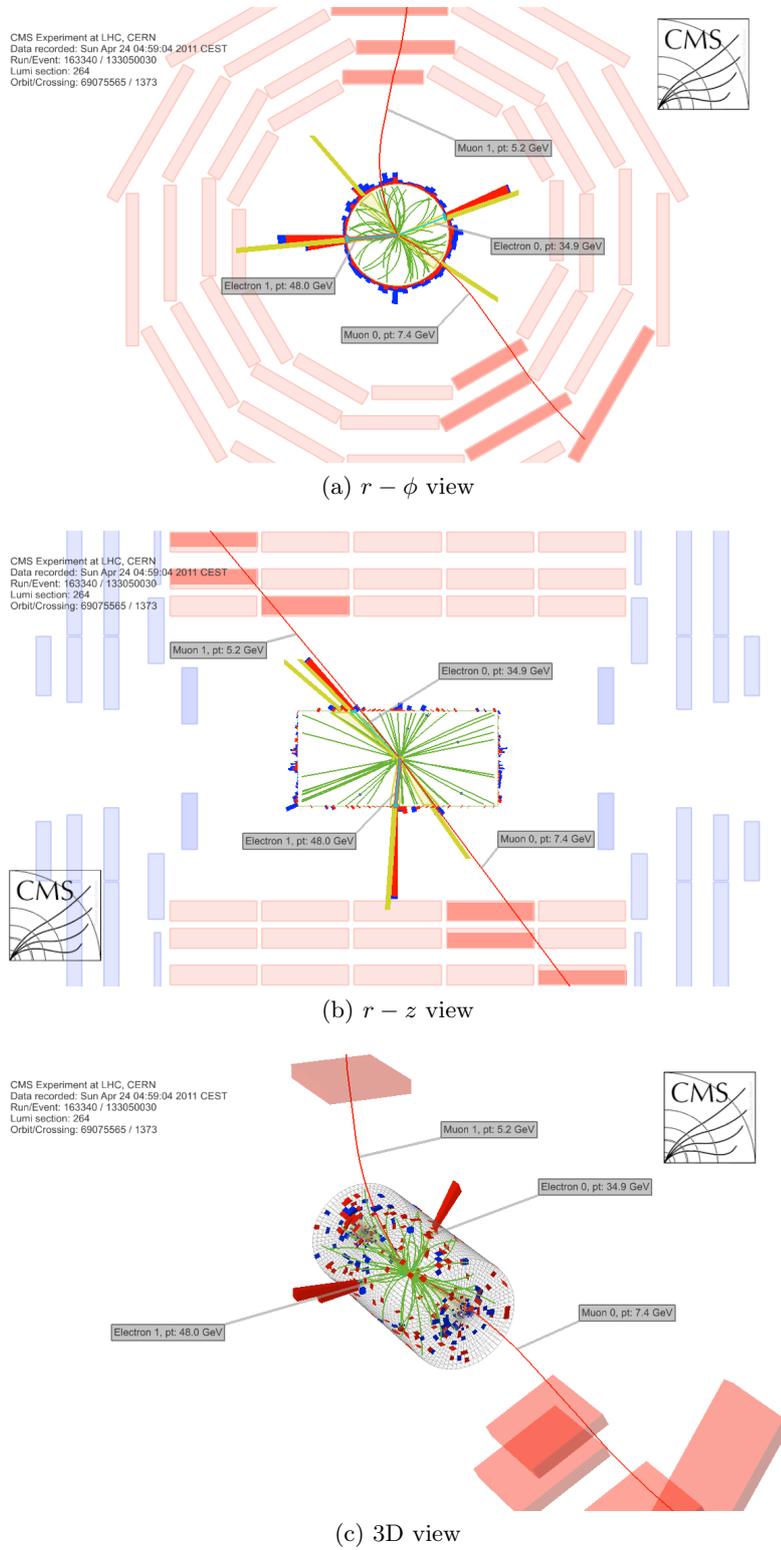
In the following pages, from figure 5.20 to 5.22 we show some multi-lepton candidate events chosen among the ones with 2 additional leptons, with the Z reconstructed either as $Z(e^+e^-)$ or $Z(\mu^+\mu^-)$.

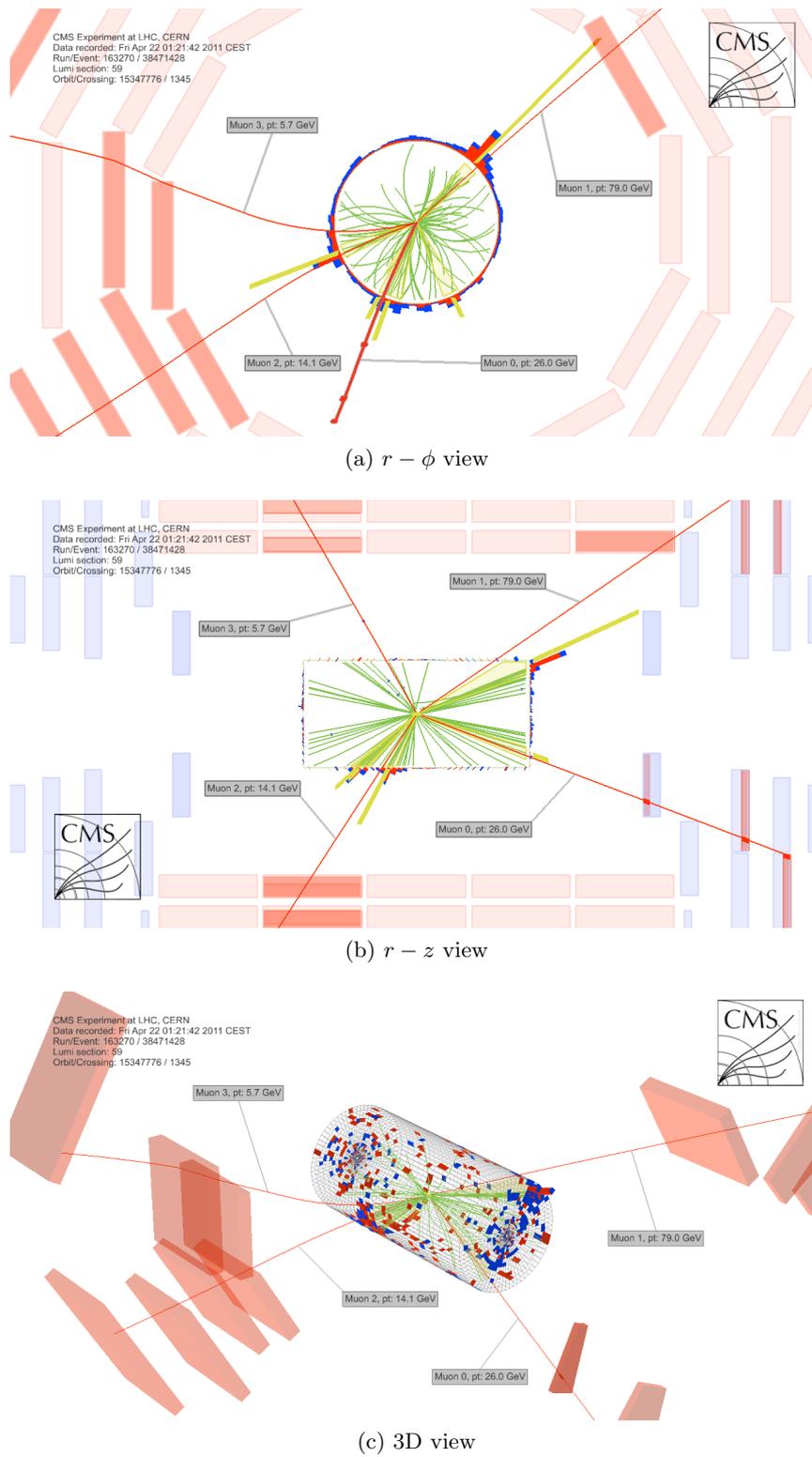
The same color conventions of section 4.4 are used. In all the events which are shown the Z is reconstructed with the two leptons with highest p_T , which are always the only ones passing the dilepton selection (see 4.2.1).

(a) $r - \phi$ view(b) $r - z$ view

(c) 3D view

Figure 5.20: A $Z(e^+e^-) + 2e$ candidate event.

Figure 5.21: A $Z(e^+e^-) + 2\mu$ candidate event.

Figure 5.22: A $Z(\mu^+\mu^-) + 2\mu$ candidate event.

Conclusions

The work described in this thesis has been carried out within the CMS group in Torino working in the analysis for the measurement of the cross section of the production of a Z boson in association with one or more b -tagged jets with the CMS detector. The main subject of this thesis is the study of $Z+b$ events with additional leptons (muons and electrons) besides those coming from the decay of the Z . Such events are background to the search of the Standard Model Higgs boson decaying into a $ZZ^{(*)}$ pair, in turn decaying into four leptons. As the most important source of additional leptons in $Z+b$ events is the semileptonic decay of the b -quark, a preliminary study of the $Z+b$ simulated sample was done in order to calibrate a selection of leptons coming from b -quark semileptonic decays. This led to the choice of the existing standard selections of leptons coming from the decays of vector bosons with looser cuts on isolation and transverse momentum. With this selection some hundreds of these "multi-lepton" events were selected with 1.14 fb^{-1} of integrated luminosity, some tens containing two additional leptons. The yields of events selected in the data were compared to the predictions with a Monte Carlo driven estimation of the background. They are found to be in good agreement, especially for events containing additional muons. The agreement is worse when considering electrons, nevertheless this is likely due to the imperfect simulation of the detector response.

Appendices

Appendix A

Factorization theorem and hadronic cross sections

The *parton model* together with the *quark model* led to the possibility of interpreting the results of hadronic high energy collisions in terms of interactions between the quarks, the elementary constituents of hadrons.

The quark model, introduced by Gell-Mann to explain the hadronic "zoo" of particles, foresees the compositeness of hadrons on the basis of a flavor symmetry. The parton model, introduced by Feynman to account for the results coming from the deep inelastic scattering of electrons over nucleons, basically states that when hadrons interact with high momentum transfer they seem to be made of point-like, quasi-free, fermions: the partons. These are the quarks and gluons of QCD. A parton distribution function (PDF) $\mathcal{F}_i^h(x)$ is associated to each parton, being the probability for the parton i to have a fraction x of the momentum of the hadron h .

QCD factorization theorems rigorously disentangle the non-perturbative properties of hadrons and the asymptotically free behavior of partons. Any hadronic observable can be written as:

$$\mathcal{O}_A(Q^2) = \sum_i f_A^i(Q^2, \mu_f) \otimes \mathcal{O}_i(Q^2, \mu_f)$$

where Q^2 is the square of momentum transfer, the \otimes is an integral convolution, μ_f is the *factorization scale* and the sum is over all contributing partons. The functions \mathcal{O}_i describe the short distance parton interaction, and if Q^2 is large can be calculated perturbatively. The functions f_A^i contains all the non-perturbative (i.e. long distance) informations on the hadronic state A , and are universal, hence independent of the process.

The factorization scale μ_f is introduced as a boundary from the perturbative and non-perturbative regime. In principle, \mathcal{O}_A does not depend on μ_f , but this is true only when all orders of the perturbative expansion are considered. Being the perturbative series always truncated, a residual dependence remains which is of the order the next term in the perturbative expansion. For this reason, μ_f is often cited about the uncertainties in the theoretical

calculations.

As a useful application of the factorization theorems, the cross section for any hadronic process can be written as:

$$\sigma(h_1 h_2 \rightarrow X) = \sum_{ij} \frac{1}{1 + \delta_{ij}} \int dx_1 dx_2 \left[\mathcal{F}_i^{h_1}(x_1, \mu) \mathcal{F}_j^{h_2}(x_2, \mu) \hat{\sigma}_{ij}(x_1, x_2, \mu) + (x_1 \leftrightarrow x_2) \right]$$

where h_1, h_2 are generic hadrons, X a generic final state and the sum runs over all contributing partonic processes initiated by partons i and j , whose cross section is $\hat{\sigma}_{ij}$. The symmetric factor placed in front of the integral is to take into account the presence of identical particles in the initial state.

Appendix B

B-tagging in CMS

Jets originated from B-hadrons can be identified thanks to the long lifetime of B-hadrons and the relatively hard fragmentation of the b-quark. In CMS several algorithms have been developed, each one defining a variable, called discriminator, exploiting specific properties of the B-hadron decay. Given the value of the discriminator, a probability that the jet is originated from B-hadrons is calculated. Here we describe two b-tagging algorithms, which are used in this analysis.

The track counting algorithm

The *track counting* (TC) algorithm belongs to the family of algorithms that use the information of the impact parameter of tracks. The impact parameter (IP) of a track is quantified by the distance of the track trajectory to the primary vertex (V) as illustrated in the figure 23. The closest point of approach of the track to the jet direction is extracted. From this point (S) the tangent of the track is determined and the impact parameter is the distance of the primary vertex to the extrapolated tangent of the track. It

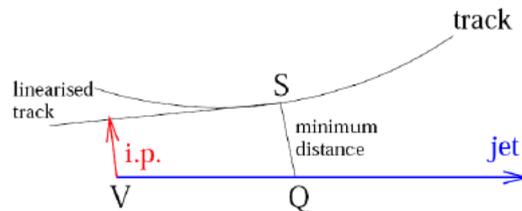


Figure 23: Representation of the impact parameter (IP) of a track.

is given a positive sign if the angle between the impact parameter and the jet direction is smaller than 90° , negative if it's larger than 90° .

To take into account the experimental resolution, the impact parameter significance is used, which is defined as the IP divided by its uncertainty.

The discriminator for the TC algorithm is the three dimensional IP significance of the n -th track, the tracks being ordered in descending significance. Different choice of n can be used, defining two versions of the discriminator: $n=2$ corresponds to high efficiency (HE) for selecting bottom jets, $n=3$ high purity (HP). In figure 24 the TC discriminator distributions are shown for $t\bar{t}$ events.

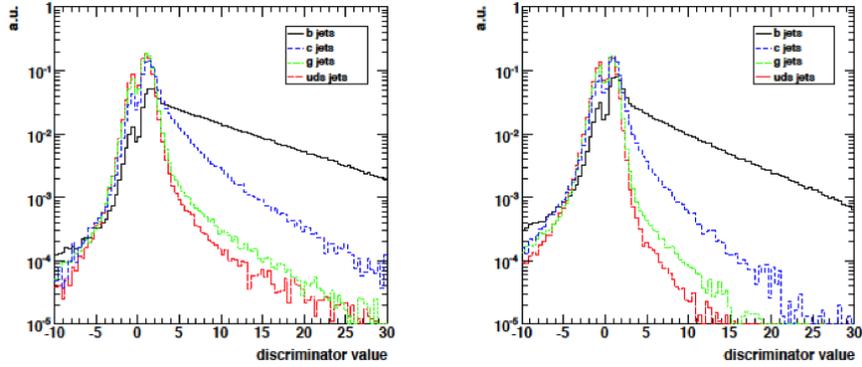


Figure 24: TC discriminator for HE (left) and HP (right) working points, for $t\bar{t}$ MC events.

The simple secondary vertex algorithm

Thanks to the good track reconstruction efficiency in CMS, displaced vertices can be identified. The *simple secondary vertex* (SSV) algorithm uses the information of a second vertex reconstructed in the event. The flight distance D is the distance between the reconstructed primary vertex and the secondary vertex, calculated in the transverse plane or in three dimensions. If the number of tracks attached to the vertex is larger than 2 (3), a ?high efficiency? SSVHE (?high purity? SSVHP) tagger is defined, respectively. The discriminator is defined as $\log\left(1 + \frac{D}{\sigma_D}\right)$ and is shown for the three-dimensional flight distance in figure 25. This SSV algorithm, although using

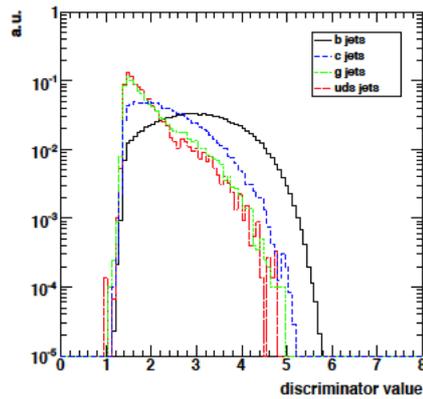


Figure 25: SSV discriminator for $t\bar{t}$ MC events.

high level objects like secondary vertices, is rather robust against possible

mis-alignment of the tracking system and that's why is often favorable for analysis.

Appendix C

The anti- k_T jet clustering algorithm

In high energy physics jet algorithms allow to establish a link between the streams of hadrons observed in the detector (jets) with the particles involved in the hard scattering (partons). The ingredients of any jet algorithm are:

- a list of input objects for the algorithm;
- a *distance* d_{ij} , defined between objects i and j ;
- a procedure to decide whether objects are to be combined or not;
- a procedure, called *recombination scheme*, which decides how the objects have to be combined.

There are two most commonly used recombination scheme: the E -scheme, which simply adds the four-momenta of the entities and produces massive jets; the E_T -scheme produces massless jets by equating the jet transverse momentum to the sum of the transverse momentum of every constituent ($\sum E_T$) and fixing the direction of the jet. The latter step can be done in two ways: $\sin\theta = \sum E_T/E$, where E is the jet energy or

$$\eta = \frac{\sum E_{T,i}\eta_i}{\sum E_T}, \quad \phi = \frac{\sum E_{T,i}\phi_i}{\sum E_T}$$

Jet algorithms are often divided in two broad categories: the cone algorithms, which cluster objects close in angle to an energetic seed and the clustering algorithm where objects with the smallest distance are iteratively combined.

Three algorithms are implemented in CMS:

- *Iterative cone*
A list of E_T ordered objects is the input. A cone in $\eta - \phi$, with size R , is built around the particle with largest transverse momentum, above a certain threshold ($E_{T,thr}$). Particles inside the cone are used to build a "proto-jet", using the E_T -scheme, with a new direction and energy. A new cone is built around the proto-jet axis, and the same procedure iteratively applies until the iteration corrects the energy for less than 1% and the direction for less than $\Delta R = 0.01$. Now the stable proto-jet is added to the list of jets, and all of its components are removed from the input list. The algorithm runs again and again until no objects above the threshold remains. Thus R and $E_{T,thr}$ are the relevant parameters for this algorithm .
- *Midpoint cone*
In contrast to the iterative cone described above, here no object is

removed from the input list, resulting in the possibility for proto-jets to overlap, i.e. a single object can belong to different proto-jets. For pair of stable proto-jets which are closer than the cone radius R , a midpoint is set with the combined four-momenta and it is used as seed in a new iteration over the list of proto-jets. With the resulting final list, the splitting and merging procedure is applied, starting from the largest E_T proto-jet. If it does not share objects with other proto-jets, it is defined as a jet and removed from the list. Otherwise, the transverse energy shared with the highest E_T neighbour proto-jet is compared to the total transverse energy of this neighbour. If the fraction is greater than f (typically 50%) the two proto-jets are merged, otherwise the shared objects are individually assigned to the proto-jet that is closest in $\eta - \phi$ space. The procedure goes on until no proto-jet is left. This algorithm uses the E_T -scheme to define the property of the proto-jets. R , $E_{T,thr}$, f and also a maximum number of proto-jets in the list used to calculate the midpoints are the parameters.

- *Inclusive k_T*

This is a clustering algorithm. For each object i and each pair ij of the input list, the following quantities are calculated:

$$d_i = k_{T,i}^2, \quad d_{ij} = \min(k_{T,i}^2; k_{T,j}^2) \frac{\Delta_{ij}^2}{R^2}$$

where R is the radius parameter, normally set to unity, $k_{T,i}$ is the transverse momentum of the object i and:

$$\Delta_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$$

where η_i and ϕ_i are the pseudo-rapidity and the azimuthal angle of the object i .

When the algorithm finds the smallest d_{ij} , objects i and j are removed from the list and replaced with one object accordingly to the recombination scheme used. When the algorithm finds the smallest d_i , the object i is removed from the list and added to the list of final jets. The procedure continues until every object is included in jets. Finally, objects are merged if $R_{ij} < R$, resulting in a list in which every jet has a distance larger than R from every other jet.

The iterative cone algorithm is faster and it is often used for software-based triggers, while the midpoint cone and the inclusive k_T are widely used in the offline analysis.

The inclusive k_T clustering algorithm can be seen as belonging to a broader class of algorithms, in which the distances are generalized as:

$$d_i = k_{T,i}^{2p}, \quad d_{ij} = \min(k_{T,i}^{2p}; k_{T,j}^{2p}) \frac{\Delta_{ij}^2}{R^2}$$

One recovers the inclusive k_T algorithm by replacing $p = 1$. For $p = 0$ one obtains the *Cambridge/Aachen* algorithm (see [18]). The "*anti- k_T* " *jet clustering* algorithm is defined by setting $p = -1$.

It's easy to see that from values of $p > 0$ to values of $p < 0$ the ordering of the particles changes, and this leads to different behaviors especially in jet boundaries with respect to soft radiation. We don't go further in technical details and address to [19] as a reference paper.

Appendix D

Event reweighting

In the 2011 Z+b analysis MC events are reweighted in order to match the pile-up (PU), the b-tag efficiency and the lepton efficiency with those observed in the data.

PU weight

In the MC samples, PU is simulated with a flat distributions of the number of interactions between 0-10 with a poissonian tail above 10 (see figure 26).

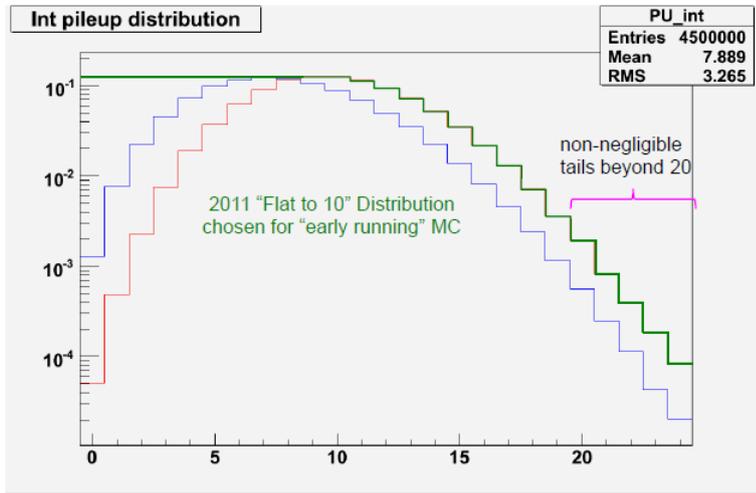


Figure 26: Distributions of PU interactions for 2011 MC samples (green). The red histogram is a Poisson distribution with a mean of 10 interactions. The other histogram (blue) in the plot is a distribution that would be obtained with a peak luminosity corresponding to 10 interactions per crossing, decreasing linearly to 5 interactions per crossing during a fill.

The PU weight is defined as :

$$w(n_{PU}^{MC}) = \frac{P^{data}(n_{PU})}{P^{MC}(n_{PU})}$$

where n_{PU}^{MC} is the number of PU events in the MC event to be weighted, and $P^{data}(n_{PU})$ and $P^{MC}(n_{PU})$ are the probabilities of having n_{PU} events as they are estimated from data and from MC respectively. In general, there are different weights depending on the period of data taking.

B-tag weight

As far as b-tagging is concerned, the measured scaling factor (SF) $\epsilon_{data}/\epsilon_{MC}$ can be used to reweight the events, where ϵ_{data} and ϵ_{MC} are the per-jet efficiency.

The b-tag weight is defined as:

$$w(n_b, n_c, n_l) = \frac{P^{data}(n_b, n_c, n_l)}{P^{MC}(n_b, n_c, n_l)}$$

where P is the probability of selecting an event as a function of the event content in terms of b, c, l quarks. It can be written as:

$$P = \sum_{n_b, tag}^{n_b} \sum_{n_l, tag}^{n_l} \sum_{n_l, tag}^{n_l} C \epsilon_b^{n_b, tag} (1-\epsilon_b)^{n_b-n_b, tag} \epsilon_c^{n_c, tag} (1-\epsilon_c)^{n_c-n_c, tag} \epsilon_l^{n_l, tag} (1-\epsilon_l)^{n_l-n_l, tag} S \quad (5)$$

where $n_{x, tag}$ is the number of tagged jets of type x , being $x = b, c, l$, C is a combinatorial factor to account for how many ways a set of b, c, l tagged jets can be obtained from the composition (n_b, n_c, n_l) , ϵ_x is the tagging efficiency for $x = b, c, l$ jets and S is a "selection" function which is 1 for selected events and 0 for rejected events. For example, for an exclusive selection of t b-jets:

$$S = \delta_{t n_{tag}}$$

where $n_{tag} = n_{b, tag} + n_{c, tag} + n_{l, tag}$ is total number of tagged jets and δ is the Kronecker delta function.

The probability $P^{data}(n_b, n_c, n_l)$ can be written in terms of MC efficiencies and scaling factors using equation (5) and:

$$\epsilon_{bcl}^{data} = SF_{bcl} \epsilon_{bcl}^{MC} \quad (6)$$

In equation (6) ϵ_{bcl}^{MC} is the tagging efficiency as it is estimated from the MC sample to be reweighted and SF_{bcl} is the scaling factor as it is defined above. For the details of the data-driven estimation of the b-tag efficiency and the mistag efficiency we address to section 2.3.4 and 2.3.4.

In general, the b-tag weight is a function of the event content, the number of tags required, the tagging efficiencies of the MC samples and the scaling factors. The MC efficiencies are in turn function of p_T and η , this dependencies have been omitted above in order to simplify the notation. In figure 27 the scaling factor is shown as a function of the p_T of jet: from this plot we derive that the MC efficiencies overestimate the tagging efficiencies, as they are derived from data.

Lepton efficiency weight

Monte Carlo events are reweighted in order to take into account the trigger efficiency of double lepton triggers. The weight assigned to each

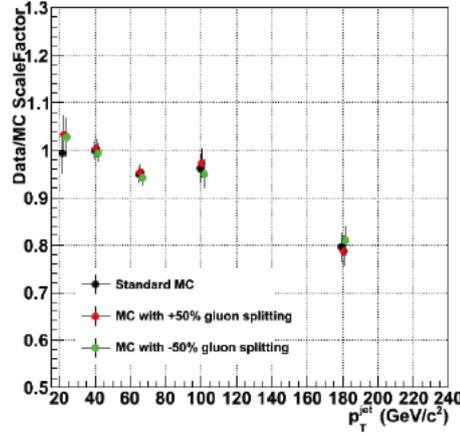


Figure 27: SF as a function of the p_T of the jet for different data-driven estimation of the efficiencies from [9].

event is:

$$w = \frac{\epsilon_{trig}^{data}}{\epsilon_{trig}^{MC}}$$

where ϵ_{trig}^{data} is derived using the "tag and probe" method.

In the Z+b analysis, trigger is required only in data sample, i.e. $\epsilon_{trig}^{MC} = 1$, hence $w = \epsilon_{trig}^{data}$. For double lepton triggers with asymmetric thresholds (omitting "data" apex):

$$w_a = \epsilon_{1,h}\epsilon_{2,l} + \epsilon_{1,l}\epsilon_{2,h} - \epsilon_{1,h}\epsilon_{2,h}$$

where 1,2 stands for lepton 1 and 2, l and h for low and high thresholds. For double lepton triggers with symmetric thresholds:

$$w_s = \epsilon_1\epsilon_2$$

where ϵ is the trigger efficiency for that threshold.

Both types of double lepton triggers have been used for 2011 analysis, hence MC events are reweighted taking into account the relative proportions of symmetric and asymmetric triggered data, respectively 18% and 82%. The final weight is then:

$$w = 0.18w_s + 0.82w_a$$

Bibliography

- [1] The CMS Collaboration, *CMS Physics Technical Design Report, Volume 1: Detector Performance and Software*, 2006
- [2] ALEPH, DELPHI, L3 and OPAL Collaborations, The LEP Working Group for Higgs Boson Searches, "Search for the Standard Model Higgs Boson at LEP", Phys.Lett.B565:61-75, 2003
- [3] The CDF, D0 Collaborations, the Tevatron New Phenomena, Higgs Working Group, "Combined CDF and D0 Upper Limits on Standard Model Higgs Boson Production with up to 8.6 fb^{-1} of Data", arXiv:1107.5518, 2011
- [4] The CMS Collaboration, *Search for standard model Higgs boson in pp collisions at $\sqrt{s} = 7 \text{ TeV}$ and integrated luminosity up to 1.7 fb^{-1}* , 2011
- [5] The CMS Collaboration, *Performance of muon identification in pp collisions at $\sqrt{s} = 7 \text{ TeV}$* , 2010
- [6] The CMS Collaboration, *Electron reconstruction and identification at $\sqrt{s} = 7 \text{ TeV}$* , 2010
- [7] The CMS Collaboration, *Jet Performance in pp Collisions at $\sqrt{s} = 7 \text{ TeV}$* , 2010
- [8] The CMS Collaboration, *Plans for Jet Energy Corrections at CMS*, 2011
- [9] The CMS Collaboration, *Performance of b-jet identification in CMS*, 2011
- [10] Simon de Visscher, Jean-Marc Gerard, Michel Herquet, Vincent Lemaitre, Fabio Maltoni, *Unconventional phenomenology of a minimal two-Higgs-doublet model*, arXiv:hep-ph/0904.0705, 2009
- [11] Simon de Visscher, *Observability of an unconventional two-Higgs-doublet model at the LHC*, Doctoral dissertation, 2009

- [12] F. Febres Cordero, L. Reina and D. Wackerroth, *W and Z boson production with a massive bottom-quark pair at the Large Hadron Collider*, arXiv:0906.1923, 2009
- [13] F. Febres Cordero, *Next-to-Leading-Order Corrections to Weak Boson Production with a Massive Quark Jet Pair at Hadron Colliders*, arXiv:0809.3829, 2008
- [14] J. Campbell, R. K. Ellis, F. Maltoni, and S. Willenbrock, *Production of a Z Boson and Two Jets with One Heavy-Quark Tag*, arXiv:hep-ph/0510362, 2008
- [15] J. Campbell, R. K. Ellis, F. Maltoni, and S. Willenbrock, *Associated Production of a Z Boson and a Single Heavy-Quark Jet*, arXiv:hep-ph/0312024, 2008
- [16] The MG/ME developing Team, <http://madgraph.hep.uiuc.edu/>
- [17] Torbjörn Sjöstrand, Stefan Ask, Richard Corke, Stephen Mrenna, Peter Skands , <http://home.thep.lu.se/~torbjorn/Pythia.html>
- [18] Y.L. Dokshitzer, G.D. Leder, S. Moretti and B.R. Webber, *Better jet clustering algorithms*, JHEP 08(1997)01 arXiv:hep-ph/9707323, 1997
- [19] Matteo Cacciari, Gavin P. Salam, Gregory Soyez, *The anti- k_t jet clustering algorithm*, JHEP04(2008)063, 2008

Acknowledgements

Spero di riuscire ad includere in queste poche righe tutte le persone (e non sono poche) che in un modo o nell'altro hanno contribuito a farmi raggiungere questo traguardo.

Mi sembra doveroso incominciare con il ringraziare il prof. Migliore, per l'infinita disponibilità e pazienza che ha dimostrato nel seguirmi in questo lavoro. Da lui ho imparato moltissime cose e non solo riguardanti la fisica.

Un ringraziamento speciale va anche al dott. Musich per il tempo che mi ha dedicato, gli aiuti che mi ha dato e per la simpatia che mi ha dimostrato.

Grazie anche al dott. Castello perchè il suo lavoro è stato prezioso per questa tesi.

Spendo un ringraziamento generale per il gruppo CMS di Torino, per l'opportunità che ho avuto di lavorare in questo ambiente piacevole e stimolante. In particolare ringrazio Cristina Botta per le interessanti discussioni, Nadia Pastrone per la simpatia che mi ha dimostrato e la prof.ssa Romero per i suoi preziosi consigli.

Ringrazio la dott.ssa Beolè per aver accettato di essere il mio controrelatore e, insieme al dott. Maina e il dott. Prino, per le interessanti discussioni e i consigli che ho ricevuto durante il seminario di metà tesi.

Grazie anche a Luca e Riccardo, compagni fedeli di banco, per la loro simpatia, per la compagnia che ci siamo tenuti tra una particella e l'altra.

Anche se probabilmente non leggerà mai questa pagina, volevo ringraziare il prof. Longhi, perchè chi mi conosce sa che se oggi studio fisica è per "colpa" sua.

Ringrazio la mia famiglia, i miei genitori in particolare, per avermi sempre sostenuto, moralmente ed economicamente, in tutti questi anni.

Infine un grazie alla persona per cui, da due anni a questa parte, tutto, anche questo lavoro, ha un po' più di senso.

