UNIVERSITA' DEGLI STUDI DI PERUGIA
DOTTORATO DI RICERCA IN FISICA
XIX CICLO
Settore scientifico disciplinare FIS-04

"Search for $H \to b\bar{b}$ in association with a $t\bar{t}$ pair at CMS"

Dott. DANIELE BENEDETTI

RELATORI:
Prof. PAOLO LARICCIA

Dott. ATTILIO SANTOCCHIA

COORDINATORE:
Prof. PASQUALE SODANO

A.A. 2005/06

To my parents

Contents

Introduction

In the last years, the precision achieved in high energy physics experiments, in particular at e^+e^- colliders (LEP and SLC), has allowed tests of the Standard Model theory to be performed at a very high level of accuracy. Nevertheless not all the building blocks of the model have been experimentally established yet. In particular, the complete confirmation of the theory requires the observation of a new particle responsible for generating the masses of the fundamental particles: the Higgs boson.

A lower bound of 114.4 GeV/c^2 at 95% Confidence Level for the Higgs boson mass has been set from the direct search performed at LEP. Upper bounds can be predicted from experimental high-precision analyses that are sensitive to energy scales beyond those which can be reached directly. Using the data collected at the e^+e^- colliders and the last measurement of the top mass performed at TeVatron (July 2006), these methods predict a Higgs mass below 166 GeV/c^2 at 95% Confidence Level. The combination of these limits leads to preferring a Higgs mass just above the excluded value from LEP search.

In the low Higgs mass range (up to $M_H = 135$ GeV/c^2) the decay mode with the highest branching fraction is $H \rightarrow b\bar{b}$. At the Large Hadron Collider (LHC), however, the cross section of QCD production of $b\bar{b}$ pairs is about eight orders of magnitude larger than the cross section $gg \rightarrow H$; the observation of the Higgs boson is therefore not feasible using this channel. In this mass range Monte Carlo studies show that it is possible to detect the Higgs boson in the channel $H \rightarrow \gamma\gamma$ as a narrow mass peak above a large background, nevertheless an alternative way to study the Higgs boson is desirable to give a more complete picture and a better understanding of the Higgs boson nature.

A potentially promising channel is the production of the Higgs in association with a $t\bar{t}$ pair. In this case the search for $H \rightarrow b\bar{b}$ is enhanced by the presence of the top quarks that allow a good reduction in possible backgrounds. However the variety and complexity of these events, and particularly the appearance of a large number of high energy jets and/or leptons

requires to rely upon and fully exploit the performance of all components of the Compact Muon Solenoid (CMS) detector. The Tracking system is of particular importance, since the detection of b-jets is crucial to the identification and reconstruction of these events.

Promising results for the $t\bar{t}H$, $H \rightarrow b\bar{b}$ channel were obtained in some early studies using only a parametric modelling of the CMS detector. In this thesis we re-investigate the discovery potential using a detailed simulation of the CMS detector. Every attempt has been made to reproduce real data acquired in proton-proton collisions at a center of mass energy of 14 TeV. Thus, online triggers have been simulated, and all the standard offline reconstruction algorithms, in their most recent stage of development, have been employed in the processing of the data. With these prescriptions the analyses are substantially more complicated than studies based on more simple approaches involving parametric detector modelling. This has the advantage of providing a more reliable picture of the complexity of the events under study, highlighting some of the limitations that might otherwise be overlooked. Even if real data are not still available, we also made an attempt to include the possible effects that systematic uncertainties can have in the $t\bar{t}H$ discovery potential.
The thesis is organised as follows:

In Chapter 1 we give a general description of the Standard Model theory and the Higgs mechanism, with particular emphasis to the theoretical and experimental limits on its mass and on the Higgs search at LHC. In Chapter 2 we describe the LHC particle collider and the CMS detector. In Chapter 3 we describe the method to reconstruct and identify muons and electrons coming from W decays. Chapter 4 is dedicated to the study of the jet clustering algorithms, to the jet energy calibrations and to the reconstruction of the missing transverse energy. In Chapter 5 we explain the b-tagging algorithm developed to identify jets coming from b quarks. The analysis method and the results obtained for the $t\bar{t}H \rightarrow b\bar{b}b\bar{b}qq'l\nu_l$ are described in detail in Chapter 6. In Chapter 7 the analysis methods for the di-lepton ($t\bar{t}H \rightarrow b\bar{b}b\bar{b}l'\mu'_l l\nu_l$) and all-hadronic ($t\bar{t}H \rightarrow b\bar{b}b\bar{b}qq'q''q'''$) are summarised and the combined final results are presented, including the systematic errors.

Chapter 1

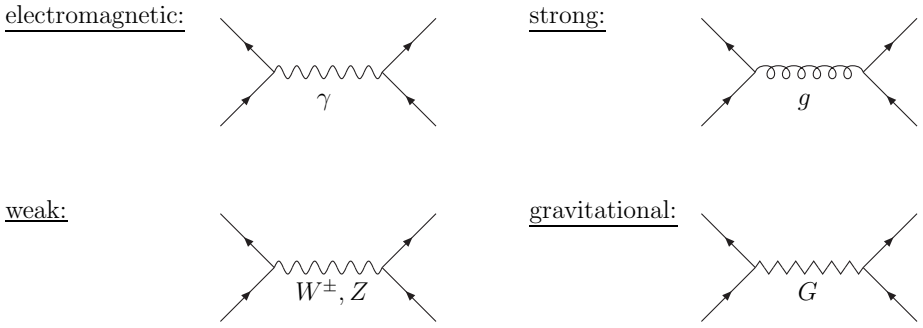
The Standard Model

The laws of Nature are summarised in the Standard Model of particle physics. The Standard Model provides a valid framework for the description of Nature, probed from microscopic scales of order 10^{-16} cm up to cosmological distances of order 10^{28} cm. The Standard Model consists of three components:

1. The basic constituents of matter are leptons and quarks [1] which are realised in three families of identical structure:

leptons	ν_e	ν_μ	ν_τ
	e^-	μ^-	τ^-
quarks	u	c	t
	d	s	b

2. Four different forces act between leptons and quarks: The electromagnetic and weak forces are unified in the Standard Model. The fields associated with these forces, as well as the fields associated with the strong force, are spin-1 fields, describing the photon γ, the electroweak gauge bosons $W^\pm$ and Z, and the gluons g. The interactions of the force fields with the fermionic constituents of matter as well as their self-interactions are described by Abelian and non-Abelian $SU(3) \times SU(2) \times U(1)$ [2] gauge theories. The gravitational interaction is mediated by a spin-2 fields, describing the graviton G, with a character quite different from spin-1 gauge fields. The gravity sec-



tor is attached *adhoc* to the other sectors of the Standard Model, not properly formulated yet as a quantum phenomenon.

3. The third component of the Standard Model is the Higgs mechanism [6]. In this sector of the theory, scalar fields interact with each other in such a way that the ground state acquires a non-zero field strength, breaking the electroweak symmetries spontaneously. The interaction of electroweak gauge bosons, leptons and quarks with this field manifest themselves as non-zero masses of these particles.

Not all building blocks of the model, however, have been experimentally established so far. In particular, the Higgs mechanism for generating the masses of the fundamental particles still lacks experimental verification up to now, even though indirect indications support this mechanism quite strongly. In the following sections the Higgs mechanism and the theoretical and experimental limits to its mass will be described in details. In the following the notation $c = \hbar = 1$ is used.

1.1 The standard electroweak theory

1.1.1 The electroweak sector

In 1961 Glashow shows that the electromagnetic and weak interaction are not two different forces, but rather two aspects of the same force: the electroweak interaction [3]. However it was still not clear why the electroweak interaction should be mediated by a massless boson for the electromagnetic interaction and three extremely massive bosons for the weak interaction. The solution to this problem was given in 1967 by Weinberg [4] and in 1968 by Salam [5] independently. They applied the Higgs mechanism [6] to the $SU(2) \times U(1)$ symmetry group. In 1973 the first neutral current was observed

by the Gargamelle Neutrino Collaboration and finally in 1983 the $W^{\pm}$ and Z bosons were observed by the UA1 collaboration [8]. However the complete confirmation of the theory expects the observation of another particle liable of the generation of the boson and fermion masses: the Higgs boson.

In the electroweak standard theory four gauge fields W^i_μ with $i = 1, 2, 3$ and B_μ are introduced by requiring the invariance of the Lagrangian under a local $SU(2) \times U(1)$ gauge transformation with the covariant derivative:

$$D_\mu = \partial_\mu + ig_1 Y B_\mu + ig_2 \frac{\tau_i}{2} W^i_\mu \tag{1.1}$$

where $\vec{\tau} = 2\vec{T}$ are the Pauli spin matrices, g_1 and g_2 are coupling constants of the interactions. The Lagrangian of the Standard Model can be written as a sum of four independent terms:

$$\mathcal{L} = \mathcal{L}_{Fermions} + \mathcal{L}_{Gauge} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa} \tag{1.2}$$

The term $\mathcal{L}_F$ describes massless fermion fields and their interaction with the gauge fields:

$$\mathcal{L}_{Fermions} = i\bar{\psi} D_\mu \psi \tag{1.3}$$

while the second term contains the kinetic energy of the gauge fields B and $\vec{W}$ and the self interaction with the $\vec{W}$ fields due to the non-Abelian nature of the $SU(2)$ group:

$$\mathcal{L}_{Gauge} = -\frac{1}{4} W^i_{\mu\nu} W^i_{\mu\nu} - \frac{1}{4} B_{\mu\nu} B_{\mu\nu} \tag{1.4}$$

where

$$W^i_{\mu\nu} = \partial_\nu W^i_\mu - \partial_\mu W^i_\nu - g_2 \epsilon^{ijk} W^i_\mu W^k_\nu \tag{1.5}$$

$$B_{\mu\nu} = \partial_\nu B_\mu - \partial_\mu B_\nu \tag{1.6}$$

In the Standard Model the mass eigenstates of the $\vec{W}$ particles are:

$$W^{\pm}_\mu = \frac{1}{2}(W^1_\mu \mp W^2_\mu) \tag{1.7}$$

while the neutral part W^3_μ and the field B are combined together to define the photon field A_μ and the field Z_μ:

$$A_\mu = B_\mu \cos\theta_W + W^3_\mu \sin\theta_W \tag{1.8}$$

$$Z_\mu = -B_\mu \sin\theta_W + W^3_\mu \cos\theta_W \tag{1.9}$$

Therefore the gauge Lagrangian without the self interaction part of the $SU(2)$ group could be write as:

$$\mathcal{L}_{Gauge} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{2}F^{\dagger}_{W\mu\nu}F^{\mu\nu}_{W} - \frac{1}{4}Z_{\mu\nu}Z^{\mu\nu} \tag{1.10}$$

where the $F^{\mu\nu}$ is the electromagnetic tensor, $F^{\mu\nu}_{W}$ is the tensor of the non-hermitian field W and $Z^{\mu\nu}$ is:

$$Z_{\mu\nu} = \partial_{\nu}Z_{\mu} - \partial_{\mu}Z_{\nu} \tag{1.11}$$

We can note that $\mathcal{L}_{Gauge}$ e $\mathcal{L}_{Fermions}$ do not contain the mass terms of the bosons and fermions because they would destroy the gauge invariance of the Lagrangian. To introduce the mass terms preserving the invariance under $SU(2) \times U(1)$ transformation we need to introduce a new scalar field as suggested by Peter Higgs (1964). Indeed the remaining terms $\mathcal{L}_{Higgs}$, $\mathcal{L}_{Yukawa}$, that we do not deal up to now, describe the interaction of the Higgs field with gauge and fermion fields respectively.

1.1.2 The spontaneous symmetry breaking

In field theory the state of lowest energy is the vacuum. Spontaneous symmetry breaking is only relevant to field theory if the vacuum state is non-unique and if a potential that create such degeneracy exists. To explain the spontaneous symmetry breaking in a easier way the main concepts of the Higgs model will be described in the $U(1)$ group [9].

The typical potential form (Fig. 1.1) that create such degeneracy is reported below:

$$V(\phi) = \mu^2(\phi^*\phi) + \lambda^2(\phi^*\phi)^2 \quad \text{con} \quad \phi = \phi_1 + i\phi_2 \tag{1.12}$$

$$\mu^2 < 0 \text{ e } \lambda > 0$$

The Lagrangian density invariant under the $U(1)$ gauge transformations:

$$\begin{aligned} \phi(x) \rightarrow \phi^{'}(x) &= \phi(x)^{-iqf(x)} \\ \phi^*(x) \rightarrow \phi^{*'}(x) &= \phi^*(x)e^{iqf(x)} \\ A_{\mu}(x) \rightarrow A^{'}_{\mu}(x) &= A_{\mu}(x) + \partial_{\mu}f(x) \end{aligned} \tag{1.13}$$

can be written:

$$\mathcal{L}_{em}(x) = [D^{\mu}\phi(x)]^*[D_{\mu}\phi(x)] - \mu^2|\phi(x)|^2 - \lambda|\phi(x)|^4 - \frac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x) \tag{1.14}$$

where $D_{\mu} = \partial_{\mu} + iqA_{\mu}(x)$ is the covariant derivative and $\frac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x)$ is the free gauge field. Taking $\mu^2 < 0$ e $\lambda > 0$ the vacuum state is not unique

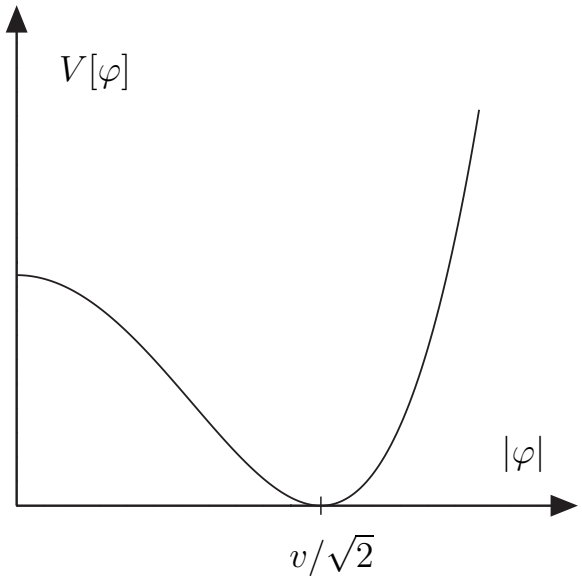


Figure 1.1: The potential function

leading to a spontaneous symmetry breaking and the potential $V(\phi)$ posses a whole circle of absolute minima at:

$$\phi_0 = -\frac{\mu^2}{2\lambda}^{1/2} = \frac{1}{\sqrt{2}}v(> 0) \tag{1.15}$$

We now introduce two real fields $\sigma(x)$ and $\eta(x)$ through the equation

$$\phi(x) = \frac{1}{\sqrt{2}}(v + \sigma(x) + i\eta(x)) \tag{1.16}$$

$\sigma(x)$ and $\eta(x)$ measure the deviations of the field ϕ from the equilibrium ground state configuration ϕ_0. In terms of these fields, the Lagrangian density becomes:

$$\begin{aligned}\mathcal{L}_{em}(x) = \tfrac{1}{2}[\partial^\mu\sigma(x)][\partial_\mu\sigma(x)] - \tfrac{1}{2}(2\lambda v^2)\sigma^2(x) \\ -\tfrac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x) + \tfrac{1}{2}(qv)^2A_\mu(x)A^\mu(x) \\ +\tfrac{1}{2}[\partial^\mu\eta(x)][\partial_\mu\eta(x)] \\ +qvA^\mu(x)\partial_\mu\eta(x) + \text{'interaction terms'}\end{aligned} \tag{1.17}$$

The direct interpretation of equation 1.17 leads to difficulties. The first line of this equation describes a real Klein-Gordon field which gives uncharged spin

0 bosons with mass $\sqrt{(2\lambda v^2)}$. However, the product term $A^\mu(x)\partial_\mu\eta(x)$ shows that $A^\mu(x)$ and $\eta(x)$ are not independents and one cannot conclude that the second and third lines describe massive vector bosons and massless scalar bosons respectively. Counting the degrees of freedom for the Lagrangian 1.14 and 1.17 shows up that the first one has four degree of freedom: two from the complex scalar field $\phi(x)$ and two from the real massless vector field A_μ, and instead the second one had five degree of freedom: the real scalar fields $\sigma(x)$ e $\lambda(x)$ each present one degree and the real massive vector field $A^\mu(x)$ contributes with three degrees. Of course, a change of variables cannot alter the number of degrees of freedom of a system. We must conclude that the Lagrangian density 1.17 contains an unphysical field which does not represent real particles. Indeed the term $A^\mu(x)\partial_\mu\eta(x)$ shows that $\eta(x)$ transforms it self in $A_\mu(x)$ giving to it the own degree of freedom. The scalar field $\eta(x)$ can be then eliminated from 1.16 and the Lagrangian 1.17 can be write as:

$$\mathcal{L}_{em}(x) = \frac{1}{2}[\partial^\mu\sigma(x)][\partial_\mu\sigma(x)] - \frac{1}{2}(2\lambda v^2)\sigma^2(x)$$
$$-\frac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x) + \frac{1}{2}(qv)^2A_\mu(x)A^\mu(x) + \mathcal{L}_I(x) \tag{1.18}$$

where $\mathcal{L}_I(x)$ describes higher-order interaction terms. In this way in the Lagrangian 1.18, $\sigma(x)$ gives a neutral scalar boson of mass $\sqrt{(2\lambda v^2)}$ and $A_\mu(x)$ gives a neutral vector boson of mass $|qv|$. This phenomenon by which a vector boson acquires mass without destroying the gauge invariance of the Lagrangian density is known as the Higgs mechanism and the massive spin 0 boson associated withe the field $\sigma(x)$ is called the Higgs boson.

In the Standard Model the gauge invariance under a local $SU(2) \times U(1)$ transformation is required. In this model the Higgs boson is defined as a weak isospin doublet:

$$\phi(x) = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{1.19}$$

with the complex fields:

$$\phi^+ = \frac{\phi_1 + i\phi_2}{\sqrt{2}} \quad \text{e} \quad \phi^0 = \frac{\phi_3 + i\phi_4}{\sqrt{2}} \tag{1.20}$$

where ϕ_i $i = 1...4$ are real fields. Thus the Higgs Lagrangian can be written as:

$$\mathcal{L}_{Higgs} = (D_\mu\phi)^\dagger(D^\mu\phi) - V(\phi) = (D_\mu\phi)^\dagger(D^\mu\phi) - \mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2 \tag{1.21}$$

where the partial derivative ∂_μ has replaced with the covariant derivative D_μ (1.1). In this case the minima for the Higgs potential are:

$$\phi_0^\dagger\phi_0 = -\frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2} \tag{1.22}$$

Hence four fields can be chosen so that the spinor:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{1.23}$$

In this way only the $\phi_3 = v+h(x)$ survives while the fields $\phi_{1,2,4}$ are unphysical fields. In the unitary gauge they are transformed away and the $W^{\pm}$ and Z are seen to acquire mass. Replacing the 1.23 to 1.21 the kinetic part becomes:

$$(D_\mu\phi)^\dagger(D^\mu\phi) = \frac{1}{2}\partial_\mu h(x)\partial^\mu h(x) + |(ig_1 Y B_\mu + i\frac{g_2}{2}\vec{\tau}\vec{W}_\mu)\frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}|^2 \tag{1.24}$$

and using the 1.7 and the 1.9 we obtain:

$$\begin{aligned}(D_\mu\phi)^\dagger(D^\mu\phi) = \tfrac{1}{2}\partial_\mu h(x)\partial^\mu h(x) + (\tfrac{1}{2}vg_2)^2 W^+_\mu W^{\mu-} \\ + \tfrac{1}{2}\left(\tfrac{1}{2}v\sqrt{{g_1}^2 + {g_2}^2}\right)^2 Z_\mu Z^\mu + ...\end{aligned} \tag{1.25}$$

while the potential part gives:

$$V(\phi) = -\lambda v^2 h(x)^2 - \lambda v h(x)^3 - \frac{1}{4}h(x)^4 = -\frac{1}{2}M_H h(x)^2 ... \tag{1.26}$$

So that we can define the mass of the $W^{\pm}$ and Z bosons:

$$M_W = \tfrac{1}{2}vg_2 \tag{1.27}$$

$$M_Z = \tfrac{1}{2}v\sqrt{{g_1}^2 + {g_2}^2} \tag{1.28}$$

Since in 1.25 there is not term proportional to $A_\mu A^\mu$ we can conclude that:

$$M_\gamma = 0 \tag{1.29}$$

and from 1.26 we can define the mass of the Higgs boson:

$$M_H = \sqrt{2\lambda}v \tag{1.30}$$

The extra terms that are not described in 1.25 and in 1.26 represent interactions of the Higgs boson with the gauge fields and self-interactions of the Higgs boson. We will not give explicitly the expression of $\mathcal{L}_Y$ that describe the interaction of the Higgs with leptons but it is important to know that the fermion mass is proportional to the coupling constant, therefore the Higgs decay preferentially into the heavier fermions available:

$$m_f = \frac{g_f v}{\sqrt{2}} \tag{1.31}$$

This is the reason why, if the Higgs boson has a mass below 135 GeV, it prefers to decay into a $b\bar{b}$ couple: they are the heaviest fermions available for that mass.

1.1.3 The boson masses

An important aspect of the electro-weak theory is that the $W^{\pm}$ and Z masses can be predicted [9]. This theory allow to related the $W^{\pm}$ and Z masses to three experimentally well known quantities: the fine structure constant

$$\alpha = \frac{g_e^2}{4\pi} = \frac{1}{137.04} \tag{1.32}$$

the Fermi coupling constant determined from μ lifetime experiments

$$G_F = (1.16632 \pm 0.00002) \times 10^{-5}\ \text{GeV}^{-2} \tag{1.33}$$

and the weak mixing angle θ_W which is determined from neutrino scattering experiments:

$$\sin^2 \theta_W = 0.227 \pm 0.014 \tag{1.34}$$

We can see that the coupling constants g_1 and g_2 are related with the electromagnetic constant by the Weinberg angle θ_W:

$$g_1 \cos\theta_W = g_2 \sin\theta_W = g_e \tag{1.35}$$

From this equation we can easily obtain that:

$$\cos\theta_W = \frac{g_2}{\sqrt{g_1^2 + g_2^2}} \text{ e } \sin\theta_W = \frac{g_1}{\sqrt{g_1^2 + g_2^2}} \tag{1.36}$$

and therefore the we can write the Z mass as a function of the W mass and the Weinberg angle:

$$M_Z = \frac{M_W}{\cos\theta_W} \tag{1.37}$$

In the Intermediate vector boson (IVB) [9] theory the Fermi coupling constant is define as:

$$\frac{G_F}{\sqrt{2}} = \left(\frac{g_W}{M_W}\right)^2 \tag{1.38}$$

where g_W is:

$$g_W = \frac{g_2}{2\sqrt{2}} \tag{1.39}$$

From these two last equations and from 1.27 we can therefore obtain:

$$v = \left(\frac{1}{\sqrt{2}G_F}\right)^{1/2} \tag{1.40}$$

Using the 1.32 the $W^{\pm}$ and Z can be predicted; before their discovery in 1983 the expected values were [9]:

$$M_W = \left(\frac{\alpha\pi}{G_F\sqrt{2}}\right)^{\frac{1}{2}} \frac{1}{\sin\theta_W} = \left(78.3 \begin{array}{c}+2.5\\-2.3\end{array}\right) GeV \tag{1.41}$$

$$M_Z = \left(\frac{\alpha\pi}{G_F\sqrt{2}}\right)^{\frac{1}{2}} \frac{2}{\sin 2\theta_W} = \left(89.0 \begin{array}{c}+2.1\\-1.8\end{array}\right) GeV \tag{1.42}$$

Unfortunately the same argument does not apply to the Higgs boson because λ remains a free parameter of the standard electroweak theory:

$$M_H = \sqrt{2\lambda}v$$

1.2 Limits on Higgs boson mass

Even if the Standard Model can not predict the mass of the Higgs boson, upper and lower bounds can be derived from internal consistency conditions and from extrapolations of the model to higher energies. More stringent limits can be derived from indirect search: the high statistics achieved in the LEP and SLC experiments has allowed tests of the electroweak theory at the quantum level. The high precision quantum analyses led to a successful prediction of the top-quark mass confirmed later in the Tevatron experiments and the same argument can be applied to the prediction of the Higgs boson mass.

1.2.1 Theoretical limits

The existence of the Higgs boson is essential not only to explain the mass of particles but also to the renormalization of the electroweak theory. Feynman diagrams involving the emission and subsequent re-absorption of Higgs particles contribute to higher-order corrections and without them the electroweak theory would not be renormalizable. As M_H becomes very small or very large, these corrections become very large, and the success of the theory in lowest-order calculations restricts M_H to the extremely wide range: 7 GeV $\lesssim M_H \lesssim$ 10^3 GeV [9].

Quite restrictive bounds on the value of the Standard Model Higgs mass follow from hypothetical assumptions on the energy scale Λ up to which the Standard Model can be extended before new physical phenomena may emerge. The key to these bounds is the fact that quantum fluctuations modify the self-interaction of the Higgs boson and the quartic coupling constant

λ is replaced by an effective energy dependent coupling $\lambda(\mu)$ [10, 11, 12, 13]. These quantum fluctuations are described by Feynman diagrams as depicted in Fig. 1.2. The Higgs loop itself gives rise to an indefinite increase of the

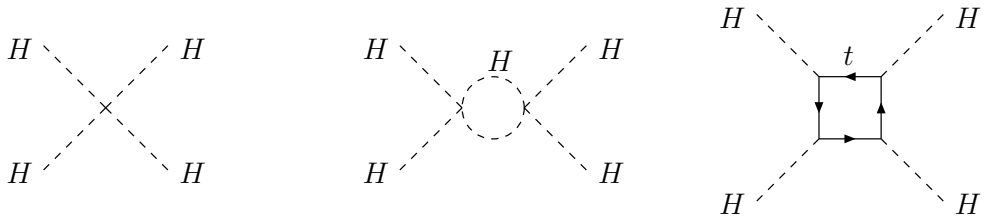


Figure 1.2: Diagrams generating the evolution of the Higgs self-interaction λ

coupling while the fermionic top-quark loop drives, with increasing top mass, the coupling to smaller values, finally even to values below zero. The variation of the effective quartic Higgs coupling $\lambda(\mu)$ and the effective top-Higgs Yukawa coupling $g_t(\mu)$ with energy may be written as:

$$\frac{d\lambda}{dlog(\mu^2)} = \frac{3}{8\pi^2}[\lambda^2 + \lambda g_t^2 - g_t^4] \quad \text{with} \quad \lambda(v^2) = M_H^2/v^2 \tag{1.43}$$

$$\frac{dg_t}{dlog(\mu^2)} = \frac{1}{32\pi^2}\left[\frac{9}{2}g_t^3 - 8g_t g_s^2\right] \quad \text{with} \quad g_t(v^2) = \sqrt{2}m_f/v \tag{1.44}$$

For moderate top masses, the quartic coupling λ rises indefinitely, $d\lambda/dlog(\mu^2) \sim \lambda^2$, and the coupling becomes strong shortly before reaching the Landau pole:

$$\lambda(\mu^2) = \frac{\lambda(v^2)}{1 - \frac{3\lambda(v^2)}{8\pi^2}log(\frac{\mu^2}{v^2})} \tag{1.45}$$

Re-expressing the initial value of λ by the Higgs mass, the condition $\lambda(\Lambda) < \infty$, can be translated to an upper bound on the Higgs mass:

$$M_H^2 \lesssim \frac{8\pi^2 v^2}{3log(\Lambda^2/v^2)} \tag{1.46}$$

A lower bound on the Higgs mass can be based on the requirement of vacuum stability. Since top-loop corrections reduce λ for increasing top-Yukawa coupling, λ becomes negative if the top mass becomes too large. To avoid the instability, the Higgs mass must exceed a minimal value for a given top mass to balance the negative contribution. This lower bound depends on the cut-off value Λ. For a top mass of 175 GeV the allowed Higgs mass values as function of the cut-off Λ are shown in Fig. 1.3. If the Standard Model is assumed to be valid up to the grand unification scale (10^{19} GeV), the Higgs mass is restricted to a narrow window between 130 and 190 GeV.

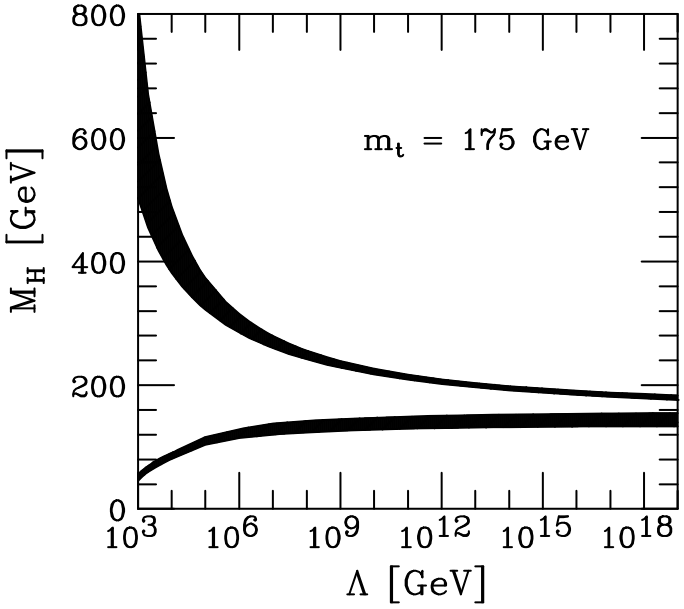


Figure 1.3: Theoretical bounds of the Higgs boson in the Standard Model.

1.2.2 Experimental limits

The high precision of the prediction on the theoretical side is matched by an equivalently high precision on the experimental side. The accuracy achieved in the e^+e^- colliders, in particular at LEP and SLC, has allowed to perform tests of the Standard Model theory at the per-cent level and in some cases down to the per-mille level. The consequence of this accuracy is that the theory can be tested at the level of quantum corrections and that the analyses are sensitives to energy scales beyond those which can be reached directly. The sensitivity to quantum fluctuations in the physical observables demands the rigorous treatment of the electroweak and QCD corrections; we will describe these corrections for the process of fermion-pair production in e^+e^- collisions (Fig. 1.4). This process is mediated by Z-boson and γ exchange in the s-channel: $\mathcal{A} = \mathcal{A}_Z + \mathcal{A}_\gamma$ (for $f = e$, there are also t-channel contributions). The cross section, included the leading logarithmic radiative correction and the $\gamma - Z$ interference contribution (ΔZ), may be written as [14]:

$$\sigma(s) = \frac{12\pi\Gamma_e\Gamma_f}{|Z(s)|^2}[1 + \Delta Z] + \frac{4\pi\alpha^2(s)}{3s}Q_f^2 N_c \tag{1.47}$$

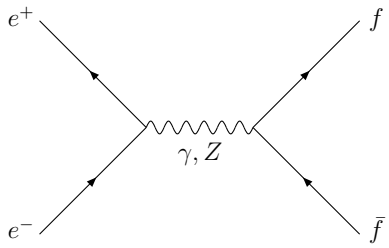


Figure 1.4: The annihilation process $e^+e^- \rightarrow f\bar{f}$ at leading order

where $|Z(s)|^2$ is the Z propagator:

$$|Z(s)|^2 = (s - M_Z^2)^2 + (M_Z\Gamma_Z)^2 \tag{1.48}$$

and Γ_f is the partial decay width defined as function of the axial (g_a) and vector (g_v) coupling constant of the fermion with the Z boson:

$$\Gamma_f = \frac{\alpha M_Z}{3}(g_{v_f}^2 + g_{a_f}^2) \tag{1.49}$$

$$\begin{aligned} g_{v_f} &= \frac{v_f}{2\sin\theta_W\cos\theta_W} = \frac{I_3^W(1 - 2\sin^2\theta_W)}{2\sin\theta_W\cos\theta_W} \\ g_{a_f} &= \frac{a_f}{2\sin\theta_W\cos\theta_W} = \frac{I_3^W}{2\sin\theta_W\cos\theta_W} \end{aligned} \tag{1.50}$$

The second term of the 1.47 describes the photon-exchange contribution.

The 1.47 is mainly corrected by two contributions: QED (photon) corrections and electroweak corrections (Fig. 1.5). Numerically the QED corrections [15] are the most important radiative corrections near the Z resonance because the radiation of photons from initial electrons and positrons shifts the energy $\sqrt{s} \rightarrow \sqrt{s'}$ away from the peak, resulting in a large modification of the cross section. In this case the cross section, including the photon radiation in the initial state, can be expressed as a convolution of the electroweak cross section 1.47 with the radiator function H:

$$\sigma_{QED(s)} = \int_0^{x_{max}} dx H(x)\sigma[(1-x)s] \tag{1.51}$$

where $H(x)$ is the probability that a photon can be radiated and X_{max} is the maximal fraction of the photon energy not resolved by the detector. The complete QED corrections reduce the peak value of the resonance cross

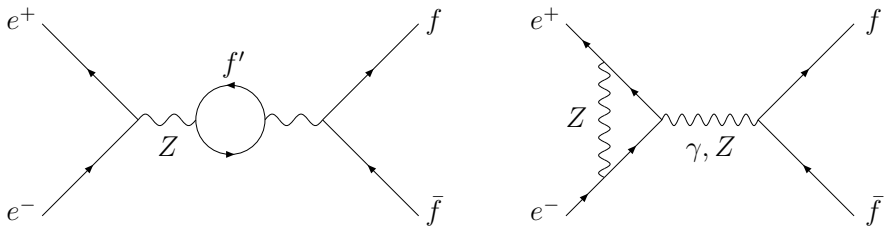


Figure 1.5: electroweak corrections to $e^+e^- \to f\bar{f}$.

section by about 30% and shift the position of the peak upward by about 100 MeV. The electroweak correction are important not only because they modify the total cross section ($e^+e^- \to f\bar{f}$) but mostly because they allow to access to two high mass scales in the model: the top quark mass and the Higgs-boson mass. These particles enter as virtual states in the loop corrections and together with other one loop corrections they lead to modify two important parameters [16]:

$$\rho = \frac{M_W^2}{M_Z^2 \cos^2\theta_W} \quad \to \quad \rho_{eff} = \rho + \Delta\rho \tag{1.52}$$

$$\sin^2\theta_W = 1 - \frac{M_W^2}{M_Z^2} \quad \to \quad \sin^2\theta_{eff} = k \sin^2\theta_W \tag{1.53}$$

$$\tag{1.54}$$

The Higgs and top virtual loop for instance modify the $\sin^2\theta_{eff}$ in this way:

$$\sin^2\theta^l_{eff}(1 - sin^2\theta^l_{eff}) = \frac{\pi\alpha}{\sqrt{2}G_F M_Z^2(1-\Delta r_Z)} \tag{1.55}$$

where Δr_Z is:

$$\Delta r_Z = \Delta\alpha - \Delta\rho^t + \Delta r_Z^H + ... \tag{1.56}$$

$\Delta\alpha$ includes the shift of the electromagnetic coupling near the M_Z pole while $\Delta\rho^t$ and Δr_Z^H are the corrections due to the leading top Higgs contributions. The top has the largest effect because the contribution is quadratic in m_t:

$$\Delta\rho^t = \frac{3G_F m_t^2}{8\pi^2\sqrt{2}} \tag{1.57}$$

Instead the Higgs contribution depend only logarithmically on the Higgs-boson mass:

$$\Delta r_Z^H = \frac{3G_F M_W^2}{8\pi^2\sqrt{2}} \frac{1 + 9\sin^2\theta_W}{3\cos^2\theta_W} \log\frac{M_H^2}{M_W^2} \tag{1.58}$$

Two important parameters that presents a good sensitivity to these corrections are the forward-backward asymmetry, which describes the difference of the production cross sections for leptons and quarks in the forward and backward hemispheres and the left-right asymmetry for longitudinally polarised electron and positron beams. These asymmetries can be expressed in terms of the electroweak parameters as [17]:

$$A_{FB} = \frac{\sigma_F - \sigma_B}{\sigma_F + \sigma_B} = \frac{3}{4}\left[\frac{1 - 4\sin^2\theta^f_{eff}}{1 + (1 - 4\sin^2\theta^f_{eff})}\right]^2 \tag{1.59}$$

$$A_{LR} = \frac{\sigma_L - \sigma_R}{\sigma_L + \sigma_R} = \frac{1 - 4\sin^2\theta^f_{eff}}{1 + (1 - 4\sin^2\theta^f_{eff})} \tag{1.60}$$

Other observables that are evaluated to scrutinise the Standard Model are the left-right-forward-backward asymmetry that allows to measure the asymmetry of the single leptons:

$$\tilde{A}^f_{FB} = \frac{(\sigma^F_L - \sigma^B_L) - (\sigma^F_R - \sigma^B_R)}{\sigma_L + \sigma_R} = \frac{3}{4}A_f \tag{1.61}$$

where f can be e, τ, μ, the partial decay widths:

$$R^0_q = \frac{\Gamma_q}{\Gamma_{had}} \quad (q = b, c) \tag{1.62}$$

and the forward-backward asymmetry near the Z pole for the b and c quarks ($A^{0,b}_{FB}$, $A^{0,c}_{FB}$) [18].

The large amount of electroweak data from e^+e^- colliders available in the Summer of 1994 (the time of quark top discovery) was so accurate to estimate the quark top mass to a value of 173^{+30}_{-33} GeV for mass Higgs varied from 60 GeV to 1 TeV [19, 20]. Known experimentally the top mass the observables of the Standard Model are now function only of the Higgs mass (Fig. 1.8):

$$N_i = f(\log M_H) \tag{1.63}$$

Using the Standard Model observables reported in the left part of Fig. 1.7 [19] it is then possible to build a χ^2 that is function of M_H and minimising the $\chi^2(M_H)$ the expected value of Higgs mass for the SM parameters can be extracted (Fig. 1.7 right):

$$\chi^2 = \sum_i \frac{(N^{th}_i - N^{ex}_i)^2}{(\triangle N)^2} \quad \text{with} \quad \chi^2 = f(\log M_H) \tag{1.64}$$

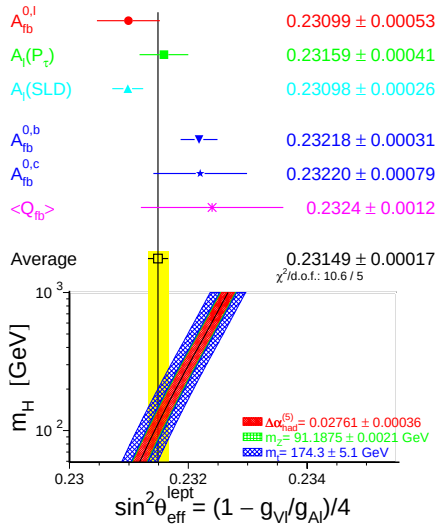


Figure 1.6: sin^2_{eff} as function of M_H obtained from the asymmetry observables.

For $m_t = 171.4 \pm 2.1$ GeV (CDF+D0 July 2006) [21] the expected value for M_H is:

$$M_H = 85^{+39}_{-28} \text{ GeV} \quad \text{at} \quad 68\% \text{ (C.L.)} \tag{1.65}$$

with $M_H < 166$ GeV at 95% C.L.

The lower limit on the Higgs mass is however determined from the directed search performed at LEP, mainly with the channel $e^+e^- \to Z^* \to ZH$. The lower limit, including the 2000 data tacking, is [22]:

$$M_H > 114.4 \text{ GeV at } 95\% \text{ C.L.} \tag{1.66}$$

However in 2000 there was a lot of rumour because ALEPH experiment found three events with an high Higgs probability and also L3 saw one good Higgs candidate but OPAL and DELPHI saw none. Overall there was not enough statistics to claim the discovery but these events together with the Higgs mass indication from theoretical and electroweak fit suggest that the Higgs can have a mass just over the lower limit. In conclusion we can say that the Standard Model seems to prefer a "light" Higgs boson.

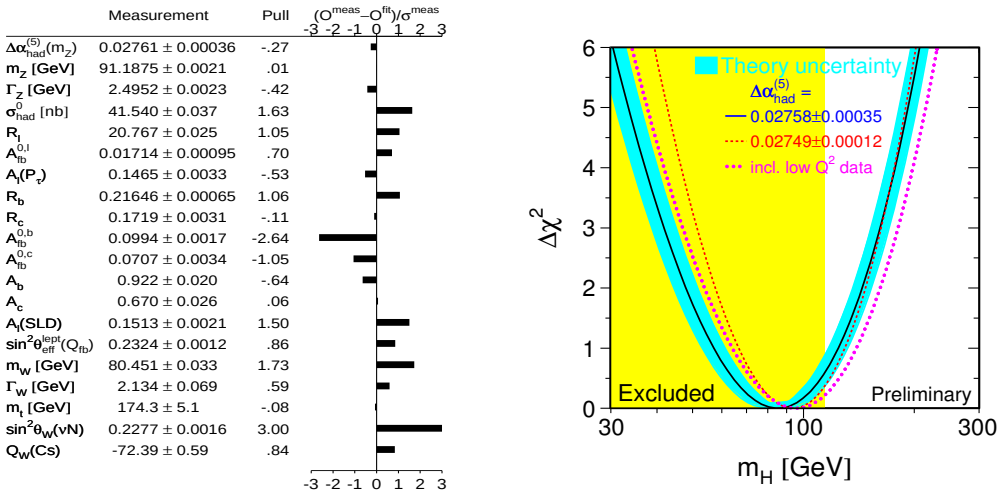


Figure 1.7: Left: experimental measure of the Standard Model observables. Right: final electroweak fit for M_H.

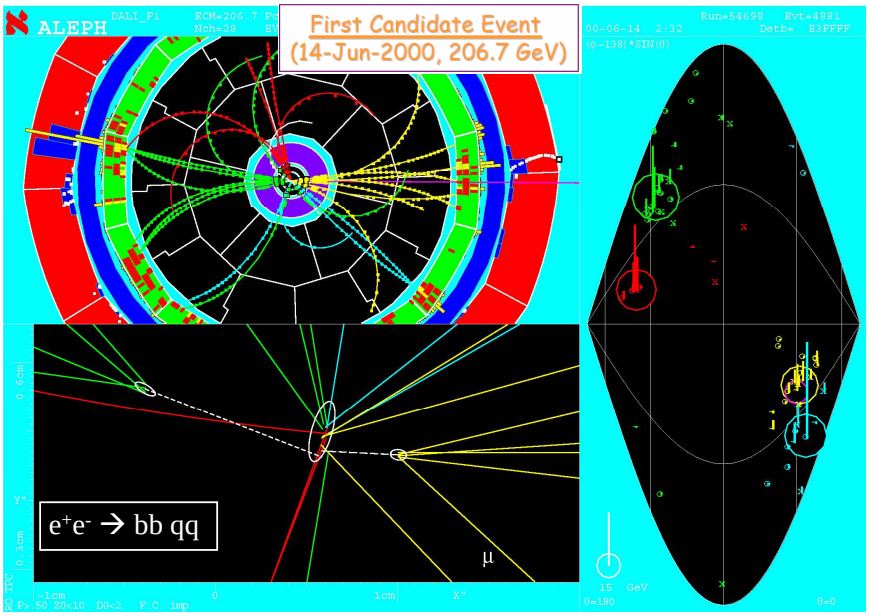


Figure 1.8: ALEPH experiment: the first candidate with am high Higgs probability.

1.3 The Higgs search at LHC

The search for the Higgs boson is one of the main goals of LHC [23]. There is not a single production mechanism to produce the Higgs boson at LHC

but rather, together with the Higgs decay modes, the number of channels is quite large [24]. The process with the largest cross section, that dominate the production mechanism in the entire relevant Higgs mass range up to about 1 TeV, is the gluon fusion process (Fig. 1.9: a) where a quark loop connects the massless gluons with the Higgs boson. The second largest contribution

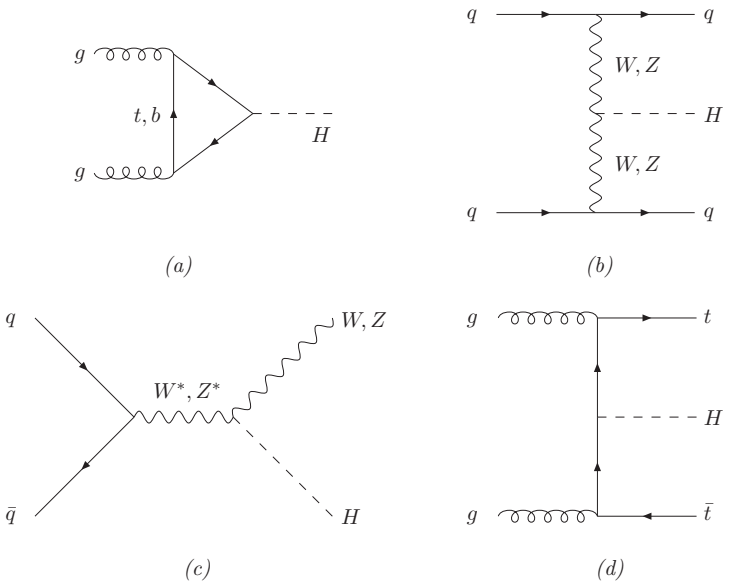


Figure 1.9: Higgs main production channel at hadron colliders.a) Gluon fusion process, b) boson fusion, c) Higgs-strahlung, d) Associated production

to LHC cross section is the WW, ZZ fusion (Fig. 1.9: b) where in additional to the Higgs boson two highly energetic quarks in forward and backward direction are produced. The third and forth Higgs production mechanisms are the Higgs-strahlung off W, Z (Fig. 1.9: c) and the associated production with a qq pairs (mainly $t\bar{t}$) (Fig. 1.9: d). These mechanisms can be important for low Higgs mass range because the additional particle produced can help to reject the backgrounds and also because these channels can allow a direct determination of the coupling constants of the Higgs with the other particles of the Standard Model. An overview of all relevant production cross sections at the LHC is given in Fig. 1.10.

The Higgs decay channels change rapidly with the assumed Higgs mass. The reason is that the Higgs boson prefers to decay to as heavy particles as kinematically possible because the coupling Higgs constant is proportional to the square of the boson masses and to the mass of the fermions. With increasing Higgs mass, new decay channels open up and the decays to lighter

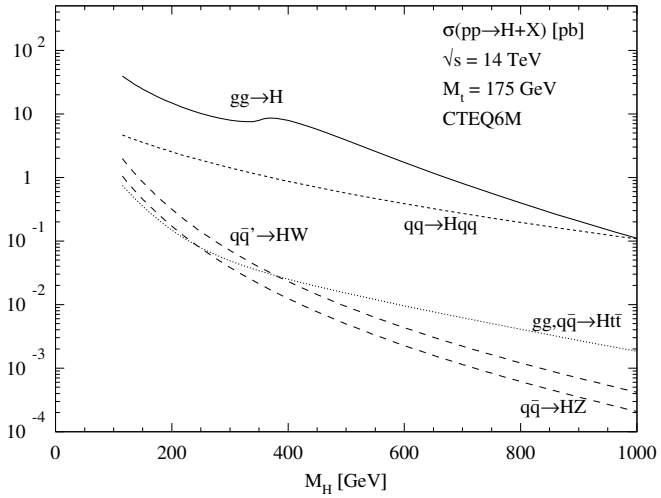


Figure 1.10: Overview of the LHC cross section for the Higgs production.

particles loose importance (Fig 1.11 right). However the Higgs mass decay can be practically divided into two mass ranges.

For $M_H > 135$ GeV the main decay modes are those into WW and ZZ pairs, where one of the vector bosons is off-shell below the corresponding kinematical threshold. The presence of leptons (μ, e) in the final state allow to reach the Higgs discovery in the first 5 fb^{-1} for an Higgs mass up to 500 GeV (Fig 1.12) [24, 25, 26, 27]. For very large Higgs masses the total decay width grows up to the order of the Higgs mass itself so that the interpretation of the Higgs boson as a resonance becomes more difficult (Fig 1.11 left).

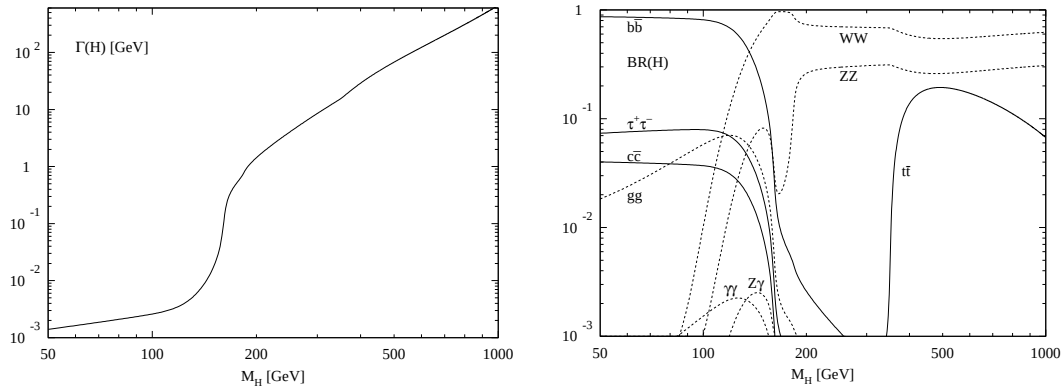


Figure 1.11: Left: total decay width of SM Higgs boson as a function of its mass. Right: branching ratios of the dominant decay modes of SM Higgs boson.

For $M_H < 135$ GeV the Higgs boson mainly decays into $b\bar{b}$ pairs. However, with this final state, direct Higgs production (gluon fusion) is almost impossible to detect due to the huge QCD cross section for $b\bar{b}$ pairs and the inability to reconstruct the Higgs mass very precisely with hadronic jets. In this mass range become then important the Higgs decay into photon pairs which is mediated by W, top and bottom quark loops. The branching fraction of this decay is very small, about 2×10^{-3}, however it has long been understood that the $H \rightarrow \gamma\gamma$ can be detected as a narrow mass peak above a large background [28]. Another possibility to detect the Higgs boson in this mass range will be presented in this work studying the Higgs decaying into $b\bar{b}$ but using the Higgs production associated with a $t\bar{t}$ pair [29, 30, 31]. The $t\bar{t}H$ holds promise because it entails substantially lower backgrounds, however the high number of jets in the final state and the presence of b-jets that have to be detected to reject the $t\bar{t} + jets$ backgrounds makes this channel one of the most challenging channels that will be studied at LHC.

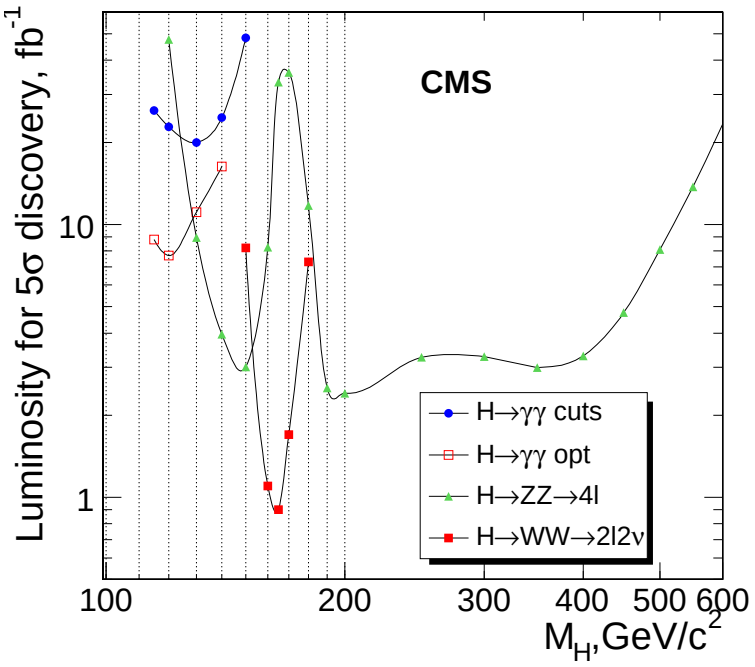


Figure 1.12: Luminosity needed for 5σ discovery of SM Higgs boson as a function of M_H.

Chapter 2

The CMS experiment at LHC

The prime motivation of the Large Hadron Collider (LHC) [23], as we learnt from the previous chapter, is to elucidate the nature of the electroweak symmetry breaking for which the Higgs mechanism is presumed to be responsible. Even if there are many indications for a "light" Higgs boson (below 200 GeV) the LHC beam energy have been chosen in order to explore all possible mass ranges up to the TeV scale. Furthermore the TeV energy scale is important to investigate the possibility of new theory beyond the Standard Model i.e. supersymmetry, extra dimensions, Technicolour etc.

The Compact Muon Solenoid (CMS) [32] experiment has been designed to meet the goals of the LHC physics programme and in particular, if we concentrate on the Higgs boson search, a good resolution on photons, leptons and b-jets is crucial to be able to explore all the Higgs mass range from 100 GeV up to 1 TeV (Sec. 1.3). To do that the requirements for the CMS experiment can be summarised as follows:

- Good muon identification and momentum resolution over a wide range of momenta in the region $|\eta| < 2.5$, good dimuon mass resolution ($\approx 1\%$ at 100 GeV), and the ability to determine unambiguously the charge of muons with $p < 1$ TeV.

- Good charged particle momentum resolution and reconstruction efficiency in the inner tracker. Efficient triggering and offline tagging of τ's and b-jets, requiring pixel detectors close to the interaction region.

- Good electromagnetic energy resolution, good diphoton and dielectron mass resolution ($\approx 1\%$ at 100 GeV), wide geometric coverage ($|\eta| < 2.5$), measurement of the direction of photons and/or correct localisation of the primary interaction vertex, π^0 rejection and efficient photon and lepton isolation at high luminosities.

- Good E_T^{miss} and dijet mass resolution, requiring hadron calorimeters with a large hermetic geometric coverage ($|\eta| < 5$) and with fine lateral segmentation ($\Delta\eta \times \Delta\phi < 0.1 \times 0.1$).

Summarising the main distinguishing features of CMS are a high-field solenoid, a full silicon-based inner tracking system, and a fully active scintillating crystals-based electromagnetic calorimeter.

2.1 The LHC machine

The LHC machine will be the most powerful hadron collider running in the next two decades. It is under construction in the already existing LEP tunnel at CERN laboratories in Geneva, Switzerland. It is situated about 100 m depth underground between French-Swiss border.

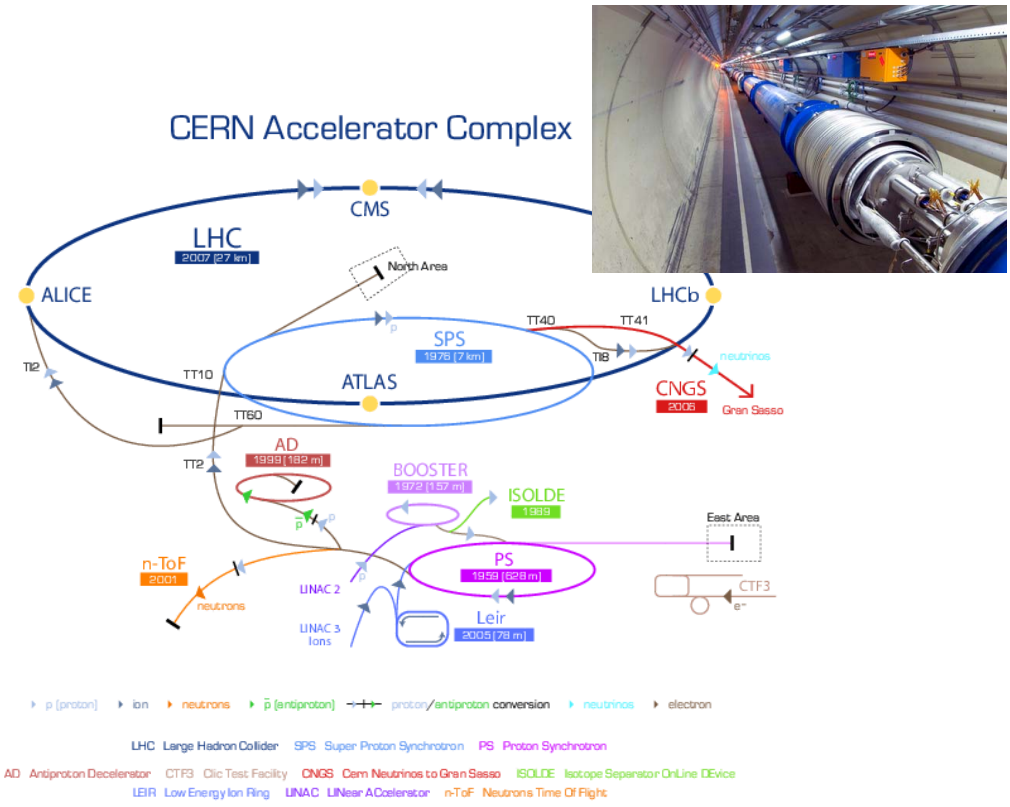


Figure 2.1: Overview of the LHC machine

The LHC accelerator will produce collisions between proton beams with $\sqrt{s} = 14$ TeV, an energy never reached before. The machine parameters relevant for the operation of CMS are listed in Table 2.1. The LHC machine

		pp	**HI**	
Energy per nucleon	E	7	2.76	TeV
Dipole field at 7 TeV	B	8.33	8.33	T
Design Luminosity*	$\mathcal{L}$	10^{34}	10^{27}	$\mathrm{cm}^{-2}\ \mathrm{s}^{-1}$
Bunch separation		25	100	ns
No. of bunches	k_B	2808	592	
No. particles per bunch	N_p	1.15×10^{11}	7.0×10^{7}	
Collisions				
β-value at IP	β^*	0.55	0.5	m
RMS beam radius at IP	σ^*	16.7	15.9	μm
Luminosity lifetime	τ_L	15	6	hr
Number of collisions/crossing	n_c	≈ 20	–	

Table 2.1: The machine parameters relevant for the LHC detectors.

* For heavy-ion (HI) operation the design luminosity for Pb-Pb collisions is given.

comprises 1232 dipole magnets, with r.f. cavities providing a "kick" that results in an increase in the proton energy of 0.5 MeV/turn. The luminosity of an hadron machine is given by:

$$\mathcal{L} = \frac{\gamma f k_B N_p^2}{4\pi\epsilon_n\beta^*} F, \tag{2.1}$$

where γ is the Lorentz factor, f is the revolution frequency, k_B is the number of bunches, N_p is the number of protons/bunch, ε_n is the normalised transverse emittance (with a design value of 3.75μm), β^* is the betatron function at the IP, and F is the reduction factor due to the crossing angle. The nominal energy of each proton beam is 7 TeV. The design luminosity of $\mathcal{L} = 10^{34}$ $\mathrm{cm}^{-2}\ \mathrm{s}^{-1}$ leads to around 1 billion proton-proton interactions per second.

The LHC is planned to produce the first pilot run in October 2007 [33], the expected evolution of its performance parameters is reported in Tab. 2.2. In 2008, following the pilot run, operation will start at 75 ns and subsequently 25 ns bunch spacing. The 75 ns operation is considered an important step in the commissioning of the LHC and the experiments. It will be especially useful for establishing synchronisation quickly. In the 75 ns mode, each beam

	Pilot run 2007	First Physics 2008
Number of bunches	43→156	936→2808
β^*	18 m→2 m	2 m→ 0.55 m
Protons per bunch	$10^{10} \rightarrow 4 \times 10^{10}$ (10^{11})	4×10^{10}
Luminosity	$3 \times 10^{29} \rightarrow 2 \times 10^{31}$ (10^{32})	$10^{32} \rightarrow 2 \times 10^{33}$
Integrated Luminosity	10 pb^{-1}	< 5 fb^{-1}

Table 2.2: Expected evolution of LHC performance parameters during 2007–2008 [34].

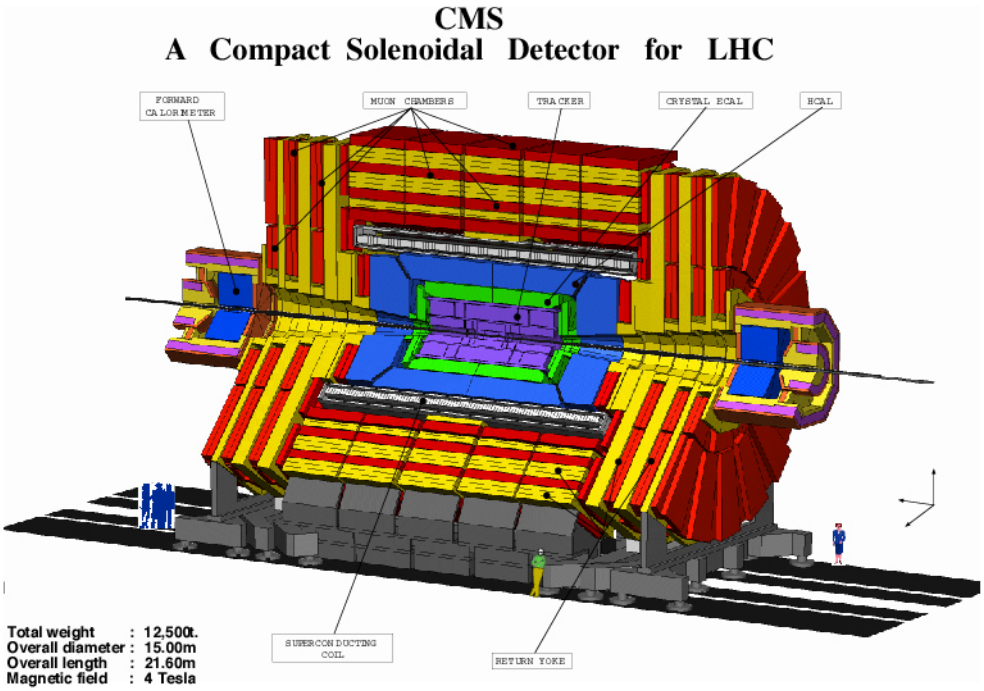
will comprise 936 bunches. A crossing angle will be needed, though it can be smaller than for the 25 ns mode. It will be desirable to move quickly to a bunch spacing of 25 ns. The number of bunches will then be the nominal number (2008) with a crossing angle of 285 μrad. The beam current cannot exceed half the nominal value as part of the beam dump and collimation systems are staged. This will limit the initial luminosity to $\mathcal{L} = 2 \times 10^{33}$ cm^{-2} s^{-1} until the 2010 run.

In the first full year of physics running the LHC should reach an integrated luminosity of about 5 fb^{-1} [33]. It will be enough for the Higgs discovery if its mass is above 135 GeV (Sec. 1.3).

2.2 The CMS detector

The Compact Muon Solenoid experiment [32], is a general purpose detector which will operate at LHC. The main feature of CMS is the 4 T superconducting solenoid that permits a compact design of the detector with a strong magnetic field. The structure of CMS is typical of a general purpose experiment designed for a collider: several cylindrical layers coaxial to the beam direction, referred to as barrel layers, closed at both ends by detector disks orthogonal to the beam pipe, the endcaps, to ensure detector hermeticity. In Fig. 2.2 a schematic view of CMS is drawn pointing out the cylindrical symmetry of the experiment, which has a full length of 21.6 m, a diameter of 14.6 m and reaches a total weight of 12 500 t.

The natural coordinate frame used to describe the detector geometry is a right-handed cartesian system with the x axis pointing to the centre of LHC ring, the z axis coincident with the CMS cylinder axis and the y axis directed almost upward along the vertical. The cylindrical symmetry of CMS design and the invariant description of pp physics drive to use a pseudo-angular reference frame, given by the triplet (r, ϕ, η), with r distance from z axis, ϕ



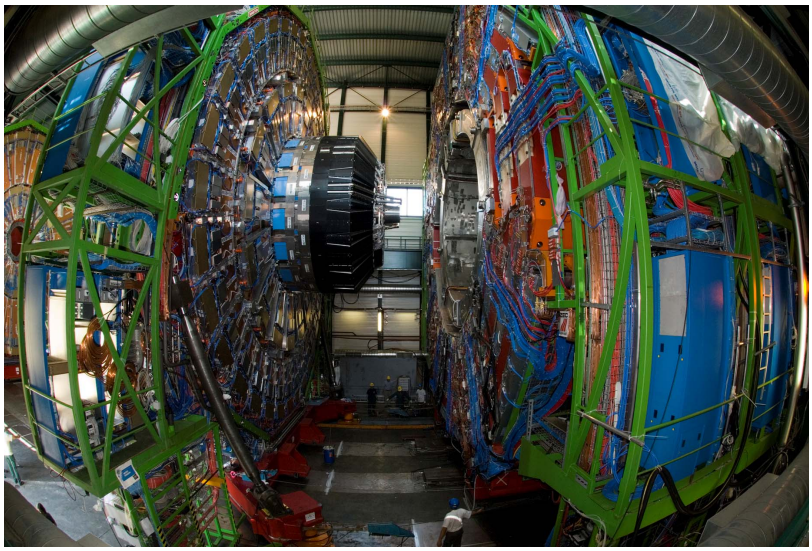


Figure 2.2: Above: overview of the CMS experiment. Below: picture of the CMS detector at P5 before its closing for the magnet test cosmic challenge; on the front there are the last ring of the CMS barrel and the first ring of the CMS endcap.

azimuthal coordinate with respect to x axis and pseudorapidity η defined as:

$$\eta = -\ln(\tan(\theta/2))$$

In this reference frame it is easy to describe the CMS subdetectors, that are installed radially from inside out as described below:

Tracker $r < 1.2$ **m,** $|\eta| < 2.5$: silicon pixel vertex detector plus 200 m^2 active area of silicon microstrip detectors to reconstruct charged particle tracks and individuate primary and secondary vertexes.

Electromagnetic Calorimeter (ECAL) $1.2 < r < 1.8$ **m,** $|\eta| < 3.0$: 61200 lead tungstate crystals to measure electrons and photons

Hadronic Calorimeter (HCAL) $1.8 < r < 2.9$ **m,** $|\eta| < 5.0$: plastic scintillator and brass absorbers to identify jet position and jet transverse energy.

Magnet $2.9 < r < 3.8$ **m,** $|\eta| < 1.5$: 4T longitudinal magnetic field supplied by a superconducting solenoid.

Muon System $4.0 < r < 7.4$ **m,** $|\eta| < 2.4$: drift tubes and cathode strip chambers to detect muon and its sign up to a momentum of ≈ 1 TeV.

2.2.1 Magnet

The CMS magnet is a large superconducting solenoid with a diameter of 5.9 m. It provides an inner uniform 4 T magnetic field obtained with a current of 20 kA. The main features of the CMS solenoid are a central flat superconducting cable, an high purity aluminium stabiliser and an external aluminium-alloy to reinforce the sheath. The superconducting cable is a Rutherford type with 40 NiTb strands and is is kept cooled by a liquid helium cryogenic system. The magnetic flux is closed in a loop via a 1.8 m thick saturated iron yoke, instrumented with four muon stations. The bore of the magnet coil is also large enough to accommodate the inner tracker and the calorimetry inside. The main parameters of the CMS magnet are given in Tab. 2.3 [35].

2.2.2 The tracker system

The silicon tracker [36] is the inner detector of CMS. It is the closest to the interaction point and represents an essential detector to address the multiplicity of LHC physics goals. To better solve the pattern recognition problem, the tracker is designed to fulfil two basic properties: low cell occupancy and large hit redundancy. By considering the charged particle flux at various radii at high luminosity, 3 regions can be delineated:

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
Hoop stress	64 atm

Table 2.3: Parameters of the CMS superconducting solenoid.

- Closest to the interaction vertex where the particle flux is the highest ($\approx 10^7$/s at $r \approx 10$ cm), pixel detectors are placed. The size of a pixel is $\approx 100 \times 150 \mu\text{m}^2$, giving an occupancy of about 10^{-4} per pixel per LHC crossing.

- In the intermediate region ($20 < r < 55$ cm), the particle flux is low enough to enable use of silicon microstrip detectors with a minimum cell size of 10 cm $\times 80$ μm, leading to an occupancy of $\approx$ 2–3%/LHC crossing.

- In the outermost region ($r > 55$ cm) of the inner tracker, the particle flux has dropped sufficiently to allow use of larger-pitch silicon microstrips with a maximum cell size of 25 cm $\times$ 180μm, whilst keeping the occupancy to $\approx$ 1%.

Even in heavy-ion (Pb-Pb) running, the occupancy is expected to be at the level of 1% in the pixel detectors and less than 20% in the outer silicon strip detectors, permitting track reconstruction in the high density environment.

The layout of the CMS tracker is shown in Figure 2.3. The outer radius of the CMS tracker extends to nearly 110 cm, and its total length is approximately 540 cm. Close to the interaction vertex, in the barrel region, are 3 layers of hybrid pixel detectors at a radii of 4.4, 7.3, and 10.2 cm. In the barrel part, the silicon microstrip detectors are placed at r between 20 and 110 cm. The barrel part is separated into an Inner and an Outer Barrel. In order to avoid excessively shallow track crossing angles, the Inner Barrel is shorter than the Outer Barrel, and there are an additional 3 Inner Disks in the transition region between the barrel and endcap parts, on each side of the Inner Barrel. Each of the two Endcap regions are instead build with 2 pixel and 9 microstrip layers.

The total area of the pixel detector is ≈ 1 m^2, while that of the silicon strip detectors is 200 m^2, providing coverage up to $|\eta| < 2.4$. The inner

tracker comprises 66 million pixels and 9.6 million silicon strips [37]. With

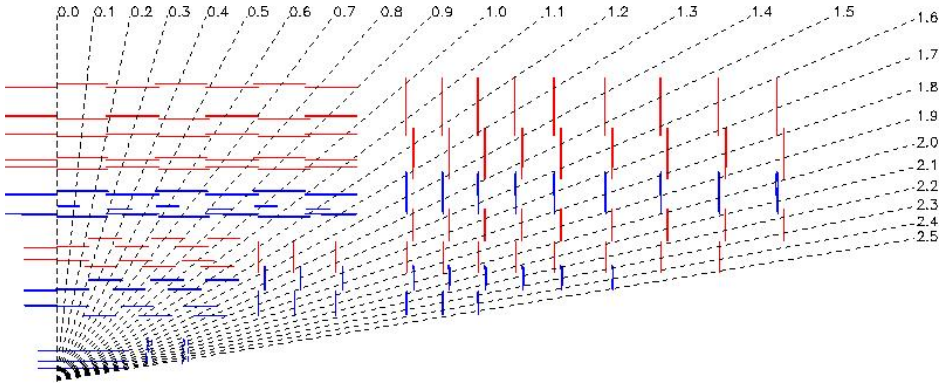


Figure 2.3: The tracker layout (1/4 of the z view).

these features the inner tracker is able to reconstruct single leptons with an efficiency close to 100% and with a lepton momentum resolution better than 3% within $|\eta| < 2.0$ for single muons with different transverse momenta, as shown in Fig. 2.4 [38].

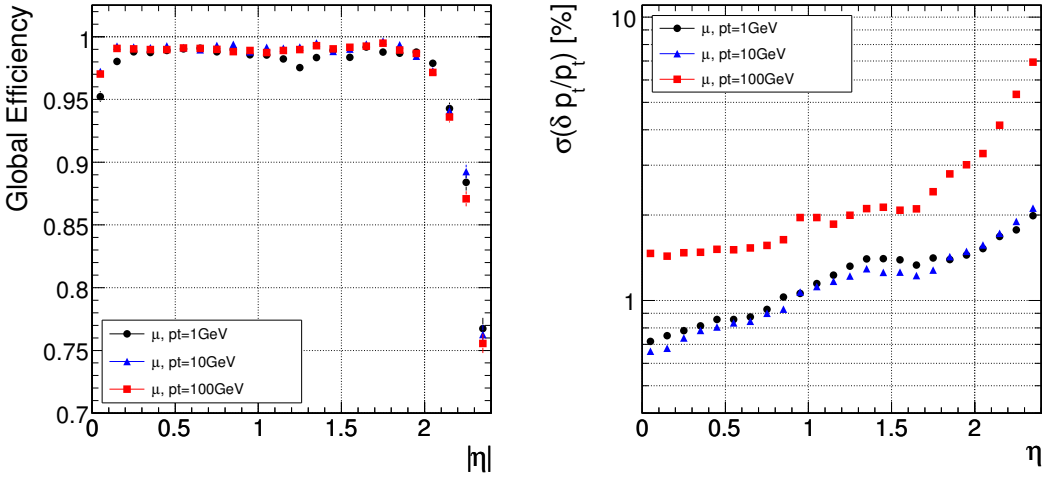


Figure 2.4: Left: global track reconstruction efficiency for muons. Right: resolution of transverse momentum for single muons with transverse momenta of 1, 10 and 100 GeV.

Pixel tracker

The pixel detector [36] Fig 2.5 is a fundamental device for b-tagging studies and impact parameter measurements. It is also important as a starting point

in reconstructing charged particle tracks. It consists of 3 barrel layers with 2 endcap disks on each side on them. The 3 barrel layers are located at mean radii of 4.4 cm, 7.3 cm and 10.2 cm, and have a length of 53 cm. The 2 end disks, extending from 6 to 15 cm in radius, are placed on each side at $|z| = 34.5$ cm and 46.5 cm, in this way the pixel tracker covers the region $|\eta| < 2.4$. At high luminosity conditions, the inner barrel layer will be substituted by an outer layer placed at $r = 13$ cm to improve resolution and limit radiation damages. In order to achieve the optimal vertex position

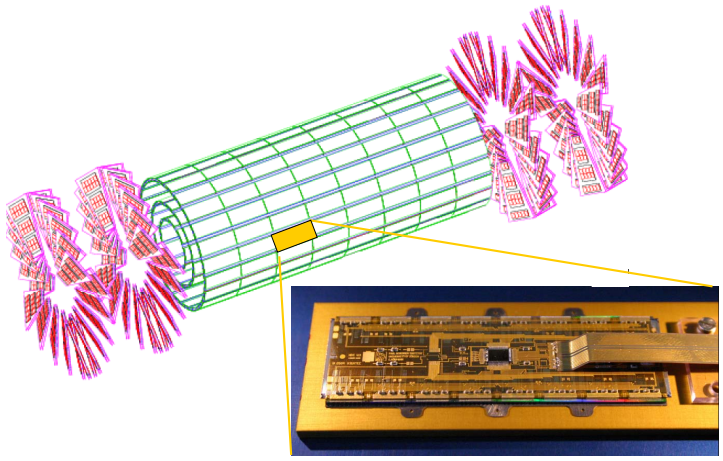


Figure 2.5: Layout of pixel detectors in the CMS tracker.

resolution, a design with an "almost" square pixel shape of $100\times150\mu\text{m}^2$ in both the (r,ϕ) and the z coordinates has been adopted. The barrel comprises 768 pixel modules arranged into half-ladders of 4 identical modules each. The large Lorentz effect (Lorentz angle is 23°) improves the r-ϕ resolution through charge sharing. The endcap disks are assembled in a turbine-like geometry with blades rotated by 20° to also benefit from the Lorentz effect. The endcap disks comprise 672 pixel modules with 7 different modules in each blade. The spatial resolution is measured to be about 10μm for the r-ϕ measurement and about 20μm for the z measurement. The detector is readout using approximately 16 000 readout chips, which are bump-bonded to the detector modules. The life time at full luminosity is 2, 5, 10 years for the inner, middle and outer layers respectively. However to reduce costs the pixel detector during the low luminosity scenario will be probably placed with only two barrel layers and one disk per each endcap region.

Strip tracker

The barrel tracker region is divided into 2 parts: a TIB (Tracker Inner Barrel) and a TOB (Tracker Outer Barrel). The TIB is made of 4 layers and covers up to $|z| < 65$ cm, using silicon sensors with a thickness of 320 μm and a strip pitch which varies from 80 to 120μm. The first 2 layers are made with "stereo" modules in order to provide a measurement in both r-ϕ and r-z coordinates. A stereo angle of 100 mrad has been chosen. This leads to a single-point resolution of between 23–34 μm in the r-ϕ direction and 230 μm in z. The TOB comprises 6 layers with a half-length of $|z| < 110$ cm. As the radiation levels are smaller in this region, thicker silicon sensors (500μm) can be used to maintain a good S/N ratio for longer strip length and wider pitch. The strip pitch varies from 120 to 180μm. Also for the TOB the first 2 layers provide a "stereo" measurement in both r-ϕ and r-z coordinates. The stereo angle is again 100 mrad and the single-point resolution varies from 35–52 μm in the r-ϕ direction and 530 μm in z.

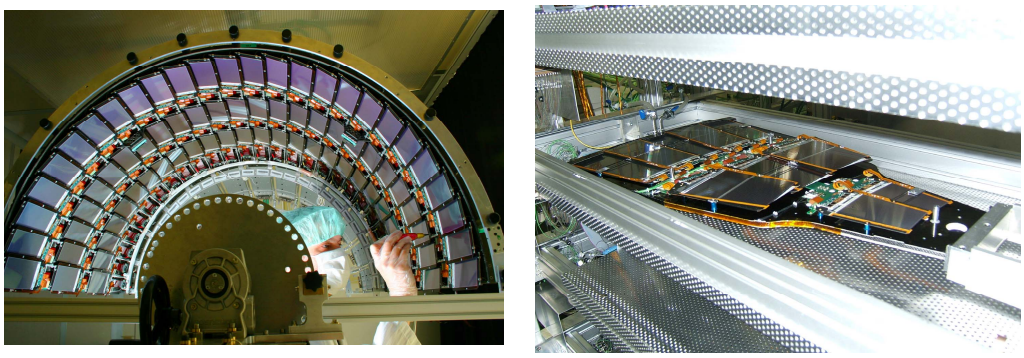


Figure 2.6: Left: Half of the Layer 4+ shell for the Tracker Inner Barrel. Right: A TEC petal of the CMS Tracker

The endcaps are divided into the TEC (Tracker End Cap) and TID (Tracker Inner Disks). Each TEC comprises 9 disks that extend into the region 120 cm$< |z| <$ 280 cm, and each TID comprises 3 small disks that fill the gap between the TIB and the TEC. The TEC and TID modules are arranged in rings, centred on the beam line, and have strips that point toward the beam line, therefore having a variable pitch. The first 2 rings of the TID and the innermost 2 rings and the fifth ring of the TEC have "stereo" modules. The thickness of the sensors is 320 μm for the TID and the 3 innermost rings of the TEC and 500 μm for the rest of the TEC.

The entire silicon strip detector consists of almost 15 400 modules, which will be mounted on carbon-fibre structures and housed inside a temperature

controlled outer support tube. The operating temperature will be around −20°C.

part	No. detectors	thickness (μm)	mean pitch (μm)
TIB	2724	320	81/118
TOB	5208	500	81/183
TID	816	320	97/128/143
TEC	2512	320	96/126/128/143
TEC(2)	3888	500	143/158/183

Table 2.4: Detector types in the silicon tracker

2.2.3 Electromagnetic calorimeter

An high performance electromagnetic calorimeter is a fundamental requirement for any general purpose LHC experiment for precise measurements on electrons and photons. In particular since for $M_H < 135$ GeV the main discovery channel for the Higgs boson is $gg \to H \to \gamma\gamma$, the $\gamma\gamma$ invariant mass resolution should be of the order of 1% to enhance the signicance of a possible signal. The CMS collaboration has chosen a hermetic, homogeneous electromagnetic calorimeter (ECAL) comprising 61 200 lead tungstate ($PbWO_4$) crystals mounted in the central barrel part, closed by 7324 crystals in each of the 2 endcaps [39] (Fig. 2.7). The lead tungstate crystals have a

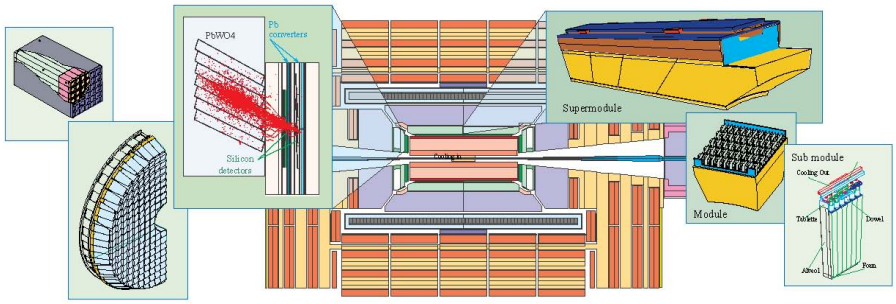


Figure 2.7: The CMS electromagnetic calorimeter

short scintillation decay time $\tau = 10$ ns that allows to collect 85% of the light in the 25 ns interval between two pp collisions. The small Molire radius of 21.9 mm and radiation length $\chi_0 = 8.9$ mm permits the shower containment in a limited space resulting in a compact calorimeter design.

The barrel section (EB) has an inner radius of 129 cm. It is structured as 36 identical "supermodules," (Fig. 2.8 left) each covering half of the barrel

length and corresponding to a pseudorapidity interval of $0 < |\eta| < 1.479$. The crystals are quasi-projective (the axes are tilted at 3° with respect to the line from the nominal vertex position) and cover 0.0174 (i.e. 1°) in $\Delta\phi$ and $\Delta\eta$. The barrel crystals have a front face of about 22×22 mm^2, which matches well the Molire radius of 21.9 mm. To limit fluctuations on the longitudinal shower leakage of high-energy electrons and photons, the crystals were chosen with a total thickness of 25.8 radiation lengths, corresponding to a crystal length of about 23 cm. An R&D programme has shown that radiation affects neither the scintillation mechanism nor the uniformity of the light yield along the crystal. It only affects the transparency of the crystals through the formation of colour centres. This light loss will be monitored by a light-injection system.

The endcaps (EE), at a distance of 314 cm from the vertex and covering a pseudorapidity range of $1.479 < |\eta| < 3.0$, are each structured as 2 "Dees", consisting of semi-circular aluminium plates from which are cantilevered structural units of 5×5 crystals, known as "supercrystals" (Fig. 2.8 right). The endcap crystals, like the barrel crystals, off-point from the nominal vertex position, but are arranged in an x-y grid (i.e. not an η-ϕ grid). They are all identical and have a front face cross section of 28.6×28.6 mm^2 and a length of 220 mm (24.7 X_0). A preshower device is placed in front of the

Figure 2.8: Left: Two ECAL barrel supermodules, installed for the MTCC. Right: The 60 mm thick aluminium backplate of an ECAL endcap Dee held on its assembly and installation frame. Fours mock supercrystals (5 × 5 crystals) are shown attached in the spacer.

crystal calorimeter over much of the endcap pseudorapidity range to help in neutral pions identification and to improve the electron tagging against minimum ionising particles. The preshower detector is a sampling calorimeter with 2 layers: lead radiators initiate electromagnetic showers from incoming photons/electrons whilst silicon strip sensors placed after each radiator

measure the energy deposited and the transverse shower profiles.

The crystal of the electromagnetic calorimeter are equipped with avalanche photodiodes in the barrel and vacuum phototriodes in the endcap, the resulting data amplified and digitised are transported off the detector via optical fibres to the readout and trigger systems in the counting room.

The performance of a supermodule was measured in a test beam. Representative results on the energy resolution as a function of beam energy are shown in Fig. 2.9. The energy resolution, measured by fitting a Gaussian function to the reconstructed energy distributions, has been parameterised as a function of energy [38]:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2, \tag{2.2}$$

where S is the stochastic term, N the noise and C the constant term. The values of these parameters are listed in the figure.

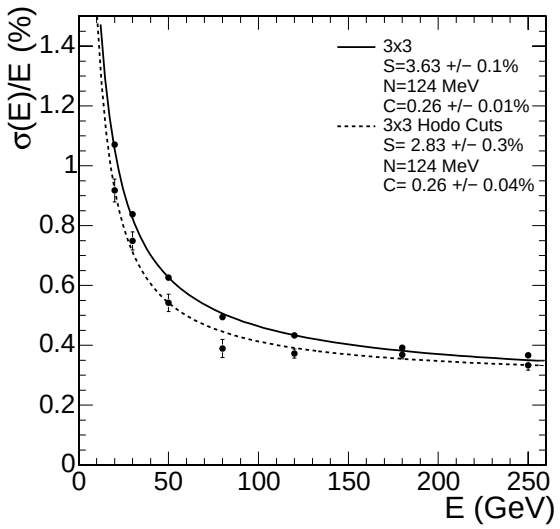


Figure 2.9: ECAL supermodule energy resolution, σ_E/E, as a function of electron energy as measured from a beam test. The upper series of points correspond to events taken with a 20×20 mm^2 trigger and reconstructed using a containment correction described in [38]. The lower series of points correspond to events selected to fall within a 4×4 mm^2 region. The energy was measured in an array of 3×3 crystals with electrons impacting the central crystal.

2.2.4 Hadronic calorimeter

The Hadronic Calorimeter (HCAL), plays an essential role in the identification and measurement of quarks, gluons, and neutrinos by measuring the energy and the direction of jets and of missing transverse energy flow in events. Missing energy forms a crucial signature of new particles, like the supersymmetric partners of quarks and gluons. For good missing energy resolution, a hermetic calorimetry coverage to $|\eta| = 5$ is required. The HCAL will also aid in the identification of electrons, photons and muons in conjunction with the tracker, the electromagnetic calorimeter, and muon systems.

The design of the hadron calorimeter (HCAL) [40] is strongly influenced by the choice of magnet parameters since most of the CMS calorimetry is located inside the magnet coil and surrounds the ECAL system. An important requirement of HCAL is to minimise the non-Gaussian tails in the energy resolution and to provide good containment and hermeticity for the E_T^{Miss} measurement. Hence, the HCAL design maximises material inside the magnet coil in terms of interaction lengths. This is complemented by an additional layer of scintillators, referred to as the hadron outer (HO) detector, lining the outside of the coil. The hadron barrel (HB) and hadron endcap (HE) calorimeters, shown in Fig. 2.10, are sampling calorimeters with 50 mm thick copper absorber plates which are interleaved with 4 mm thick scintillator sheets. Copper has been selected as the absorber material because of its density.

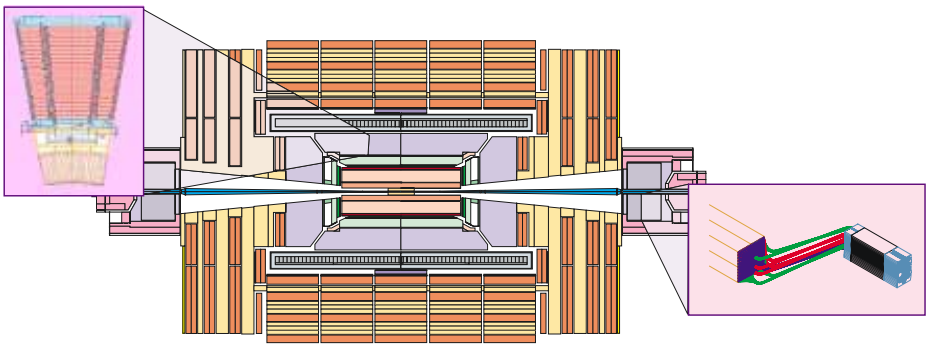


Figure 2.10: The CMS Hadronic Calorimeter

The HB is constructed of two half-barrels each of length 4.3 m. It consists of 32 towers covering the pseudorapidity region $-1.4 < \eta < 1.4$, resulting in 2304 towers with a segmentation $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ (Fig. 2.11 left). The HB is read out as a single longitudinal sampling. There are 15 brass plates, each with a thickness of about 5 cm, plus 2 external stainless steel plates for mechanical strength.

The hadron outer (HO) detector contains scintillators with a thickness of 10 mm, which line the outside of the outer vacuum tank of the coil and cover the region $-1.26 < \eta < 1.26$. The tiles are grouped in 30°-sectors, matching the ϕ segmentation of the DT chambers. They sample the energy from penetrating hadron showers leaking through the rear of the calorimeters and so serve as a "tail-catcher" after the magnet coil. They increase the effective thickness of the hadron calorimetry to over 10 interaction lengths, thus reducing the tails in the energy resolution function. The HO also improves the E_T^{Miss} resolution of the calorimeter.

The HE consists of two large structures, situated at each end of the barrel detector and within the region of high magnetic field. Each hadron endcap (HE) consists of 14 η towers with 5° ϕ segmentation, covering the pseudorapidity region $1.3 < |\eta| < 3.0$ (Fig. 2.11 right). For the 5 outermost towers (at smaller η) the ϕ segmentation is 5° and the η segmentation is 0.087. For the 8 innermost towers the ϕ segmentation is 10°, whilst the η segmentation varies from 0.09 to 0.35 at the highest η. The total number of HE towers is 2304.

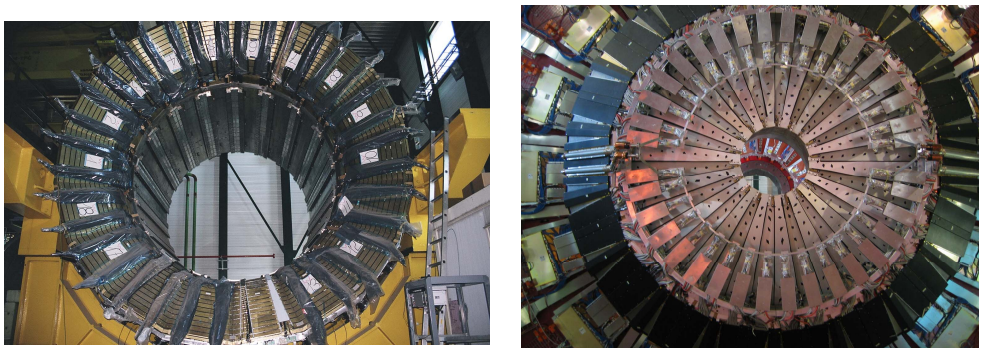


Figure 2.11: Left: Assembled hadron barrel HB. Right: Assembled hadron endcap HE in the surface hall at SX5.

To read the HCAL the blue-violet light emitted from the tiles, is absorbed by the wave shifting fibres which fluoresce in the green, then the waveshifted light is conveyed via clear fibre waveguides to hybrid photodiodes (HPDs).

There are two hadronic forward (HF) calorimeters, one located at each end of the CMS detector, which complete the HCAL coverage to $|\eta| = 5$. The HF detector, situated in a harsh radiation field, is built of steel absorber plates and radiation-resistant quartz fibres, of selected lengths, which are inserted into the absorber plates. The energy of jets is measured from the Cherenkov light signals produced as charged particles pass through the quartz

fibres. These signals result principally from the electromagnetic component of showers, which results in good directional information for jet reconstruction. The front face is located at 11.2 m from the interaction point. The depth of the absorber is 1.65 m. The diameter of the quartz fibres is 0.6 mm and they are placed 5 mm apart in a square grid. The quartz fibres, which run parallel to the beam line, have two different lengths (namely 1.43 m and 1.65 m) which are inserted into grooves, creating 2 effective longitudinal samplings. There are 13 towers in η, all with a size given by $\Delta\eta \approx 0.175$, except for the lowest-η tower with $\Delta\eta \approx 0.1$ and the highest-η tower with $\Delta\eta \approx 0.3$. The ϕ segmentation of all towers is 10°, except for the highest-η one which has $\Delta\phi = 20°$. This leads to 900 towers and 1800 channels in the 2 HF modules. Fibre optics convey the Cherenkov signals to photomultiplier tubes which are located in radiation shielded zones.

The performance of the HCAL is obtained comparing the simulated single particle energy response with test beam data from all 3 geographic parts of the HCAL. However, for gauging the performance of the HCAL, it is usual to look at the jet energy resolution and the missing transverse energy resolution. The granularity of the sampling in the 3 parts of the HCAL has been chosen such that the jet energy resolution, as a function of E_T, is similar in all 3 parts. This is illustrated in Figure 2.12 [38].

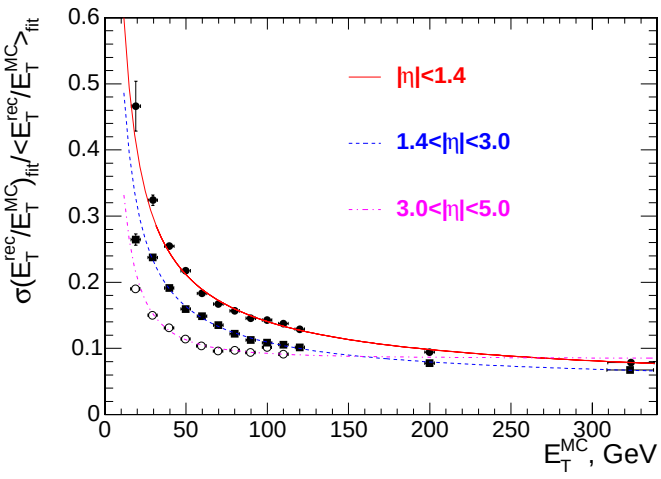


Figure 2.12: The jet transverse energy resolution as a function of the simulated jet transverse energy for barrel jets ($|\eta| < 1.4$), endcap jets ($1.4 < |\eta| < 3.0$) and very forward jets ($3.0 < |\eta| < 5.0$). The jets are reconstructed with the iterative cone $R = 0.5$ algorithm [38].

2.2.5 Muon system

The CMS detector is a general purpose detector specifically optimised for muon measurement. Three types of gaseous detectors are used to identify and measure muons [41]. The choice of the detector technologies has been driven by the very large surface to be covered and by the different radiation environments. In the barrel region ($|\eta| < 1.2$), where the neutron induced background is small, the muon rate is low and the residual magnetic field in the chambers is low, drift tube (DT) chambers are used. In the two endcaps, where the muon rate as well as the neutron induced background rate is high, and the magnetic field is also high, cathode strip chambers (CSC) are deployed and cover the region up to $|\eta| < 2.4$. In addition to this, resistive plate chambers (RPC) are used in both the barrel and the endcap regions. The layout of one quarter of the CMS muon system for initial low luminosity running is shown in Figure 2.13. In the Muon Barrel (MB)

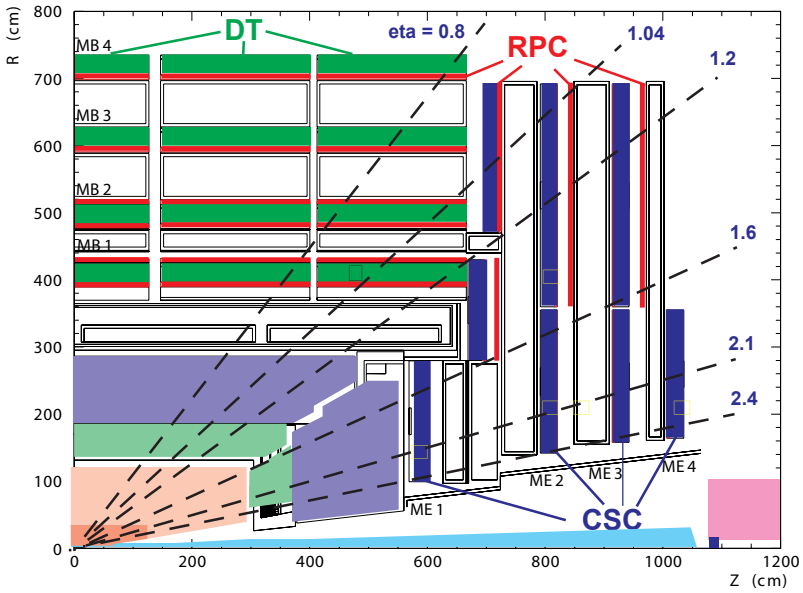


Figure 2.13: Layout of one quarter of the CMS muon system for initial low luminosity running. The RPC system is limited to $|\eta| < 1.6$ in the endcap, and for the CSC system only the inner ring of the ME4 chambers have been deployed.

region, 4 stations of detectors are arranged in cylinders interleaved with the iron yoke. The segmentation along the beam direction follows the 5 wheels of the yoke (labelled YB-2 for the farthest wheel in $-z$, and YB$+2$ for the

farthest is $+z$). In each of the endcaps, the CSCs and RPCs are arranged in 4 disks perpendicular to the beam, and in concentric rings, 3 rings in the innermost station, and 2 in the others. In total, the muon system contains order of 25 000 m^2 of active detection planes, and nearly 1 million electronic channels.

The Barrel Detector consists of 250 chambers organised in 4 layers (stations labeled MB1, MB2, MB3 and MB4 with the last being the outermost) inside the magnet return yoke, at radii of approximately 4.0, 4.9, 5.9 and 7.0 m from the beam axis. Each of the 5 wheels of the Barrel Detector is divided into 12 sectors, with each covering a 30° azimuthal angle. The MB1, 2 and 3 chambers consist of 12 planes of aluminium drift tubes; 4 r-ϕ measuring planes in each of the 2 outermost "superlayers," separated by about 20 cm and sandwiching a z-superlayer comprising 4 z-measuring planes (Fig. 2.14 left). The MB4 station does not contain the z-measuring planes. The maximum drift length is 2.0 cm and the single-point resolution

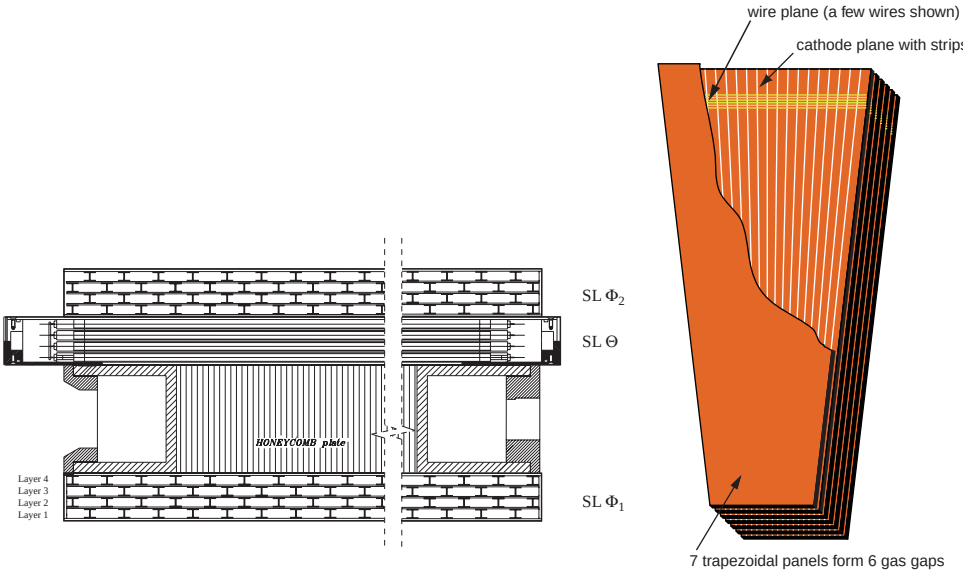


Figure 2.14: Left: the layout of a DT chamber inside a muon barrel station. Right: schematic view of a CSC chamber.

is $\approx 200\mu$m. Each station is designed to give a muon vector in space, with a ϕ precision better than 100μm in position and approximately 1 mrad in direction. Each DT chamber has 1 or 2 RPCs coupled to it before installation, depending on the station. In stations MB1 and MB2, each package consists of 1 DT chamber sandwiched between 2 RPCs. In stations MB3 and MB4, each package comprises 1 DT chamber and 1 RPC, which is placed on the innermost side of the station. A high-P_T muon thus crosses up to 6 RPCs

and 4 DT chambers, producing up to 44 measured points in the DT system from which a muon-track candidate can be built.

The Muon Endcap (ME) system comprises 468 CSCs in the 2 endcaps. CSCs are multiwire proportional chambers with segmented cathode readout. High precision coordinate along the wire is obtained by extrapolation of charges induced on several adjacent cathode strips. In CMS the strip width varies from 3.2 to 16 mm. Obtained resolution is in the range between 80 μm and 450 μm for one layer. CMS chambers have trapezoidal shape. One chamber consists of six detecting layers (Fig 2.14 right). The layers are separated by 16 mm thick polycarbonate plastic honeycomb panels which make the chamber stiff and provide a lever arm necessary to measure angle of the tracks. In each layer the strips are running radially. In angular units the strip width $\Delta\phi$ varies from 2.0 to 4.3 mrad and the length $\Delta\eta$ from 0.35 to 0.60 η-units. Combined off-line resolution of six layers approaches 50 μm. The wires are perpendicular to the strips, except ME1/1 where the wires are tilted by 25°. This is to compensate the Lorentz effect in high magnetic field (almost 4T) to which the chamber is exposed.

RPC chambers are located both in the barrel and in the endcaps. These RPCs are operated in avalanche mode to ensure good operation at high rates (up to 10 kHz/cm^2) and have double gaps with a gas gap of 2 mm. RPCs provide a fast response with good time resolution but with a coarser position resolution than the DTs or CSCs. RPCs can therefore identify unambiguously the correct bunch crossing.

Centrally produced muons are measured 3 times: in the inner tracker, after the coil, and in the return flux. Measurement of the momentum of muons using only the muon system is essentially determined by the muon bending angle at the exit of the 4 T coil, taking the interaction point (which will be known to $\approx 20\mu$m) as the origin of the muon. The resolution of this measurement (labelled "muon system only" in Figure 2.15) is dominated by multiple scattering in the material before the first muon station up to P_T values of 200 GeV, when the chamber spatial resolution starts to dominate. For low-momentum muons, the best momentum resolution (by an order of magnitude) is given by the resolution obtained in the silicon tracker ("inner tracker only" in Figure 2.15). However, the muon trajectory beyond the return yoke extrapolates back to the beam-line due to the compensation of the bend before and after the coil when multiple scattering and energy loss can be neglected. This fact can be used to improve the muon momentum resolution at high momentum when combining the inner tracker and muon detector measurements ("full system" in Figure 2.15) [38].

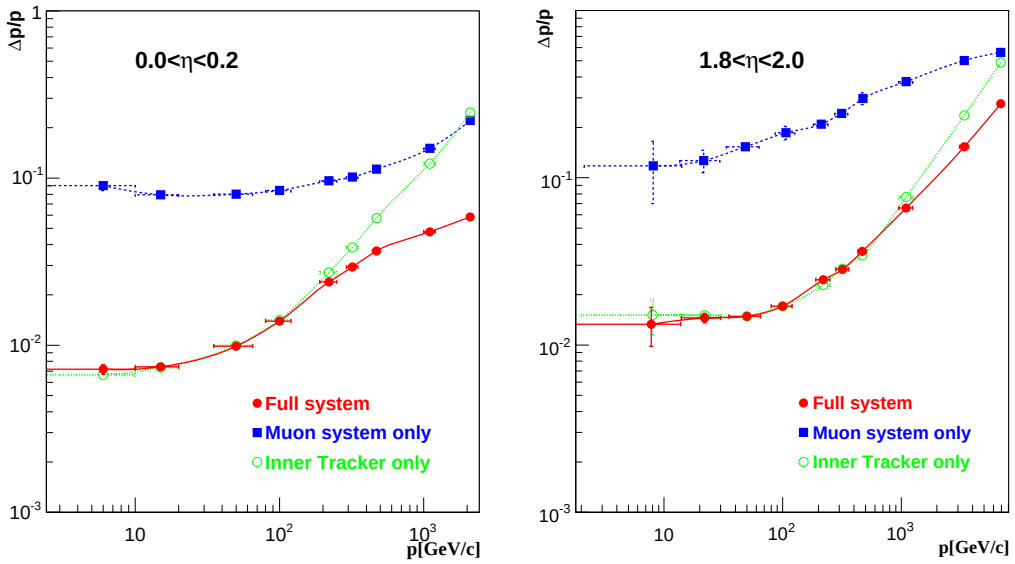


Figure 2.15: The muon momentum resolution versus p using the muon system only, the inner tracker only, or both ("full system"). a) barrel, $|\eta| < 0.2$; b) endcap, $1.8 < |\eta| < 2.0$.

2.3 The trigger system

The LHC bunch crossing rate of 40 MHz, considering 20 inelastic pp collisions at the nominal high luminosity (10^{34} cm^{-2}s^{-1}), leads to $\approx 10^9$ interactions per second. Because each event takes about 1 MB of zero-suppressed data and that the expected total storage capability for CMS is $O(10^2)$ MB/s, it is clear that a trigger reduction factor of $\sim 10^7$ is needed. To obtain this reduction factor the strategy adopted by CMS is to split the selection in two main entities. In the first step (Level-1 trigger) [42] a dedicated hardware is used to reduce at minimum the dead time and to take a very fast accept/reject decision to cut down from 40MHz to almost 100KHz the data rate. However at the LHC startup the CMS Level-1 output rate will be reduced to only 50 kHz for low luminosity (2×10^{33} cm^{-2}s^{-1}) and it will be raised to the designed 100 kHz at the full LHC luminosity. In case of positive decision, data are temporarily stored and passed to the second step, the High-Level trigger (HLT) system [43]. The CMS trigger scheme is represented in Fig. 2.16. The HLT relies on commercial processors, organised in a farm of personal computers. Many dedicated software algorithms will run to select events on physics basis and will represent the first step of physics analysis selection. The expectation that the sophistication of the algorithms

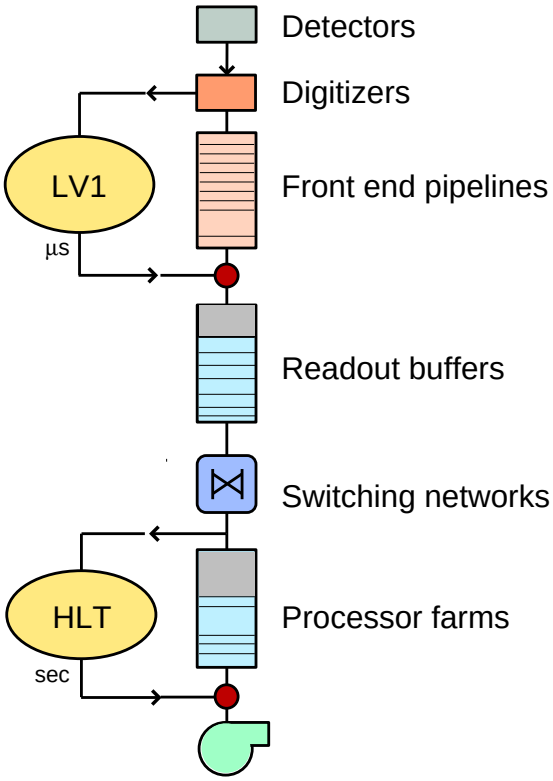


Figure 2.16: Data flow in the Trigger/DAQ system. The software-based High-Level trigger (HLT) filters the events after the hardware-based Level-1 trigger.

employed by the HLT will demand a mean processing time of $O(10^{-2})$s, along with the maximum event rate of 10^5 Hz, implies that of $O(10^3)$ processors must be employed for this processing stage.

The Level-1 trigger selection is based exclusively on calorimeter and muon chamber information, processed with hardware logical circuits, though with coarse granularity. The time available for processing in the Level-1 trigger is limited by the resources available in the front-end electronics that store the detector data during the Level-1 decision-making process. Channel count and technology considerations dictate that some of the front-end electronics can store the data from at most 128 continuous bunch crossing, i.e. the equivalent of approximately 3 μs of data. This 3 μs is the maximum time interval for a Level-1 decision to be received by the front end electronics. Because they include the latency of the propagation of the information, the

resulting time available for actually processing the data is no more than 1 μs. For this reason the Level-1 trigger can only processes data from a subset of CMS subdetectors. The two elements of the Level-1 trigger, calorimetry and muon system, work in parallel and analyse the data locally, combine the information and produce the output passed to the Data Acquisition System. The calorimetric information is synthesised into calorimetric towers by individual Trigger Primitive Generator (TPG) circuits for ECAL, HCAL and HF. The towers are clusters of signals and are sent to the Regional Calorimeter Trigger (RCT), which combines the tower information to reconstruct jets, and leptons/photons. All these objects have a raw measurement of energy and position and are collected by the Global Calorimeter Trigger (GCT). The Level-1 calorimetric trigger also provides a map of inactive calorimeter regions to improve the isolation of the muons, which are reconstructed in the meantime independently by the two muon subsystems, RPC and DT+CSC. The reconstructed muon segments are combined together by the Global Muon Trigger (GMT), implemented with a programmable logic. The GMT resolves ambiguities and uses inactive calorimetric regions to remove fakes and finally passes the Level-1 muon candidates to the Global Trigger. The information of GCT and GMT passed to the Global Trigger (GT) is combined to provide a first estimation of the missing transverse energy E_T^{miss} and determine some regions where the HLT should focus on.

The allocation of the Level-1 Trigger bandwidth must be optimised in order to ensure the widest possible physics reach for the experiment, while including all technical triggers intended for calibration and monitoring of the detector. The procedure used for this optimisation starts with a determination of the maximum Level-1 trigger rate that can be allocated, which is 16 kHz at low luminosity and 33 kHz at high luminosity. This safety factor of three respect to the total bandwidth is used to account for all uncertainties in the simulation of the basic physics process, the CMS detector, and the beam conditions. The Level-1 Trigger table that will be used for analysis is presented in Tab 2.17 for low luminosity and in Tab 2.18 for high luminosity scenario. As we can see from the tables an amount of about 1 kHz is left for calibration and monitoring purposes with random triggers to obtain a uniform sample of Minimum Bias events.

At the High-Level trigger (HLT) selection step the pixel hits are available together with tracker signals, hence primary vertex reconstruction and track finding should be possible using algorithm similar to the offline one even if with some limitation in number of hits and in the region of tracker where the hits are searched. These conditions, called partial and regional track reconstruction, are fundamental to reduce the time to the track building. Hence in the HLT will be algorithms both to reconstruct better the raw

Trigger	Threshold [GeV or GeV/c]	Expected Rate [kHz]	Cumulative Rate [kHz]
Inclusive isolated e/γ	29	3.3	3.3
ee/$\gamma\gamma$	17	1.3	4.3
Inclusive μ	14	2.7	7.0
$\mu\mu$	3	0.9	7.9
Single τ-jet	86	2.2	10.1
Two τ-jets	59	1.0	10.9
1 jet	177	1.0	11.4
3 jets OR 4 jets	86,70	2.0	12.5
1 jet AND E_{T}^{miss}	88$\otimes$46	2.3	14.3
e AND jet	21$\otimes$45	0.8	15.1
Minimum Bias (calibration)		0.9	16.0
Total			16.0

Figure 2.17: Level-1 trigger table at low luminosity.

Trigger	Threshold [GeV or GeV/c]	Expected Rate [kHz]	Cumulative Rate [kHz]
Inclusive isolated e/γ	34	6.5	6.5
ee/$\gamma\gamma$	19	3.3	9.4
Inclusive μ	20	6.2	15.6
$\mu\mu$	5	1.7	17.3
Single τ-jet	101	5.3	22.6
Two τ-jets	67	3.6	25.0
1 jet	250	1.0	25.6
3 jets OR 4 jets	110,95	2.0	26.7
1 jet AND E_{T}^{miss}	113$\otimes$70	4.5	30.4
e AND jet	25$\otimes$52	1.3	31.7
μ AND jet	15$\otimes$40	0.8	32.5
Minimum Bias (calibration)		1.0	33.5
Total			33.5

Figure 2.18: Level-1 trigger table at high luminosity.

Level-1 physics objects together with streams dedicated to particular physics channel identification, for instance related to b physics using the tracker

information. The summary of the High-Level trigger threshold and rate are reported in Tab. 2.19 for low luminosity scenario.

Trigger	Threshold [GeV or GeV/c]	Expected Rate [Hz]	Cumulative Rate [Hz]
Inclusive e	29	33	33
ee	17	1	34
Inclusive γ	80	4	38
$\gamma\gamma$	40$\otimes$25	5	43
Inclusive μ	19	25	68
$\mu\mu$	7	4	72
Inclusive τ-jets	86	3	75
Two τ-jets	59	1	76
1 jet AND E_{T}^{miss}	180$\otimes$123	5	81
1 jet OR 3 jet OR 4 jet	657,247,113	9	89
e AND jet	19$\otimes$52	1	90
Inclusive b-jets	237	5	95
Calibration and other events (10%)		10	105
Total			105

Figure 2.19: High-Level trigger requirements at low luminosity. There is no actual threshold in the HLT selection for τ-jets, so the threshold shown is that of the corresponding Level-1 trigger requirement.

Chapter 3

Leptons

In the Higgs production associated to a $t\bar{t}$ pair the presence of a lepton from $t \to bW \to bl\nu$ is very important both for trigger purpose and for top reconstruction. When the wrong lepton in the final state of the event is matched with the b-quark jet, the inferred information on the four-momentum of the top quark is randomly biased and therefore useless. Moreover in $t\bar{t}H$ events the wrong identification of the lepton can lead to the wrong identification of the b-quark jet matched for the top reconstruction and then to the wrong b-quark jets associated to the Higgs boson. In $t\bar{t}H$ the presence of extra leptons, besides that one coming from W decay, is enhanced by the semileptonic decay of b-quarks. Hence the probability to find more than one lepton in semileptonic decay of $t\bar{t}H \to bWbWH \to bq\bar{q}\bar{b}l\nu b\bar{b}$ is not negligible.

In this chapter the construction of a method which can efficiently identify the correct lepton, being electron or muon, will be described.

3.1 Muons

3.1.1 Muon reconstruction

For this analysis muons are reconstructed using the official CMS experiment software and the algorithm called "global muon reconstruction" [38, 44]. The global/Level-3 muon reconstruction consists in extending the muon trajectories to include hits in the silicon tracker.

Starting from a standalone reconstructed muon, the muon trajectory is extrapolated from the innermost muon station to the outer tracker surface, taking into account the muon energy loss in the material and the effect of multiple scattering. Silicon layers compatible with the muon trajectory are then determined, and a region of interest within them is defined in which to

perform regional track reconstruction. The determination of the region of interest is based on the track parameters and their corresponding uncertainties of the extrapolated muon trajectory, obtained with the assumption that the muon originates from the interaction point. This constraint is removed at the end of the track fitting, when all the hits in the muon chambers from the original standalone reconstruction together with the hits in the silicon tracker are used.

To study the performance of this algorithm a simulated single-muon samples with both fixed and continuous values of p_T (between 10 GeV and 1 TeV), and with a flat distribution in η and ϕ are used. In Fig. 3.1 the reconstruction efficiency and the resolution ($\sigma(1/p)$) are shown on the left and on the right respectively.

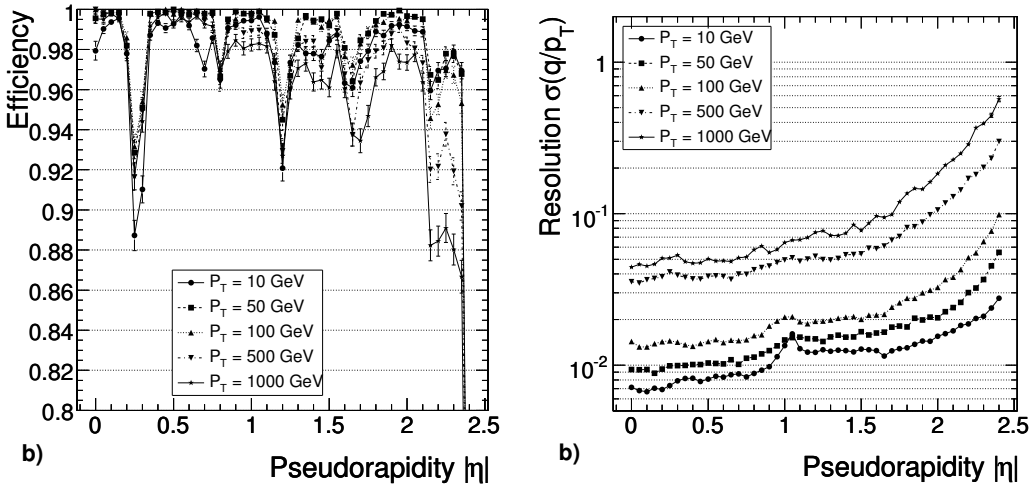


Figure 3.1: "Global muon reconstruction algorithm". Left: muon reconstruction efficiency as a function of pseudorapidity for various values of p_T. Right: the $1/p_T$ resolution as a function of pseudorapidity.

The reconstruction efficiency achieved is typically 95–99%, except in the pseudorapidity regions around $|\eta| = 0.25$ and $|\eta| = 0.8$: regions between 2 DT wheels, and around $|\eta| = 1.2$: transition region between the DT and CSC chambers, where the efficiency drops. The transverse momentum resolution was obtained by a Gaussian fit to the distribution of the quantity:

$$\frac{q^{rec}/p_T^{rec} - q^{gen}/p_T^{gen}}{q^{gen}/p_T^{gen}}$$

where q is the charge and p_T^{rec} and p_T^{gen} are the reconstructed and generated transverse momenta, respectively.

3.1.2 Muon identification

The muon coming from W have a transverse momentum peak around 30 GeV (Fig. 3.2) while in the momentum range between 10–30 GeV, the main source of muons are from b and c decays. Moreover, for low-p_T muons,

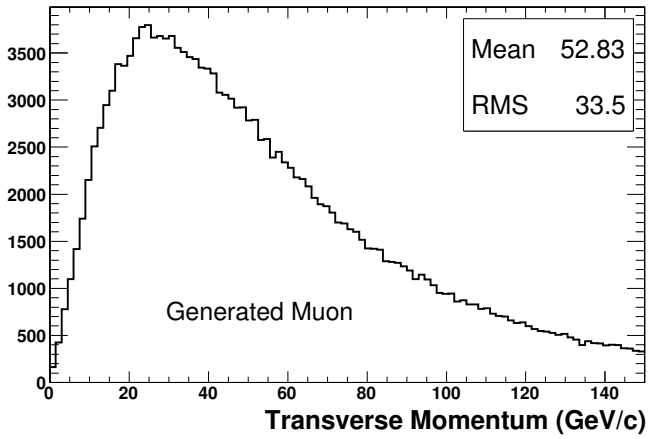


Figure 3.2: Transverse momentum distribution of the generated muons coming from W decay.

another important contribution is given by muonic K and π decays. To distinguish muons from W, which will be referred to as *signal* muons, from muons coming from other sources, defined as *background* muons, several variables have been considered [31].

The muon p_T can not be the only variable used to distinguish background from signal muons, however it plays an important role because high p_T muons have a larger probability to come from W with respect to semileptonic b and c decays.

The main observable to distinguish signal from background muons is the energy collected around the muon direction. The background muons are in fact produced in usual soft jets and thus accompanied by nearby particles, while signal muons are mainly isolated or accompanied only by particles from pile-up and by uncorrelated particles from the underlying event.

The isolation algorithms that have been used rely on the comparison of the total energy deposited in a cone around the muon. Both the transverse energy deposited in the calorimeter ($IsoCalo$) and the sum of the transverse momenta of reconstructed charged-particle tracks ($IsoTk$) are used for this purpose (Fig. 3.3). Another isolation observable that has been tried is the distance between the muon and the closest jet: $\Delta R_{\mu-jet}$. This variable however has a high correlation with the $IsoCalo$ variable and using them at the

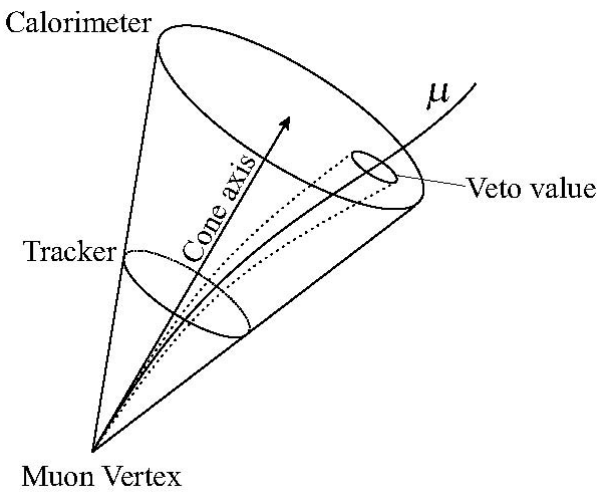


Figure 3.3: Schematic illustration of the isolation cone. The energy deposit ($\sum p_T^{Tk}$, $\sum E_T^{Calo}$) is computed in the isolation cone, the muon contribution is removed by excluding a small area around the muon (veto cone).

same time does not help in discriminating signal and background muons.

The last observable considered is the significance on the transverse impact parameter of the muon track. This helps because muon from b and c quarks often have an higher impact parameter with respect to the primary vertex of the interaction. Summarising, the observables used to identify signal from background muons are:

- Transverse momentum, p_T
- Track isolation, $IsoTk$
- Calorimeter isolation, $IsoCalo$
- Distance between the muon and the closest jet, $\Delta R_{\mu-jet}$
- Significance of the track impact parameter, $S_{ip} = d/\sigma_d$

The probability distribution functions (PDFs) associated with these variables for signal and background muons are obtained by matching to generated muons and they are shown in Fig. 3.4.

The PDFs are constructed using a sample of $t\bar{t}H$ events with $M_H = 120$ GeV in which one and only one of the W bosons decays to a muon and neutrino, while the other one decay hadronically. A reconstructed muon is identified as originating from a W boson decay when the separation ($\Delta R =$

$\sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$) from the generated muon in azimuth ϕ and pseudorapidity η space is less than 0.01. In this way from $t\bar{t}H$ events two samples of signal and background muons are extracted.

Each of the selected variables provide good discriminating power between signal and background muons (Fig. 3.4). The PDFs are combined into the following likelihood ratio.

$$L = \Pi_i \frac{P_i^{sig}(x_i)}{P_i^{sig}(x_i) + P_i^{bkg}(x_i)} \tag{3.1}$$

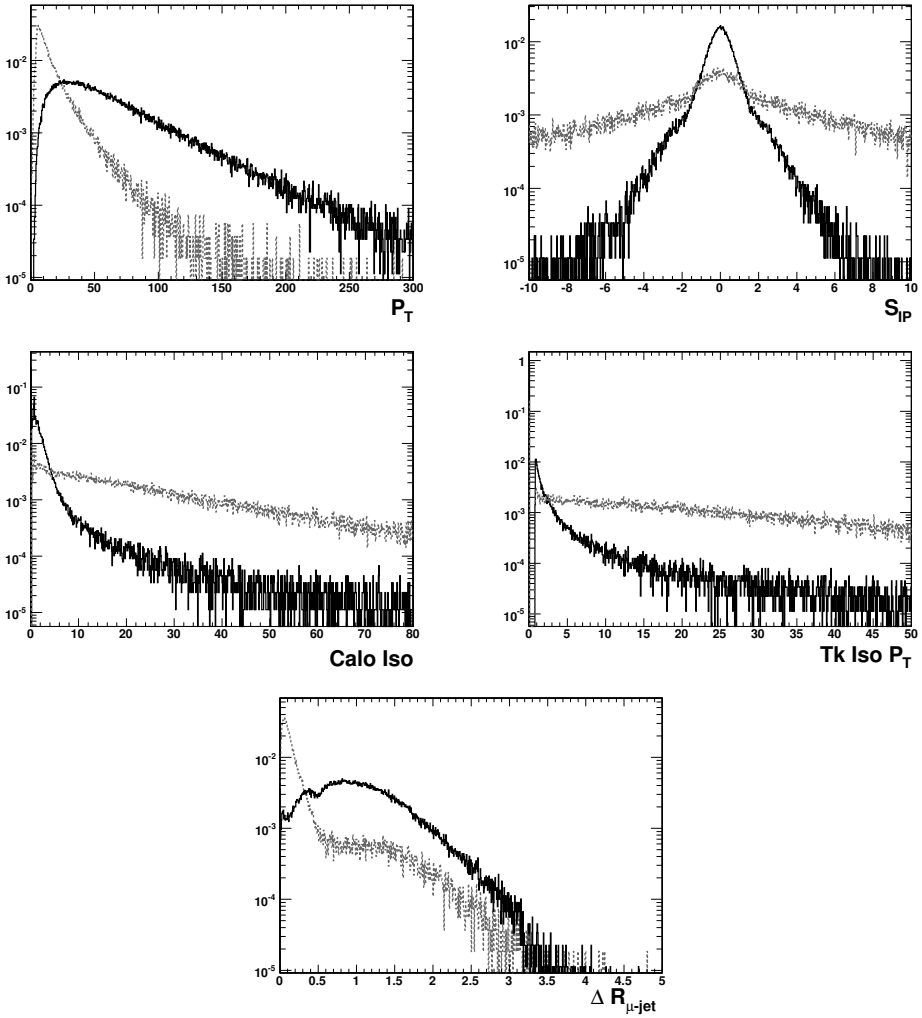


Figure 3.4: Probability distribution functions used in the construction of the muon likelihood: the black line refers to signal muons while the grey line refers to background muons, as defined in the text.

where P_i^{sig} and P_i^{bkg} are the PDF of the observable x_i for signal and background muons, respectively. Different combinations of the PDFs have been tested combining three or four variables. The performance obtained are shown on Fig. 3.5. In the best case at signal muon efficiency of 90%, only 1% of background muons are selected. It has been obtained using the combination of p_T, $IsoTk$, $IsoCalo$, $S_{ip} = d/\sigma_d$. The corresponding minus logarithmic distribution of the combined likelihood for signal and background muon is shown in Fig 3.6.

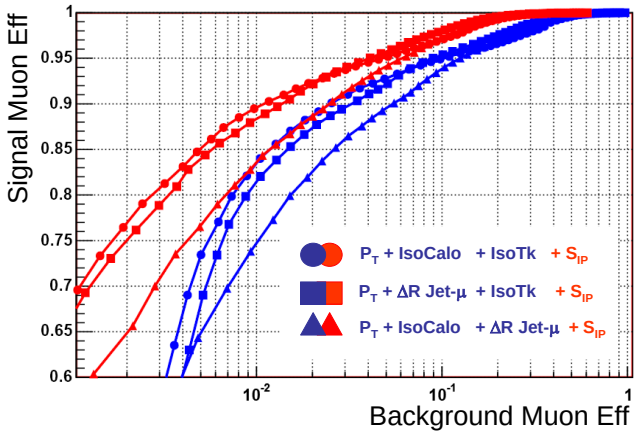


Figure 3.5: Performance of the muon likelihood discriminator for the signal and background sample obtained from $t\bar{t}H$ events. Different combinations of PDFs have been tried both with three and four variables.

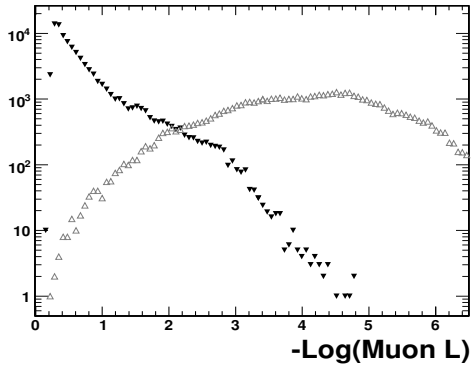


Figure 3.6: Distribution of $-Log$ of muons likelihood for the best discriminating combination for signal and background, p_T, $IsoTk$, $IsoCalo$, $S_{ip} = d/\sigma_d$.

To explore the performance of the likelihood ratio method the $t\bar{t}H$ muon selection efficiency versus the QCD ($\hat{p}_t$ >170 GeV) efficiency has been also studied. Here the performance of the likelihood selection has been compared to the performance of the HLT single muon stream selection. In this case the signal efficiency do not reach the value of 100% because part of the muon from W decay can go out of the muon chamber acceptance or because a little part of the generated muons can not be reconstructed. The single step reconstruction efficiency is shown in Tab. 3.1.

	ϵ %
Generated muons inside the chambers acceptance	92.5
Events with at least 1 muon reconstructed	91.4
Events with one reconstructed muon matched to generated one	90.2

Table 3.1: Muon reconstruction effieciency in the semi-leptonic muon $t\bar{t}H$ channel.

Fig. 3.7 shows that if the likelihood selection is used in place of the HLT, at the same QCD efficiency the likelihood gives a signal efficiency of 74% compared to 63% for the HLT single muon stream. Alternatively, if the likelihood selection is used after the HLT, dramatic improvement in QCD rejection is possible at little or no loss in signal efficiency. For example, a small drop in signal efficiency from 63% to 60% lowers the QCD efficiency by more than a factor of 3 (i.e. from 0.07% to 0.02%).

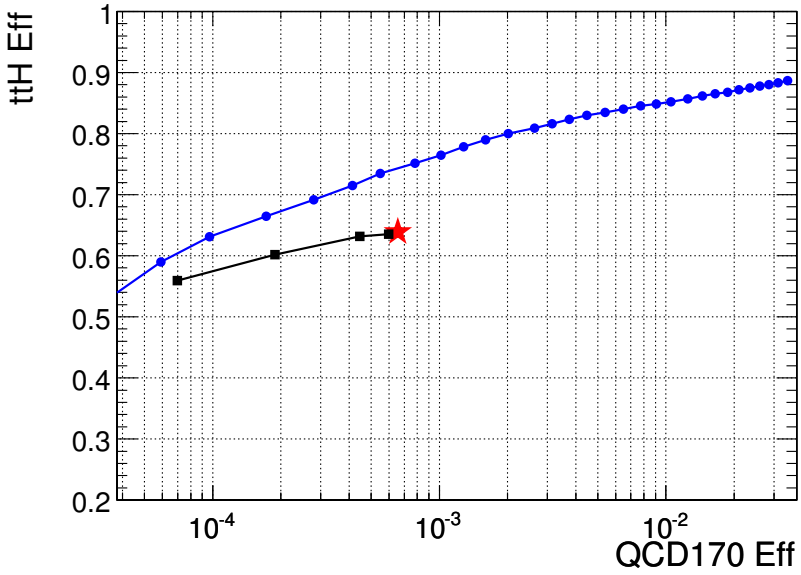


Figure 3.7: $t\bar{t}H$ signal efficiency versus QCD ($\hat{p}_t$ >170 GeV) efficiency. Circles: likelihood performance without HLT selection; Star: HLT selection; Squares: likelihood performance after the HLT selection.

3.2 Electrons

3.2.1 Electron reconstruction

Electron shower deposit its energy in several crystals in the ECAL. Approximately 94% of the incident energy of a single electron is contained in 3×3 crystals, and 97% in 5×5 crystals. Summing the energy measured in such fixed arrays give the best performance for electrons in test beam. The presence of material in front of the calorimeter results in bremsstrahlung and photon conversions. Because of the strong magnetic field the energy reaching the calorimeter is spread in ϕ. The bremsstrahlung effect is highlighted in Fig. 3.8: 30% of the electron with a transverse energy below 50 GeV, lost the 70% of their initial energy. The spread energy is clustered by building a cluster of clusters, a "supercluster", which is extended in ϕ. This algorithm to reconstruct electrons in CMS is called the "Island algorithm" [45]. It starts by a search for seeds which are defined as crystals with an energy above a certain threshold. Starting from the seed position, adjacent crystals are examined, scanning first in η and then in ϕ. Along each scan line, crystals are added to the cluster until a rise in energy or crystal that has not been read out is encountered. In much the same way as energy is clustered at the level

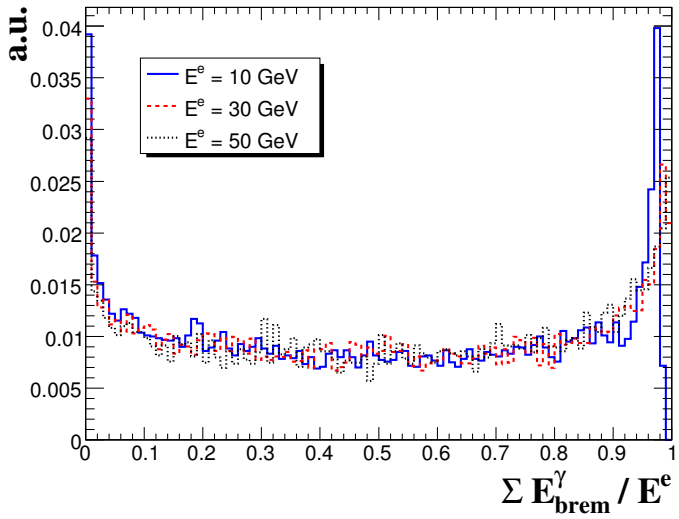


Figure 3.8: Distribution of the fraction, $\sum E^{\gamma}_{brem}/E^{e}$, of generated electron energy (E^{e}) radiated as bremsstrahlung photons for electrons of 10,20 and 50 GeV.

of calorimeter cells or crystals, non-overlapping Island clusters can in turn be clustered into superclusters. The procedure is seeded by searching for the most energetic cluster and then collecting all the other nearby clusters in a very narrow η-window, and a much wider ϕ-window.

After the calorimeter reconstruction the electron trajectory building is matched with the tracker, with a procedure similar to that one described for muons. The electromagnetic superclusters drive the finding of the track hits ("seeds") in the pixel detector. Hits in the pixel layers are predicted by propagation of the energy weighted mean position of the supercluster backward through the magnetic field under both charge hypotheses toward the pixel detector. The efficiency of the pixel finding for single electrons, averaged over the full ECAL barrel and endcaps range, reaches about 90% for electrons at $p_T = 10$ GeV. The 2 pixel hits found serve as seeds for the building and fitting of electron tracks in the Silicon Tracker Detectors.

The track reconstruction used in this work is based on a simple Kalman Filter procedure [46]. However, it has been demonstrated in CMS that a dedicated track reconstruction strategies for electrons that are affected by non-Gaussian fluctuations due to bremsstrahlung emission along their trajectory, can slightly improve the η, ϕ measurements for $p_T = 5$ GeV up to at least 30 GeV [38]. The electron track reconstruction efficiency as a function of p_T and the electron σ/E_e performance are shown in Fig. 3.9.

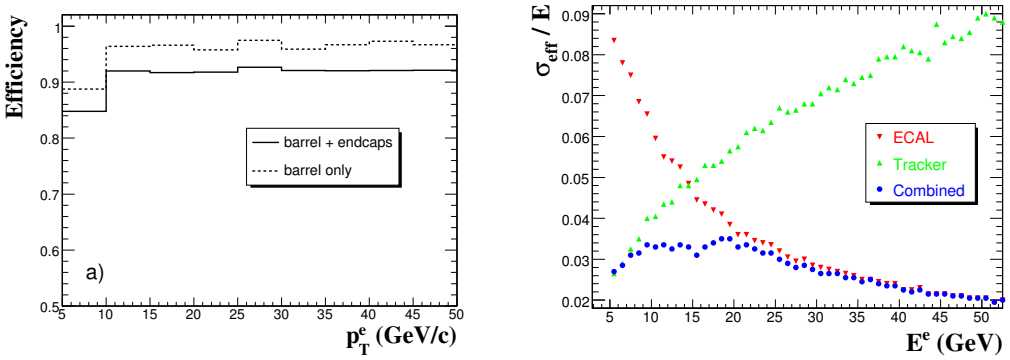


Figure 3.9: Left: Electron track reconstruction efficiency as a function of the electron p_T. Right: Energy resolution for electrons measured with only supercluster energy (downward triangles), reconstructed track momentum (upward triangles) or combined information (circles)

3.2.2 Electron identification

The identification of electrons coming from W decays respect to the electrons coming mainly from b, c, K and π is based on the same arguments of the muon identification described in Par. 3.1.2 [31].

Electrons coming from the W boson decays are typically characterised by isolated high transverse energy clusters. These electrons are thus efficiently identified by means of an isolation requirement applied to the electron candidate with respect to other reconstructed tracks in the event.

Isolation is defined by means of two variables. The first is the transverse momentum p_T sum of tracks inside an *isolation cone* of radius $\Delta R = 0.3$ in the η-ϕ plane around the candidate electron's direction, as defined by the calorimeter. A *veto cone* of radius $\Delta R = 0.015$ around the electron direction is defined in order to exclude the electron p_T from this sum. The second variable is the distance $\Delta R = \sqrt{((\phi_{ele} - \phi_{track})^2 + (\eta_{ele} - \eta_{track})^2)}$ between the electron candidate and the closest track in the $\eta - \phi$ plane outside the *veto cone*.

To further optimise the electron selection, three additional variables are considered:

- transverse momentum of the electron candidate, p_t;
- the ratio between the cluster energy and the track momentum, E/p;
- the ratio between the hadronic and the electromagnetic energy of the cluster, H/E.

All these five discriminating variables are used in order to construct a likelihood hypothesis for an isolated electron. The relevant PDFs are constructed in the form of reference histograms and combined into a likelihood in a manner completely analogous to that which was described earlier for muons in Eq 3.1. Also, as in the muon case, the reference histograms for signal and background electrons are again constructed by matching to generator level electrons in $t\bar{t}H$ events with Ws decaying semileptonically. Unit-normalised distributions for signal and background electrons are compared in Fig. 3.10. Proceeding from Figure 3.10a to 3.10e, distributions of p_t, E/P, H/E, $p_t isolation$ and ΔR are displayed respectively. The $-Log(L_e)$ curves

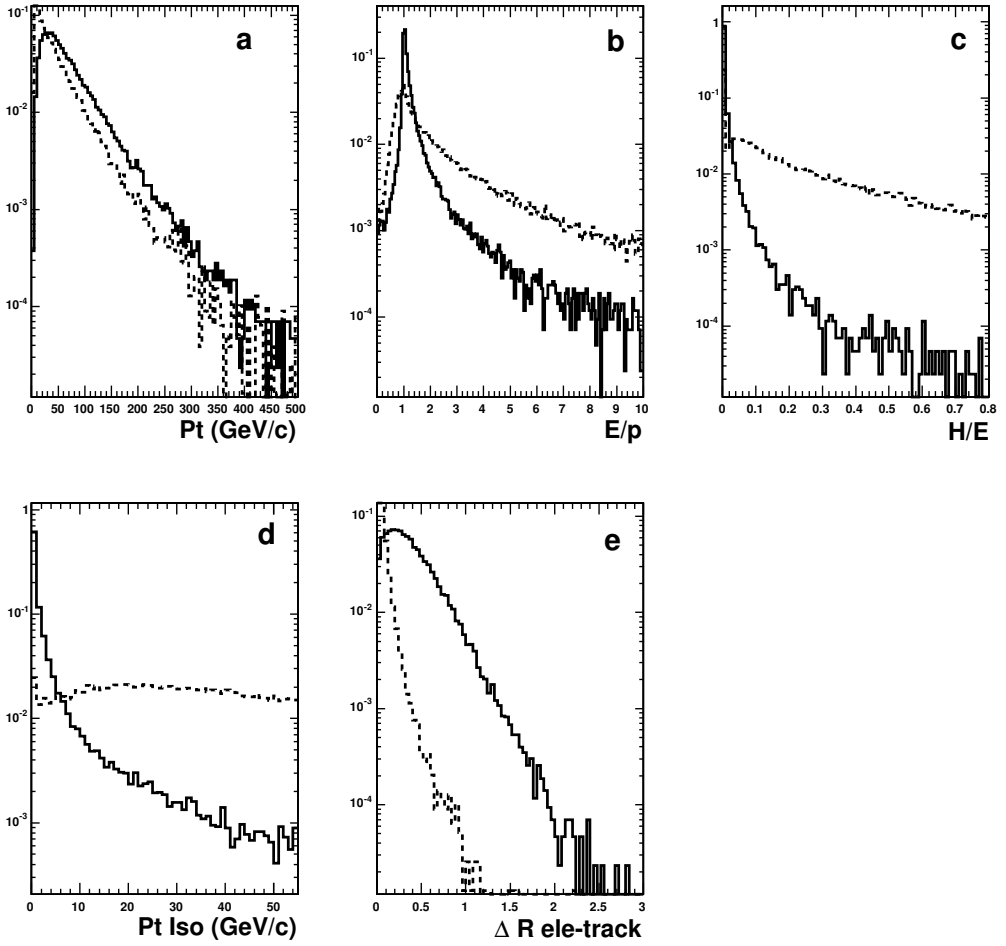


Figure 3.10: Normalized distributions of the five electron variables combined in the likelihood.

for signal (continuous line) and background electrons (dashed line) are su-

perimposed on the left of Figure 3.11.

An appropriate choice of likelihood cut value has been studied by comparing signal versus background electron efficiencies as shown on right of Fig. 3.11. Figure 3.11 shows, for example, that it is possible to achieved

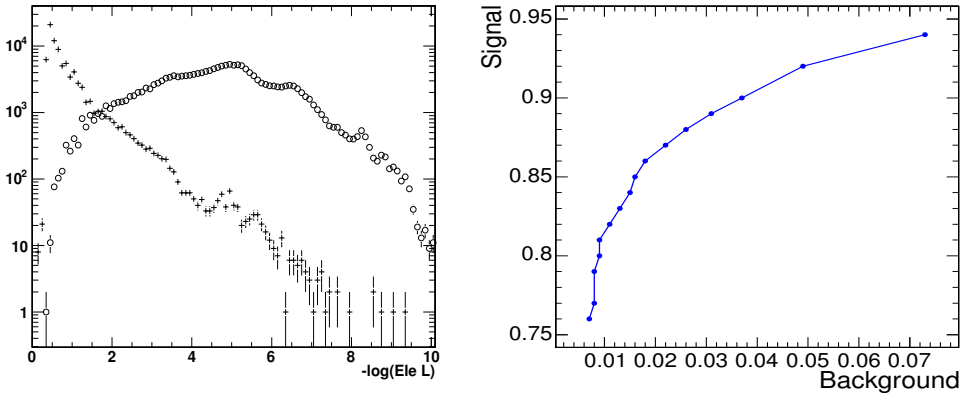


Figure 3.11: Left: $-Log$ of the electron likelihood distributions for *signal* (crosses) and *background* (circles) electrons. Right: signal versus background electron efficiencies for likelihood values ranging from 0.006 (the upper point) with a step size of 0.006, (i.e. approximately in the range $1.0 < -Log(L_e) < 2.0$.

a 90% signal electrons efficiency against only 3.7% of electron background efficiency.

Also in the electron case the performance of the likelihood method has been evaluated using a sample of QCD background with $\hat{p}_t$ >170 GeV. In Figure 3.12 the signal efficiency is plotted versus QCD background electron efficiency for various choices of likelihood cut values. The star identifies the efficiency levels obtained by applying only the HLT trigger selection. We have to remark that in the electron case the changes carried out in the HLT selection in [47] are not included in this work because not available in the official CMS software at the time when this studied has done. In particular, as in the electron HLT described in the "TDR Data Acquisition and High Level Trigger" [43] and used in this work, no tracker isolation is required at HLT level for low luminosity scenario. This is highlighted in Fig. 3.12 where it is possible to note that the electron efficiency for QCD with $\hat{p}_t$ >170 GeV is quite higher than (2.1%) the muon selection efficiency (0.07%) that uses isolation at HLT. Moreover in the same figure it is possible to note that the likelihood approach, if used in place of the HLT, increase the $t\bar{t}H$ efficiency from 54% to 88%.

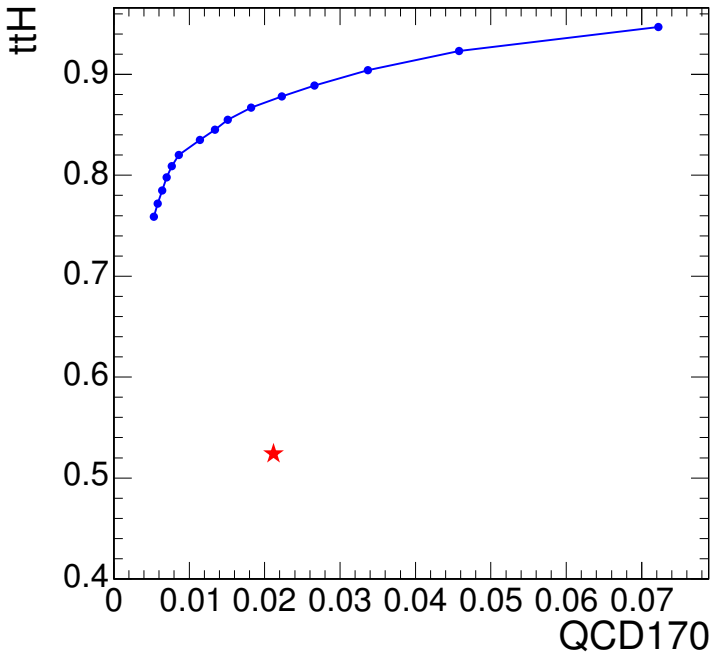


Figure 3.12: Signal efficiency versus QCD efficiency for likelihood cut values ranging from 0.006 (the upper point) with a step of 0.006 (i.e. approximately in the range $1.0 < -Log(L_e) < 2.0$).

Furthermore, as in the muon sample, maintaining roughly constant the signal efficiency, the likelihood cut in combination with the HLT trigger yields a reduction in the QCD background selection efficiency, in this case of about an order of magnitude.

Chapter 4

Jets and Missing Transverse Energy

A wide spectrum of new physics topologies, as well as known processes like top quark production, will have quarks in the final state of proton-proton collisions at the LHC. When reconstructing the quark's kinematics, jet reconstruction is of major importance. Ambiguities in the jet definition do not only arise from the theoretical point of view if higher order corrections are taken into account, but also experimentally, due to the magnetic field, the calorimeter response and different jet clustering algorithms. Moreover in all sample in which one W decay leptonically, the missing transverse energy measure plays an important role for the correct reconstruction of the top invariant mass.

In this chapter different aspects of jets will be discussed and analysed. In particular this study concentrates on the algorithmic task of clustering the input object(e.g. simulated particles or calorimeter cells) into jets, on the jet calibration and on the missing transverse energy measurement.

It is clear that to optimise the reconstruction of the $t\bar{t}H$ channel, jets play a fundamental role.

4.1 Introduction

The jet formation can be factorised in two main step [48]:

- The perturbative QCD step: shower evolution
- The non-perturbative QCD step: hadronization

In the perturbative QCD step, starting from a basic $2 \rightarrow 2$ process this kind of correction will generate $2 \rightarrow 3$, $2 \rightarrow 4$ and so on, final-state topologies.

The perturbative correction to objects, like quarks or gluons, both in the initial (ISR) and final (FSR) state is done mainly using two approaches. The first one is the matrix-element method (ME), in which Feynman diagrams are calculated, order by order. The problem of this approach is that calculations become increasingly difficult in higher orders corrections and divergences arise when soft or collimate partons are emitted from the primary one. The second possible approach is the parton-shower (PS). Here an arbitrary number of branching of one parton into two, three or more may be combined, to yield a description of multijets events. Because in this case the full matrix-element expression are not used, no explicit upper limit on the number of partons emitted could be used, giving thus a good description of the substructure of jets. However the limit of the shower approach is that it has limited predictive power for the rate of well-separated jets. In practise, especially in the last years, an hybrid approach is used trying to produce in the matrix element step, hard and well separated partons and in the shower step the description of the substructure of jets. In this case difficulties arise in possible double parton emission, both from the ME step and from the PS step. Some Monte Carlo generators try to recover the problem using matching criteria between the emitted PS and ME partons, splitting the phase-space into two parts [49, 50].
When the distance between quarks and gluons in the final state increase, QCD becomes strongly interacting and perturbation theory breaks down. In this confinement regime, the coloured partons are transformed into colourless hadrons, a process called either hadronization or fragmentation. Because the fragmentation process has yet to be understood starting from the QCD Lagrangian, the description of this part is given by different phenomenological models. The main models used for this step are: the string fragmentation (SF), used in this work, the independent fragmentation (IF) and the cluster fragmentation (CF). When hadrons in the final state have been formed the last step of the jet modelling is the decay of the unstable particles. The stable hadrons are then detected mainly by the hadronic and electromagnetic calorimeter and their energy and direction is measured. In Fig. 4.1 the shower and hadronization step are schematized for a $t\bar{t}H$ event. From this brief description of the jet modelling, two main aspects arise from the $t\bar{t}H$ reconstruction point of view. How much is the efficiency to reconstruct an event with n partons in the final state? How much the energy collected in the calorimeter have to be corrected to have the jet energy close to the primary parton energy?

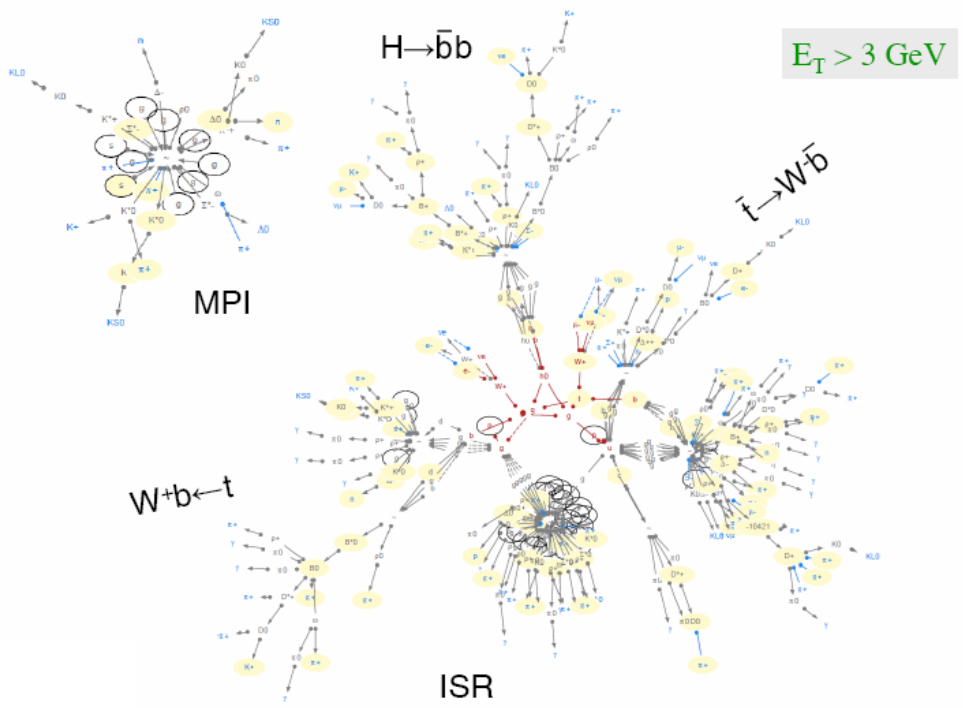


Figure 4.1: A schematic view of the parton shower and hadronization in a $t\bar{t}H$ events with Ws decaying leptonically.

4.2 Study of jet clustering algorithm

To answer the first question of the previous section a study on jet clustering is performed in the context of an analysis perspective [51], which means that the jet clustering is considered to be optimal if the reconstruction efficiency of the complete kinematics of the primary quark event topology is maximised. This reconstruction efficiency will be defined in terms of quality criteria or quality markers, relative to the performance of a typical analysis like the reconstruction of resonance masses. The distance between the generated quark i and the reconstructed jet j, and therefore the error of the jets, should be minimised in both energy and momentum (angular) space, for example $\epsilon_\theta = \theta_j^{jet} - \theta_i^{quark}$.

Physics effects like pile-up (PU) [1], underlying event (UE) [2] enlarge this

[1]Multiple interactions in a bunch crossing but in different proton-proton collisions.

[2]The UE in a hadron-hadron interaction is everything else accompanying the hard scattering component of the collision. The main sources of UE are initial (ISR) and final (FSR) state gluon radiation and multiple parton interactions in the same proton-proton collision (MPI)

mean error. The scope of this study is to find the most efficient jet finding setup in the presence of these effects, in order to maximise the fraction of events for which all quarks have smaller errors than some predefined criteria. Hence, events suffering from a large amount of hard gluon radiation will be rejected.

To disentangle detector effects from pure algorithmic and physics effects, the study is performed with simulated particle information as input to the jet finding algorithms.

4.2.1 Jet clustering algorithm

The following jet reconstruction algorithms are considered in this study: the *iterative cone* algorithm (IC), the inclusive k_T algorithm (k_T) and the *Mid-Point Cone* algorithm (MC) [52].

In the iterative cone algorithm, an E_T-ordered list of input objects (particles or calorimeter towers) is created. A cone of size R in η, ϕ space is cast around the input object having the largest transverse energy above a specified seed threshold. The objects inside the cone are used to calculate a proto-jet direction and energy. The computed direction is used to seed a new proto-jet. The procedure is repeated until the energy of the proto-jet changes by less than 1% between iterations and the direction of the proto-jet changes by $R < 0.01$. When a stable proto-jet is found, all objects in the proto-jet are removed from the list of input objects and the stable proto-jet is added to the list of jets. The whole procedure is repeated until the list contains no more objects with an E_T above the seed threshold. The cone size and the seed threshold are parameters of the algorithm.

The inclusive k_T jet algorithm is a cluster-based jet algorithm. The cluster procedure starts with a list of input objects, stable particles or calorimeter cells. For each object i and each pair (i, j) the following distances are calculated:

$$\begin{aligned} d_i &= (E_{T,i})^2 R^2 \\ d_{i,j} &= min(E^2_{T,i}, E^2_{T,j}) \Delta R^2_{i.j} \quad \text{with} \quad \Delta R^2_{i.j} = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2 \end{aligned}$$

where R^2 is a dimensionless parameter. The algorithm searches for the smallest d_i or d_{ij}. If a value of type d_{ij} is the smallest, the corresponding objects i and j are removed from the list of input objects. They are merged using one of the recombination schemes and filled as one new object into the list of input objects. If a distance of type d_i is the smallest, then the corresponding object i is removed from the list of input objects and filled into the list of final jets. The procedure is repeated until all objects are included

in jets. The algorithm successively merges objects which have a distance $R_{ij} < R$. It follows that $R_{ij} > R$ for all final jets i and j.

The midpont-cone algorithm was designed to facilitate the splitting and merging of jets. The midpoint-cone algorithm also uses an iterative procedure to find stable cones (proto-jets) starting from the cones around objects with an E_T above a seed threshold. Contrary to the iterative cone algorithm described above, no object is removed from the input list. This can result in overlapping proto-jets (a single input object may belong to several proto-jets). To ensure the collinear and infrared safety of the algorithm, a second iteration of the list of stable jets is done. For every pair of proto-jets that are closer than the cone diameter, a midpoint is calculated as the direction of the combined momentum. These midpoints are then used as additional seeds to find more proto-jets. When all proto-jets are found, the splitting and merging procedure is applied, starting with the highest E_T proto-jet. If the proto-jet does not share objects with other proto-jets, it is defined as a jet and removed from the proto-jet list. Otherwise, the transverse energy shared with the highest E_T neighbour proto-jet is compared to the total transverse energy of this neighbour proto-jet. If the fraction is greater than f (typically 50%) the proto-jets are merged, otherwise the shared objects are individually assigned to the proto-jet that is closest in η, ϕ space. The procedure is repeated, again always starting with the highest E_T proto-jet, until no proto-jets are left. The parameters of the algorithm include a seed threshold, a cone radius, a threshold f on the shared energy fraction for jet merging, and also a maximum number of proto-jets that are used to calculate midpoints.

For all algorithms the energy recombination scheme and η, ϕ metric is used. The main parameters that are varied for the different algorithms are: the cone radius for the *Iterative Cone* algorithm; the R-parameter that reflects a radius-like role for the k_T algorithm; the cone radius and the shared energy fraction threshold for merging for the *MidPoint Cone* algorithm.

For all algorithms generated and stable final state particles are used as input. Muons and neutrinos are excluded, and the effects of the magnetic field are not taken into account. All particles are assumed to emerge from the primary vertex, where the clustering is performed.

4.2.2 Event generation

For this study, performed together with the CMS top standard model group, a large event topology for physics processes at LHC has been investigated. Processes with two, four, six and eight primary quarks in the final state (i.e from dileptonic top decay in $t\bar{t}$ events to fully hadronic top decay in $t\bar{t}H$)

have been considered.

Proton collisions at 14 TeV have been generated at a luminosity of 2×10^{33} cm^{-2}s^{-1}. The $t\bar{t}$ events were generated using PYTHIA version 6.2 [48] and the $t\bar{t}H$ events were generated with compHEP version 41.10 [53], interfaced to PYTHIA version 6.215. For the leptonic decay of the W boson, only electrons and muons are considered.

4.2.3 Event selection

A realistic event selection is applied. The reconstructed jets are required to have a transverse energy larger than 20 GeV, and to be within the tracker acceptance for a proper b-tagging performance ($|\eta| < 2.4$ for the CMS experiment). Isolated signal leptons from the W-decay are removed from the jet finding input. Only if the number of jets passing these criteria is larger than or equal to the number of primary partons the event is considered for the analysis.

An iterative procedure is used to match the reconstructed jets to the generated quarks based on the ΔR distance in the (η,ϕ) plane. For each possible jet-quark couple the ΔR-value is calculated, and the smallest value is considered as a correct jet-quark matching and is removed from the list for the next iteration. When more jets have a minimal ΔR-value with the same quark, the couple with the lowest ΔR-value is taken. This procedure is iterated until all jets have their respective quark match.

4.2.4 Description of the quality markers

In order to obtain an efficient reconstruction of the kinematics of the primary partons, the selected jets should match both in energy and direction the primary partons. Variables called quality markers are defined to quantify the goodness of the event reconstruction from that perspective. Although physics effects of pile-up, gluon radiation and underlying event will degrade the overall event reconstruction efficiency, their magnitude is equal for all considered jet definitions. Hence, the relative comparison between jet definitions is meaningful.

Event selection efficiency "ϵ_s"

This efficiency is defined as the fraction of events that pass the event selection i.e. the events with a number of jets greater then the number of partons with $E_T > 20$ GeV and $|\eta| < 2.5$. When the selection is applied on quark level, the efficiency is equal to 80% for the two quarks final state, 62% for the four

quarks final state, 61% for the six quarks final state and 52% for the eight quarks final state.

Angular distance between jet and parton "Frac α_{jp}^{max}"

A jet is considered to be well reconstructed, if the ΔR distance between its direction and its best matched quark direction, α_{jp}, is sufficiently small. For each event, this results in a list of increasing α_{jp}^{i}-values, $\{\alpha_{jp}^{1}, ..., \alpha_{jp}^{n} = \alpha_{jp}^{max}\}$, where n is the amount of primary quarks in the considered event topology. Hence, α_{jp}^{max} is defined as the maximum α_{jp}^{i}-value of all i jet-quark pairs in the event. The α_{jp}^{i} distributions for a four quarks final state are shown in Fig. 4.2. The last one of these plots represents the α_{jp}^{max} variable. To quantify

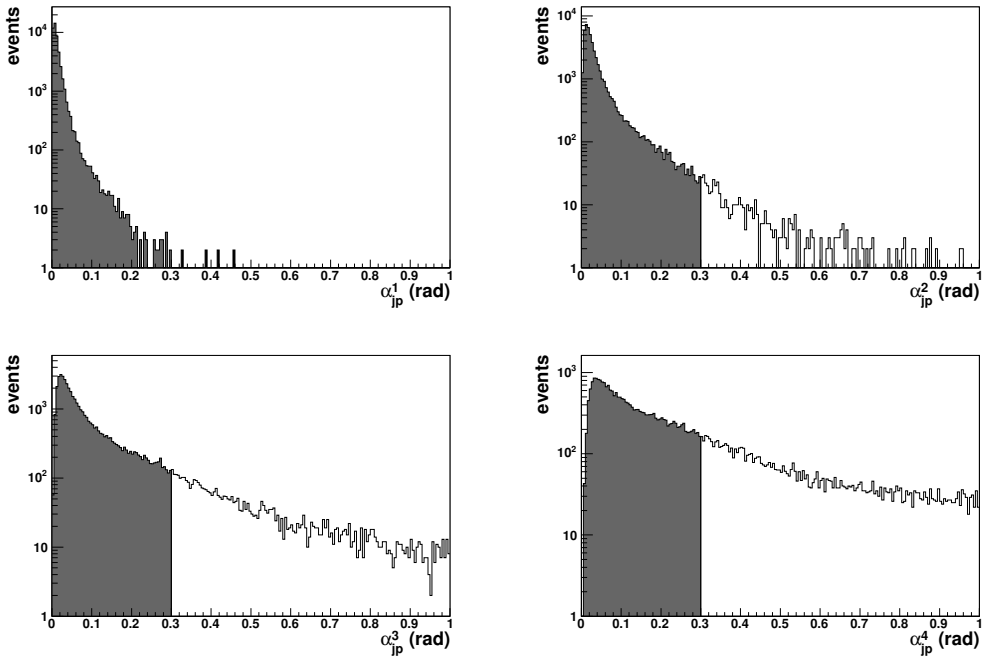


Figure 4.2: Distributions of α_{jp}^{i} in increasing order for the IC algorithm with a cone radius of 0.4 in the case of a final state with four quarks. The 0.3 rad criteria as discussed in the text is indicated.

the angular reconstruction performance of a particular jet definition, a quality marker is defined as the fraction of events with a α_{jp}^{max} value lower than 0.3 and denoted as "Frac α_{jp}^{max}". The choice of the worst jet is motivated by the reasoning that the direction of all primary quarks in the event are required to be well determined.

Energy difference "Frac β_{jp}^{max}"

The reconstructed energy of the primary quarks is usually biased and has a broad resolution. Figure 4.3 left shows the average fraction of the quark energy that is reconstructed for a specific algorithm configuration as a function of the reconstructed transverse jet energy. Such a calibration curve can be interpreted as an estimator for the expected reconstructed energy. For this plot only well matched (α_{jp} <0.3), non-overlapping jets were taken into account. For the iterative cone algorithm, a jet is considered to be non-overlapping, if its ΔR distance to any other jet is larger than twice the value of the cone radius parameter of the algorithm. It is the aim of jet calibration studies to determine these average corrections to be applied on the reconstructed jet energies. Therefore the remaining component is the energy resolution.

The β_{jp}^{i} values are defined for each primary quark i as the distance from the expected energy fraction (deduced from the fitted function in Fig. 4.3 left) in units of standard deviations. For each selected event, the primary quark with the highest β_{jp}^{i} value, called β_{jp}^{max} is considered to be the one with the worst reconstruction performance from the energy point of view. An example for the β_{jp}^{max} distribution is shown in Fig. 4.3 (on the right). An energy related quality marker is defined as the fraction of events with a β_{jp}^{max} lower than 2 standard deviations, and denoted as "Frac β_{jp}^{max}".

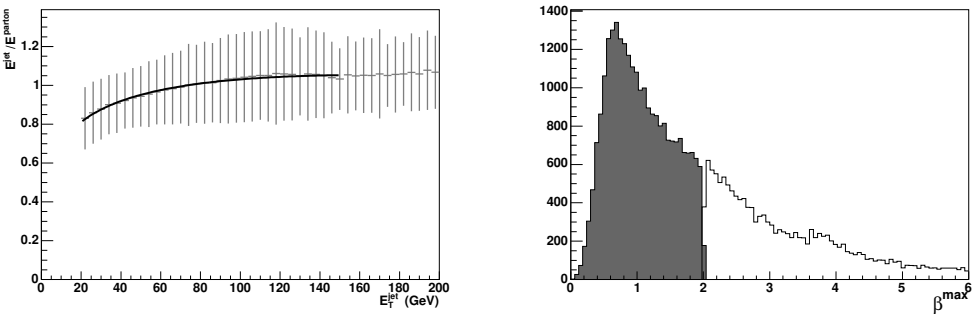


Figure 4.3: Left: example of a $\frac{E^{jet}}{E^{parton}}$ vs. E_T^{jet} curve for the IC algorithm with a cone radius of 0.4, applied on a final state with four primary quarks. The vertical bars illustrate the resolution. Right: distribution of β_{jp}^{max} for the IC algorithm with a cone radius of 0.4, applied on a final state with four primary quarks.

Combined variable "Frac(α_{jp}^{max}+β_{jp}^{max})"

A combined variable is defined as the fraction of events in which both the direction and the energy of the n primary quarks are well reconstructed following the definitions described above. The correlation between α_{jp}^{max} and β_{jp}^{max} is shown in Fig. 4.4 left, where both quality criteria define a rectangular area in which the kinematics of the primary quarks are sufficiently well reconstructed from the analysis performance point of view. As an illustration of the separation power of this combined variable, the reconstructed spectrum of the hadronic top quark mass in the semileptonic $t\bar{t}$ final state is shown in Fig. 4.4 right. The black histogram refers to the events in which the jets are reconstructed with $\alpha_{jp}^{max} < 0.3$ and $\beta_{jp}^{max} < 2$ (events inside the box of Fig. 4.4 left). The grey histogram refers to the events in which the kinematics of the primary quarks are badly reconstructed based on the combined variable (events outside the box of Fig. 4.4 left).

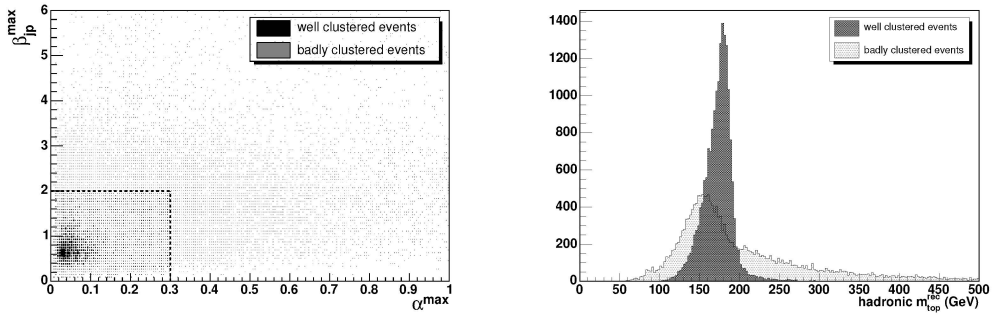


Figure 4.4: Left: box plot of β_{jp}^{max} vs. α_{jp}^{max} for the IC algorithm with a cone radius of 0.4, applied on a final state with four primary quarks. Right: distribution of the hadronic top quark mass, using jets clustered with the IC algorithm with a cone radius of 0.4, applied on a final state with four primary quarks.

Overall quality marker "FracGood"

The fraction of selected and well reconstructed events, i.e. the selection efficiency ϵ_s, multiplied by the combined variable Frac(α_{jp}^{max}+β_{jp}^{max}) is defined as "FracGood".

This last quality marker is interpreted as an estimate for the reconstruction efficiency of the kinematics of the primary quarks of the complete event, and therefore used to compare different algorithms and setups. Although this variable gives a powerful overall indication of a reasonable jet definition,

it is sometimes useful to consider the partial information of the individual quality markers. Depending on the priorities of a specific physics analysis, one would be interested in the average number of reconstructed jets, or the energy resolution for non-overlapping jets, or the efficiency of the angular matching between primary quark and jet. The average number of jets gives an idea of the sensitivity to pile-up, underlying event, and the rate of fake jets, while the energy resolution can be linked to the issue of jet calibration.

4.2.5 Results

In this subsection the most important observations for each jet clustering algorithm considered are summarized.

Iterative cone algorithm

Fig. 4.5 right shows the "FracGood" variable as a function of the cone radius. The algorithm configuration with the maximum fraction of well reconstructed events, and as a consequence the optimal configuration from analysis point of view, is obtained for a cone radius varying from 0.3 to 0.5, depending on the event topology. The dependence of the fraction of well reconstructed events on the minimal transverse energy of the jet seed is found to be negligible. It has to be remarked that a stronger dependence as well as a larger optimal cone radius is however expected when the jet input is changed from simulated to reconstructed particles.

Another important observation is the decrease of the optimal cone radius for increasing jet multiplicity. This behaviour can be explained by the higher probability of overlapping jets for higher jet multiplicities. The generally lower selection efficiency (Fig. 4.5 right) for high multiplicities is due to the higher probability for hard radiation, a higher probability for overlapping jets and the p_t spectrum of the jets. Furthermore, a lower selection efficiency ϵ_s is observed for very low jet radii. This can be explained by the transverse energy cut of 20 GeV which is more severe for small opening angles.

Both effects will result in a more difficult jet clustering task for high jet multiplicities. Compared to 55% of well clustered events in the two quark final state, only 6% of the events in an eight quark topology pass all the criteria. For these numbers the IC algorithm with the optimal cone radius was applied.

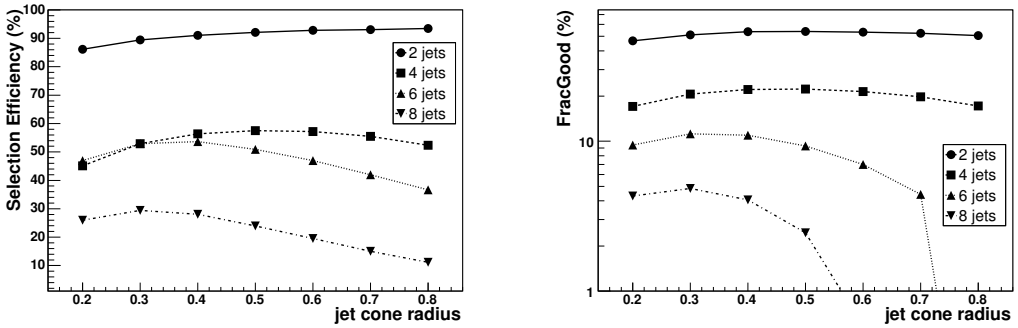


Figure 4.5: Left: fraction of selected events versus the cone radius (IC algorithm). Right: fraction of well clustered and selected events versus the cone radius (IC algorithm).

Inclusive k_T Algorithm

Fig. 4.6 shows the result for the scan of the R-parameter of the inclusive k_T algorithm. Again, a strong dependence on the jet multiplicity is observed. For the two quark topology, the best performance is reached at 0.6, while this value is reduced to 0.3 for the 8 quark topology. This behaviour is expected keeping in mind that the R-parameter plays a comparable role for the inclusive k_T algorithm as the jet radius does for the *Iterative Cone* algorithm. Compared to the optimal configuration of the *Iterative Cone* algorithm, this algorithm performs almost identical for the two quark case, but is able to get higher reconstruction efficiencies for larger jet multiplicities.

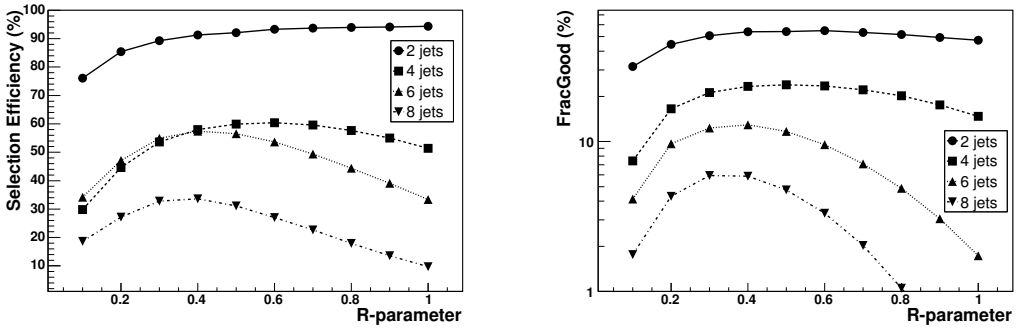


Figure 4.6: Right: fraction of selected events versus the R-parameter (k_T algorithm).Left: fraction of well clustered and selected events versus the R-parameter (k_T algorithm).

Midpoint cone algorithm

The scan of the cone radius is shown in Fig. 4.7 left and the dependence on the shared energy fraction threshold for merging is shown in Fig. 4.7 right. For high jet multiplicities, the *MidPoint Cone* algorithm is able to reach

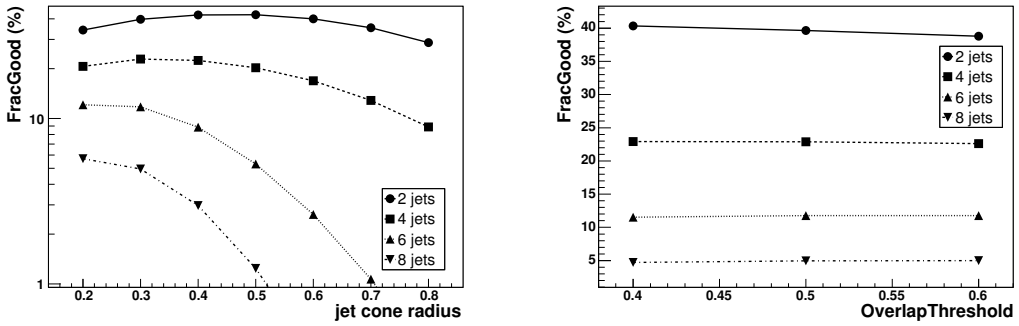


Figure 4.7: Left: fraction of well clustered and selected events versus the cone radius for a merging threshold of 0.5 and a cone area fraction of 0.25 (MC algorithm). Right: fraction of well clustered and selected events versus the threshold for merging for a cone radius of 0.3 and an area fraction of 0.25 (MC algorithm).

slightly higher efficiencies than the *Iterative Cone* algorithm. Surprisingly, almost no dependence on the shared energy fraction threshold for merging has been found. This behaviour might be related to the fact that simulated particles have been used as input or to the implemenation in the CMS framework which was not yet mature enough at the time of this study. A new investigation of the performance of this algorithm should be performed as soon as it becomes available in the new CMS software framework, because the experience of other experiments with this algorihm are very promising.

Summary of the main observations

Table 4.1 summarizes the optimal parameter values for the three jet clustering algorithms, and for each of the considered event topologies. For each optimal jet configuration, the respective estimate of the fraction of well reconstructed events is given.

Correlation between optimized configurations

The correlation between the optimized *Iterative Cone* algorithm and the inclusive k_T algorithm for the final state with four primary quarks is shown in

	IC jet radius		k_T R-parameter		MC jet radius		MC Overlap Threshold	
	Value	*FracGood*	*Value*	*FracGood*	*Value*	*FracGood*	*Value*	*FracGood*
2 quarks	0.5	53.9	0.6	54.9	0.5	42.4	0.40	40.3
4 quarks	0.5	22.3	0.5	23.8	0.3	22.8	0.40-0.50	22.9
6 quarks	0.3	11.2	0.4	12.9	0.2	12.1	0.50-0.60	11.8
8 quarks	0.3	4.85	0.3	5.93	0.2	5.72	0.60	5.0

Table 4.1: Overview of the optimal parameter values with their respective estimate of the fraction of well reconstructed events.

Fig. 4.8 for the α_{jp}^{max} and β_{jp}^{max} variables. Fig. 4.8 together with the results summarised in Tab. 4.1 emphasize that both algorithms give similar results and that both are suitable to reconstruct events with high jet multiplicities.

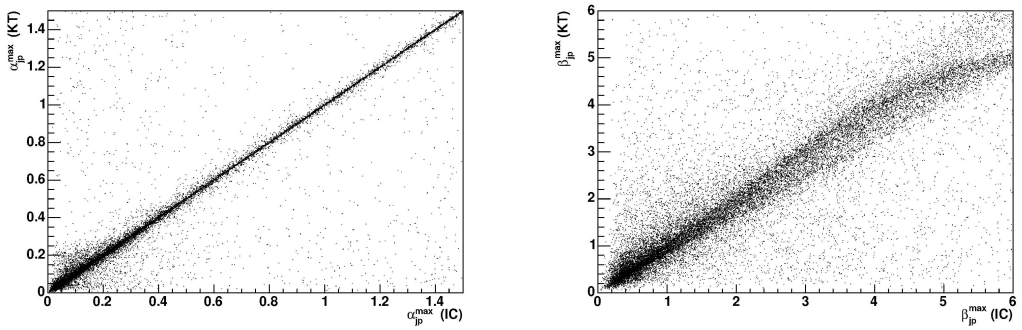


Figure 4.8: Left: correlation between the IC and k_T algorithms for the α_{jp}^{max} variable in the case of the final state with four primary quarks. Right: correlation between the IC and k_T algorithms for the β_{jp}^{max} variable in the case of the final state with four primary quarks.

Robustness of the method against hard radiation

The sensitivity of the overall observations to the radiation of gluons with a large transverse momentum relative to their mother quark, or from the initial state proton system, is investigated in the following. The distributions of the α_{jp}^{i}-values ordered by their magnitude within an event are shown in Fig. 4.9 for a sample without initial and final state radiation[3]. This has to be compared directly to Fig. 4.2 which shows the same plots including final state radiation. Obviously, the long tails are not present in the case without radiation which indicates that the ΔR cut of 0.3 for the worst jet is not

[3]PYTHIA parameters $MSTP$ 61 and 71 were switched off.

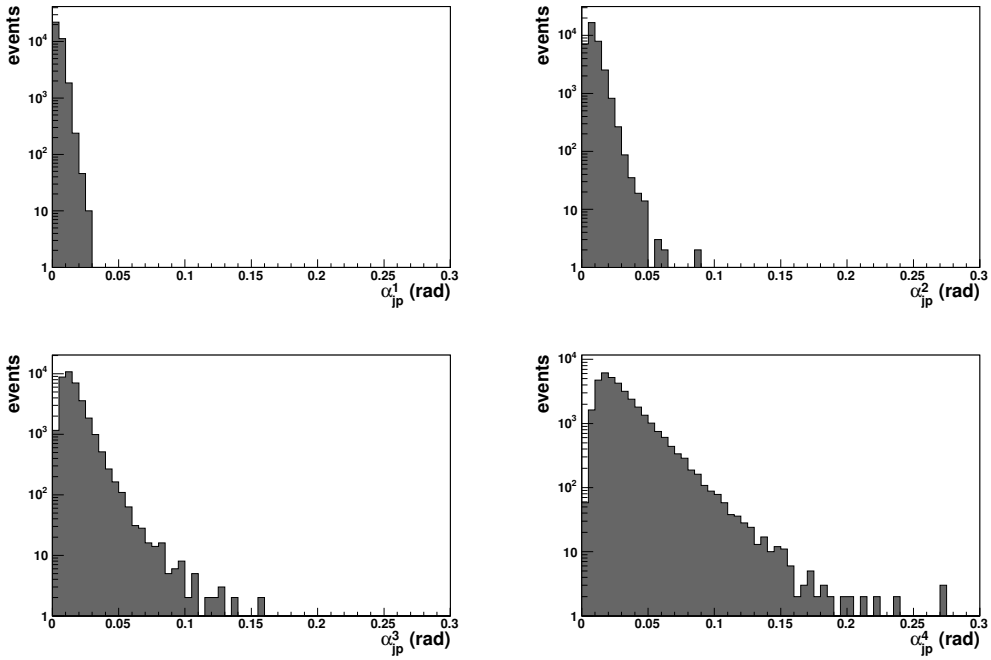


Figure 4.9: Distributions of α_{jp}^i in increasing order of magnitude for the IC algorithm in the case of a final state with four primary quarks which do not radiate hard gluons.

expected to have an effect in this case. The observation is indeed, that the Frac(α_{jp}^{max}+β_{jp}^{max}) quality marker has a flat distribution, but not the selection efficiency and therefore the "FracGood" quality marker.

Fig. 4.10 (left) shows the fraction of selected, well clustered semileptonic $t\bar{t}$ events with and without initial and final state radiation for the *Iterative Cone* algorithm. The addition of radiation results in an overall lower efficiency, but the optimal cone radius and the shape of the curve are robust. A similar observation was obtained for the inclusive k_T algorithm in Fig. 4.10 (right).

In order to quantify the effect of radiation on the resolutions, Fig. 4.11 shows the two cases for the *Iterative Cone* and the inclusive k_T algorithm for the case with four partons in the final state. The energy resolution is defined as the RMS divided by the mean value of the E^{jet}/E^{quark} distribution, and the angular resolution is defined by the width of a gaussian fit to the symmetrized ΔR distribution. As expected, the overall resolutions are better in the case without radiation, but the shape of the curves remains invariant.

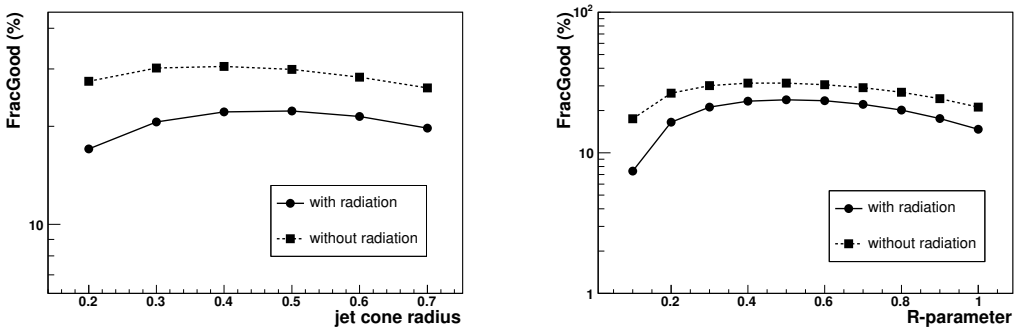


Figure 4.10: Left: influence of hard gluon radiation on the fraction of selected, well clustered events, as a function of the IC cone radius in the case with four primary quarks in the final state. Right: influence of hard gluon radiation on the fraction of selected, well clustered events, as a function of the k_T R-parameter in the case with four primary quarks in the final state.

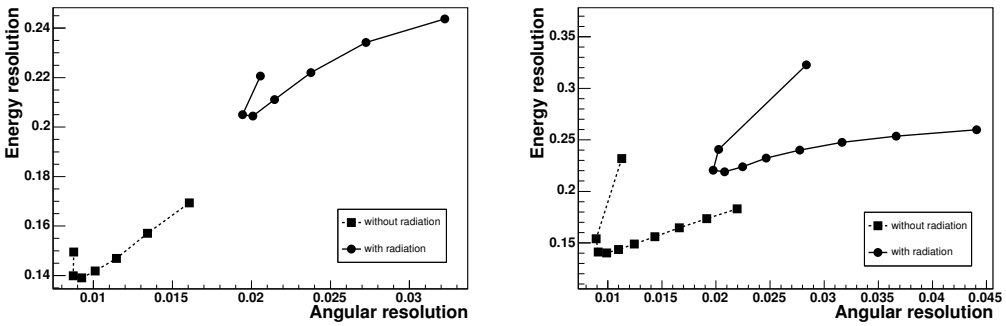


Figure 4.11: Energy resolution versus angular resolution (ΔR distance between jet and quark) for the IC algorithm (left) and k_T algorithm (right) in the case of four jets in the final state.

4.3 Monte Carlo jet calibrations

The calibrations used in this work are done in two parts. First, reconstructed jets in the calorimeter are corrected to the particle jet, i.e. the jet built with simulated stable particles after hadronization. The comparison is made by applying the same jet algorithm to stable particles (excluding neutrinos and muons) and calorimeter cells, respectively. In this step of calibration the effects due to the detector design is taking into account and the energy is corrected with respect to the simulated particle reaching the calorimeter. In CMS this step of the calibration is named *particle-level* correction. In the second calibration step the energy of the particle jet is corrected to the primary parton energy. This is needed, especially, since some particles of the parton shower can lie outside the jet cone. This effect is increased when small cone size is used. This second step of calibration is named in CMS *parton-level* correction. In Fig. 4.12 the two step of calibration are shown.

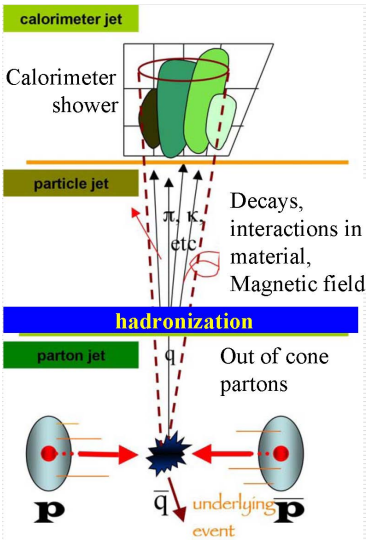


Figure 4.12: The two main step calibration for hadronic jets. The first set of calibration is done from the calorimeter cells to the particle jet, the second step correts the particle jet to the primary parton energy.

A matching criterion, based on the distance $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ is used to associate the calorimeter jets to the particle jets. In the official CMS calibrations [54] the reconstructed jet is considered associated to a particle jet if the distance $\Delta R < 0.2$. The CMS calibrations are available for the *Iterative Cone* algorithm with a jet cone radius of $\Delta R = 0.5$ or 0.7, and for the k_T algorithm with $R = 1$. Unfortunately the *Iterative Cone* algorithm with a

cone radius of $\Delta R = 0.7$ and the k_T algorithm with $R = 1$ are not suitable to reconstruct events with six partons in the final state ($t\bar{t}H$ with Ws decay semileptonically) as shown in the previous section. The ratio E_T^{Rec}/E_T^{MC} for different value of E_T^{MC} is shown in Fig. 4.13 as a function of pseudorapidity (η) before (left) and after correction (right) for jets reconstructed by the *Iterative Cone* algorithm with a cone radius of $\Delta R = 0.5$. Fig. 4.13 underlines that for particle jet with E_T below 50 GeV a calibration of about 50% has to be applyied to correct the calorimeter jet E_T, pointing out the large systematic error that can arise from calibrations. Moreover in Fig 4.14 is shown the parton level correction for jets reconstructed both with a jet cone of 0.5 and 0.7. It has to be noted that for jets with a large jet cone is not necessary to apply parton-level correction (almost all the parton shower is contained in the particle jet cone) while this correction are not negligible for jet with a cone radius of 0.5. Furthermore the correction are different for jets originated from quarks and from gluons. Because the determination of the relative ratio of quarks and gluons in the data is expected to be difficult, this would lead to additional systematic errors on the jet energy scale [38]. It is clear that the reconstruction and calibration of soft jets at LHC will be a very critical task and it will be the major source of systematic error for events with jets in the final state. As we will see in the chapter 6 this is especially critical for the $t\bar{t}H$ channel because soft jets reconstruction is needed to recover signal efficiency.

For the $t\bar{t}H$ decay fully hadronically a set of calibration has been done for smaller values of jet radius, $\Delta R = 0.3, 0.35...0.5$ (*Iterative Cone* algorithm) and for a smaller value of the R parameter, $R = 0.4$ (k_T algorithm) [55]. In this case the Monte Carlo calibration taking into account also the different flavor of jet composition. At the particle-level correction a distinction has been introduced to evaluate different detector effect on b-jet and not-b-jet. The major motivation is due to the higher probability to have μ-leptons and neutrinos from semi-leptonic decays of b hadrons in b-jets with respect to the light-quark and gluon jets. These calibrations have been done using directly a sample of $t\bar{t}H$ events. This allows to have both a rich sample of b-jets and a specific parton-level correction for the $t\bar{t}H$ channel. The different correction needed at particle-level between b-jets and not-b-jets is shown in Fig. 4.15. Also in this case a major difference for b-jets and not-b-jets exists especially at low transverse energy, pointing out the difficulty to calibrate soft jets correctly.

The different behaviour of the *ad-hoc* $t\bar{t}H$ and the official CMS calibration performed by the JetMET group is displayed in Fig 4.16 as a function of the JetMET calibrated jet E_T and for three different region of η. For this comparison a sample of QCD di-jet events with $\hat{p_T}$ up to 600 GeV has been

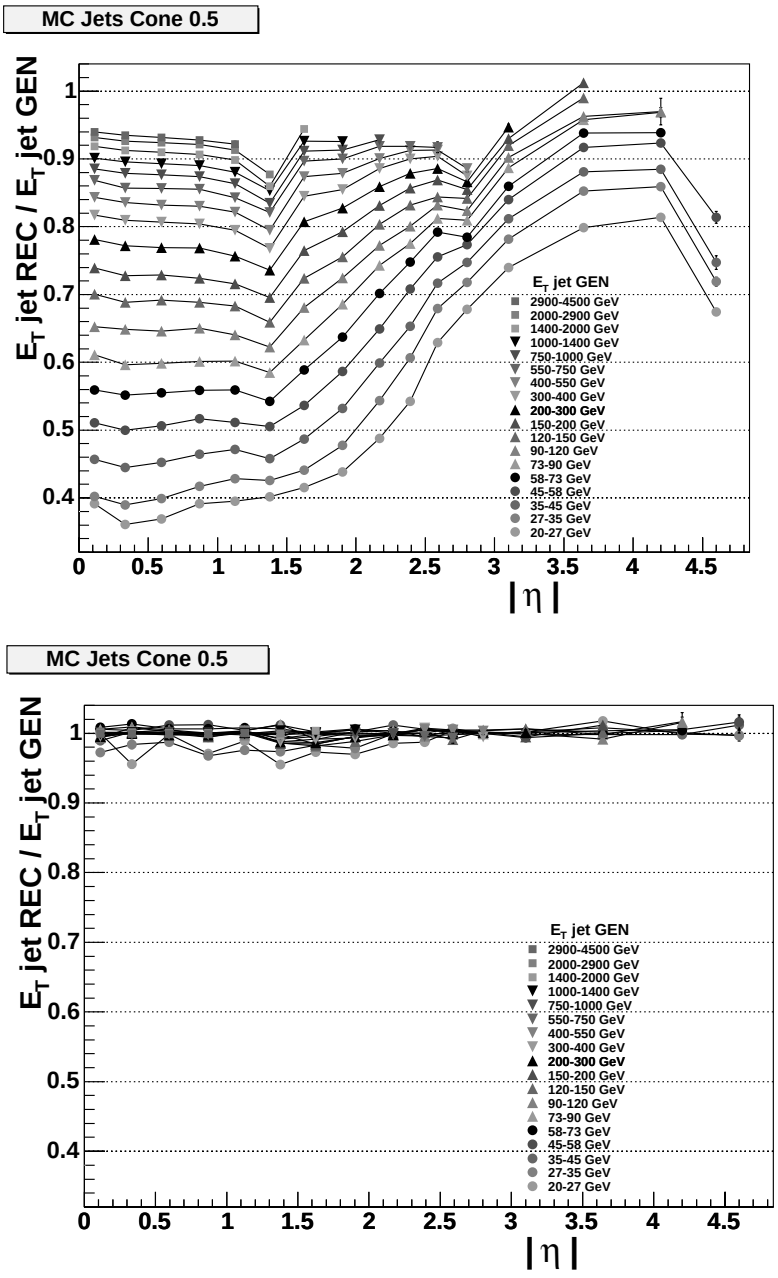


Figure 4.13: On top: the ratio of the reconstructed jet transverse energy E_T^{Rec} to the generated transverse energy E_T^{MC} as a function of pseudorapidity of generated jet $|\eta|$ of jets with different E_T^{MC} reconstructed by the *Iterative Cone* algorithm with a cone radius of $\Delta R = 0.5$. Below: the ratio E_T^{Rec}/E_T^{MC} after correction.

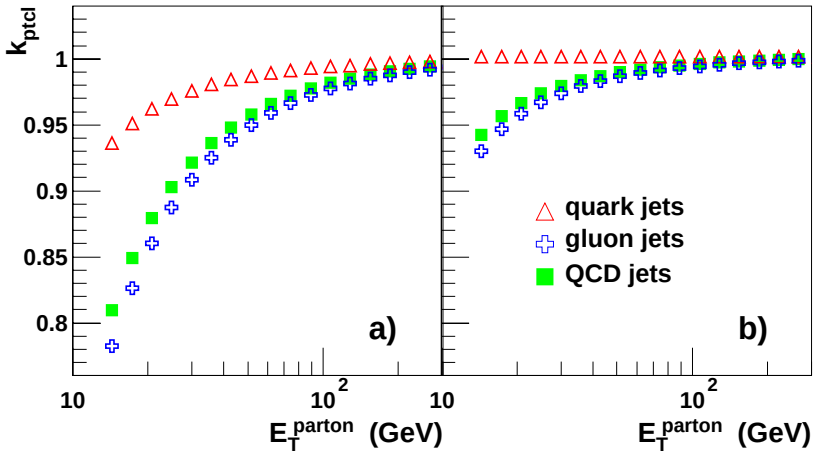


Figure 4.14: Ratio of transverse momenta of particle jet to the transverse momenta of the initial parton for quark, gluon or QCD jets as a function of the transverse parton energy. In *a* the ratio is shown for a cone radius of 0.5 and in *b* for a cone radius of 0.7. The correction are more important for a smaller jet cone size and for gluon-jets respect to quark-jets

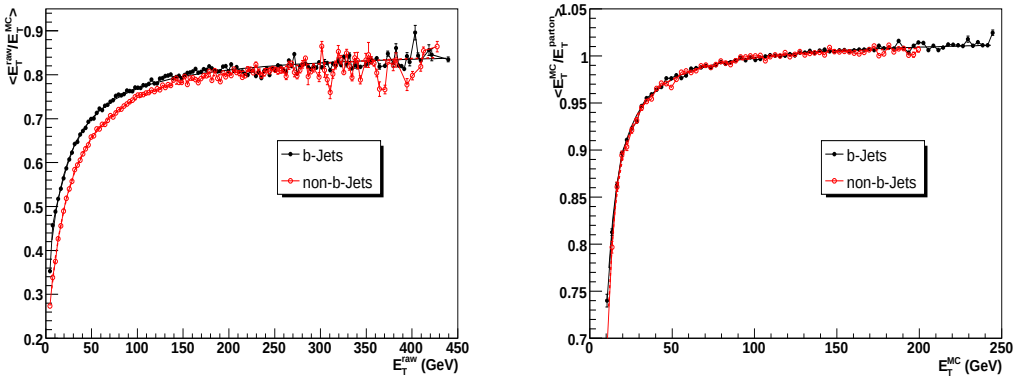


Figure 4.15: Left: E_T^{reco}/E_T^{MC} ratio for *b*-jets (red) and not-*b*-jets (black) as a function of E_T^{reco}. Right: E_T^{MC}/E_T^{Parton} ratio for *b*-jets (red) and not-*b*-jets (black) as a function of E_T^{MC}. In the parton-level correction the difference between *b*-jets and not-*b*-jets is negligible.

used and the jets are reconstructed using the *iterative cone* algorithm with a cone radius of $\Delta R = 0.5$. In the tracker acceptance region the difference between the two set of calibrations is within 5% for non-*b*-jets with $E_T > 40$ GeV while for softer jets the $t\bar{t}H$ calibration gives always harder jets. The $t\bar{t}H$ calibration for *b*-jets produces as expected higher transverse energy.

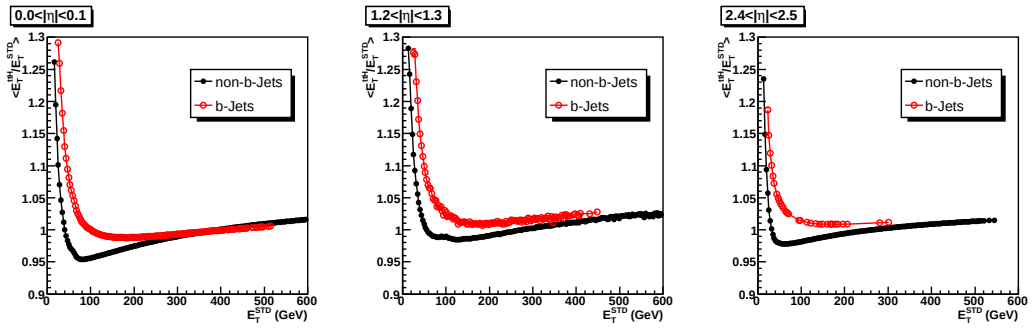


Figure 4.16: Comparing of the official CMS jet calibration with the *ad-hoc* $t\bar{t}H$ jet calibration for three different η bins.

4.4 Missing tranverse energy

The measure of missing transverse energy (E_T^{miss}) at LHC will be very important both for top physics when a W decay leptonically and for "new physics" discovery if massive vector bosons exist. In CMS the measurement of E_T^{miss} will be helped by excellent cell segmentation, hermeticity and forward coverage performed by the very forward calorimeters [56]. On the other hand the presence of pile-up at LHC, and the bending of tracks by the 4 T magnetic field in CMS will degrade the measurement of E_T^{miss}. In this work the missing transverse energy of the event is computed as [57]:

$$E_t^{miss} = \sum_i E_t^{tower} - (\sum_j E_t^{RawJet} - \sum_k E_t^{CaliJet}) + \sum_m E_t^{Muon} \qquad (4.1)$$

where the sum with index i runs over the calorimetric towers, that with j runs over raw jets, k runs over calibrated jets, and m runs over the reconstructed muons of the event. Equation 4.1 thus takes into account the corrections due to jet calibration and the contributions of muons that are not measured in the calorimeter system. The distribution of the reconstructed missing transverse energy of semi-leptonic $t\bar{t}H$ events and the missing transverse energy resolution are shown in Fig. 4.17 and Fig. 4.18. When jet correction is taken into account the E_T^{miss} resolution has a mean value centred to zero and a lower RMS [31].

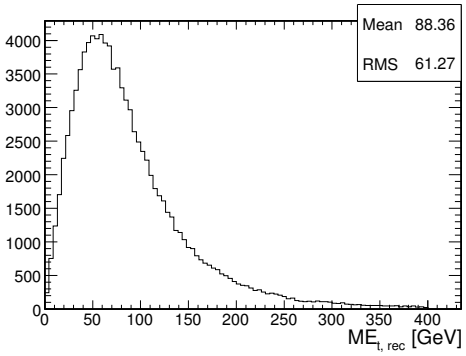


Figure 4.17: Distribution of the reconstructed missing transverse energy.

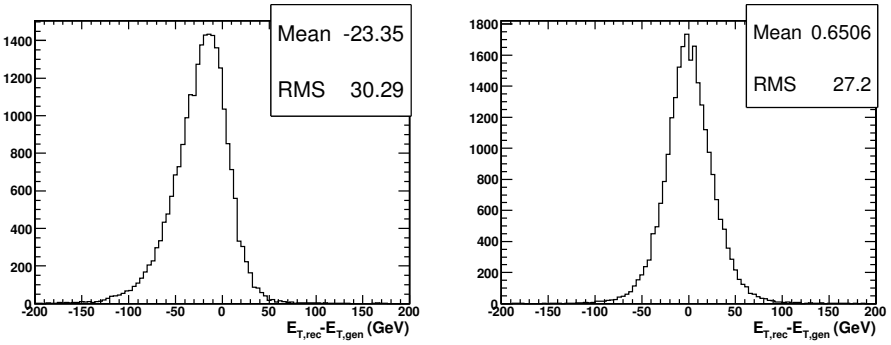


Figure 4.18: Resolution of the reconstructed missing transverse energy in $t\bar{t}H$ events with semi-leptonic W decays. Right: E_T^{miss} corrected with jet calibration as in 4.1

Chapter 5

b-tagging

The identification of b-jets in $t\bar{t}H$ events is very important for two reasons: first it is fundamental to distinguish the b-jets coming from the top and Higgs decays from the light jets coming from W decay or extra jets from initial and final state radiation. Secondly b-tagging is practically the only tool that will be able to reduce backgrounds that have more than 2 jets associated to a pair of quark tops. In this study we use a combined secondary vertex b-tagging; several topological and kinematical secondary vertex related variables as well as information from track impact parameters are combined into a single tagging variable to discriminate between jets originating from b-quark and those from other sources. A summary of the parameters used and of the method is given in the following, a detailed description is published in Ref. [58].

5.1 Combined secondary vertex tag

Identifying b-jets relies on the properties of the production and the weak decay of b-hadrons. The relatively large lifetime of b-hadrons of about 1.5 ps ($c\tau \approx 450$ μm) corresponding to a flight distance that is observable using high resolution tracking systems. This leads to secondary vertices displaced from the primary event vertex and charged particles tracks incompatible with the primary vertex. For this reason efficient track reconstruction, and in particular precise spatial reconstruction close to the interaction point are key ingredients.

5.1.1 Algorithm description

Tracks are reconstructed using a Kalman filter [46] technique and the following track selection cuts are applied to improve the quality of the selected tracks:

- At least 8 reconstructed hits in total (pixel and silicon strip detectors).
- At least 2 reconstructed hits in the pixel detectors.
- Transverse momentum $p_T > 1$ GeV/c.
- χ^2/ndf of the track fit < 10.
- Transverse impact parameter with respect to the reconstructed primary vertex < 2 mm to reject charged particle tracks having their origin from sources showing much larger displacement from the primary vertex (e.g. V^0 decays, photon conversions and nuclear interactions in the beam pipe or the first layers of the pixel detector).

Tracks are associated to jets if the distance in the η, ϕ space ($\Delta R_{jet-track} = \sqrt{\Delta\eta^2_{jet-track} + \Delta\phi^2_{jet-track}}$) between the jet direction and the track is less than 0.3. For the following b-tagging studies jets are reconstructed using the *Iterative Cone* algorithm with a cone radius of $\Delta R = 0.5$ and the jets are calibrate using the official CMS calibration as described in 4.3. The displacement of the track respect to the primary vertex is computed using the impact parameter divide by its error and the sign of the track impact parameter is defined positive if the track is reconstructed downstream of the primary vertex with respect to the jet direction, negative otherwise.

Secondary vertex are reconstructed using the Trimmed Kalman Vertex Finder (TKVF) described in Ref. [59]. During a first iteration the fitter is applied to the complete input set of tracks, yielding as outputs a vertex candidate and a set of tracks that are incompatible with that vertex candidate. During the subsequent iterations, the same procedure is applied to the set of incompatible tracks identified in the previous iterations. In this way the TKVF is sensitive to primary and secondary vertices, so a vertex filter is used to select secondary vertex candidates. The following cuts are applied to select secondary vertex candidates coming from b-hadrons decays:

- The distance from the primary vertex to the secondary vertex in the transverse plane has to exceed 100 μm and must not exceed 2.5 cm.
- The distance from the primary vertex to the secondary vertex in the transverse plane divided by its error has to be greater than 3: $L_t/\sigma_{L_t} > 3$.

- The invariant mass of charged particles associated to the vertex must not exceed 6.5 GeV/c^2.

- The vertex must not be compatible with a V^0 decay.

Based on the results of the secondary vertex reconstruction, the algorithm of the combined secondary vertex b-tagging split the jets candidate in 3 vertex categories:

"RecoVertex" : At least one secondary vertex candidate has been reconstructed and passed the selection criteria. All tracks from all accepted vertices are used for the computation of the vertex related variables if there is more than one accepted secondary vertex.

"PseudoVertex" : If no reconstructed secondary vertex candidate has been found, a so called PseudoVertex is created from charged particle tracks not being compatible with the primary event vertex, having a signed transverse impact parameter significance greater than 2, if at least 2 such tracks are present in the jet.

"NoVertex" : If neither "RecoVertex" or "PseudoVertex" are fulfilled.

The distribution of the categories for the different jet flavours are shown in Fig. 5.1. It can be seen that the presence of a secondary vertex alone is already discriminating between b-quark jets and other jets.

5.1.2 Input variables and combined discriminating variable

To get optimal performance, several topological and kinematic variables are combined. The selection of variables entering into the combination depends on the vertex category. For the category "RecoVertex" the following variables are defined:

1. The invariant mass of charged particles associated to the secondary vertex. For secondary vertices in b-jets, the vertex mass can be significantly above the mass of charm hadrons, allowing to suppress efficiently this background.

2. The multiplicity of charged particles associated to the secondary vertex; b-hadron decays show a significantly larger track multiplicity than e.g. charm hadron decays.

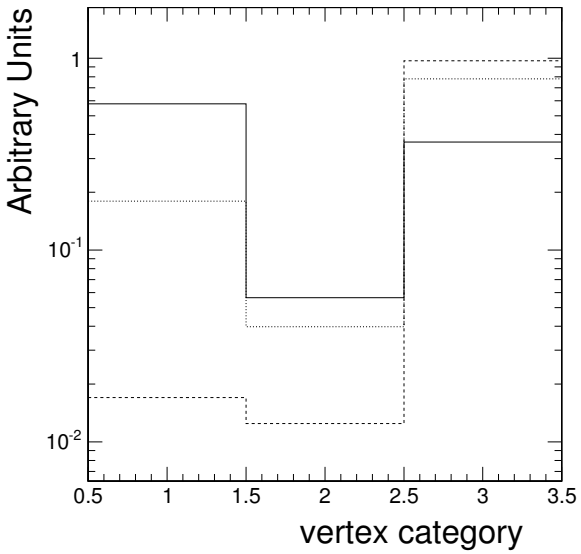


Figure 5.1: Distribution of the vertex category, for *b*-jets (solid), *c*-jets (dotted) and *uds*-jets (dashed) with transverse momenta between 50 GeV/c and 80 GeV/c.

3. The distance between the primary vertex and the secondary vertex in the transverse plane divided by its error. This is sensitive to the large flight path of *b*-hadrons.

4. The energy of charged particles associated to the secondary vertex divided by the energy of all charged particles associated to the jet. This quantity is sensitive to the hard fragmentation function of b and c-quarks.

5. The rapidities of charged particle tracks associated to the secondary vertex with respect to the jet direction: $y = \frac{1}{2} \cdot \frac{E+p_\parallel}{E-p_\parallel}$. This variable enters for n tracks, where n is the secondary vertex multiplicity.

6. The track impact parameter significances in the transverse plane.

7. The track impact parameter significance of the first track exceeding the charm threshold as described in the following. Tracks are sorted in decreasing order of their impact parameter significance. The invariant mass is computed for tracks 1 to n. If this mass exceeds a value related to the mass of charm hadrons, the impact parameter significance of the track moving the n-track mass above this threshold can be added to the discriminator. The mass threshold is set to 1.5 GeV. This is

lower than typical charm hadron masses because neutral particles are missed and not all charged particles from the decay are reconstructed and accepted.

For vertex category 2 ("PseudoVertex"), most of these variables still offer some discriminating power between b and *non-b* jets. Since there is no attempt to fit the geometrical position of the "PseudoVertex", the significance of the flight distance (the third variable defined in the list above) is not used for this category, whereas the other six variables enter into the list of input variables. For vertex category 3 ("NoVertex"), no additional variables are defined. The distribution of some of this variables for the category "RecoVertex" are shown in Fig 5.2, for the other categories see reference [58].

The variables described above are combined into a single discriminating variable using a Likelihood ratio technique. Since the distributions of most of the variables look significantly different for c-jets and $udsg$-jets, the Likelihood ratio contains two parts for the discrimination against these two backgrounds. The Likelihood function is defined as:

$$\mathcal{L}^{b,c,q} = f^{b,c,q} \times \sum_i f_\alpha^{b,c,q}(x_i) \tag{5.1}$$

and the discriminator variable, d, is defined as:

$$d = f_{BG}(c) \times \frac{\mathcal{L}^b}{\mathcal{L}^b + \mathcal{L}^c} + f_{BG}(q) \times \frac{\mathcal{L}^b}{\mathcal{L}^b + \mathcal{L}^q} \tag{5.2}$$

where α denotes the vertex category as defined above, x_i are the individual variables, q stands for u, d, s-quarks and gluon jets, $f_{BG}(c)$ and $f_{BG}(q)$ are the expected a-priori probabilities for the c and q content in non-b-jets ($f_{BG}(c) + f_{BG}(q) = 1$), $f^{b,c,q}$ is the probability for flavour b, c, q to fall into category α and $f_\alpha^{b,c,q}(x_i)$ is the probability density function for variable (x_i) for category α and flavour b, c, q. In this case the $f_{BG}(c) = 0.25$ and then $f_{BG}(q) = 1 - f_{BG}(c) = 0.75$. The choice was made because in hadronic decays of W-bosons the expected fraction of c-jets is about 0.25. This value also represents a good compromise for other compositions of the *non-b*-jet sample. The probability density function are extracted from a statistically independent sample of simulated QCD events and on a sample of semileptonic $t\bar{t}$ events.

The quantity d is used as final discriminator and, for a sample of QCD events for jets with underlying moment between 50 and 80 Gev, the d distribution for all categories together is shown if Fig. 5.3.

The misidentification rate of charm and light flavour jets as a function of the b-tagging efficiency is shown in Fig. 5.4 for the $t\bar{t}H$ and $t\bar{t}2j$ samples, respectively [31]. The performance of the algorithm is given using two different

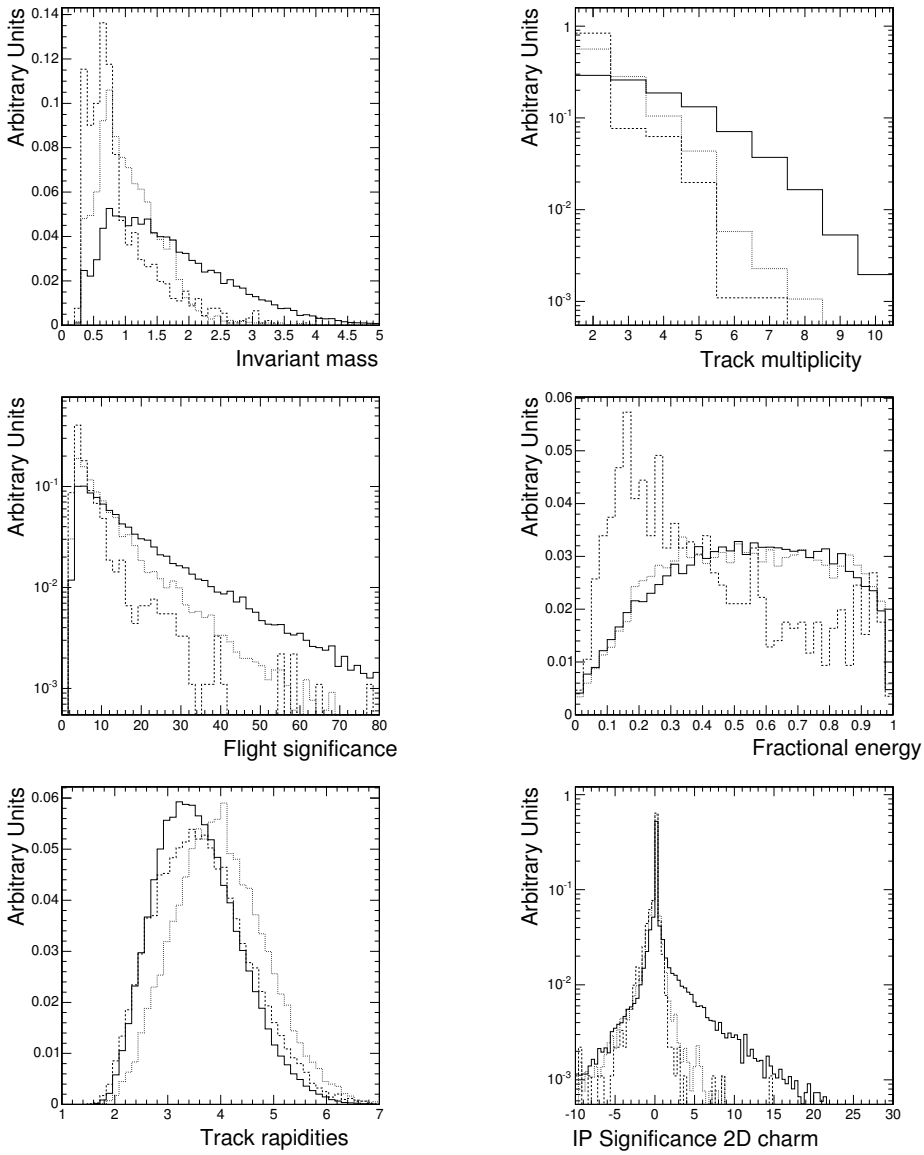


Figure 5.2: The distributions for the input variables for the vertex category "RecoVertex" for b-jets (solid), c-jets (dotted) and uds-jets (dashed).

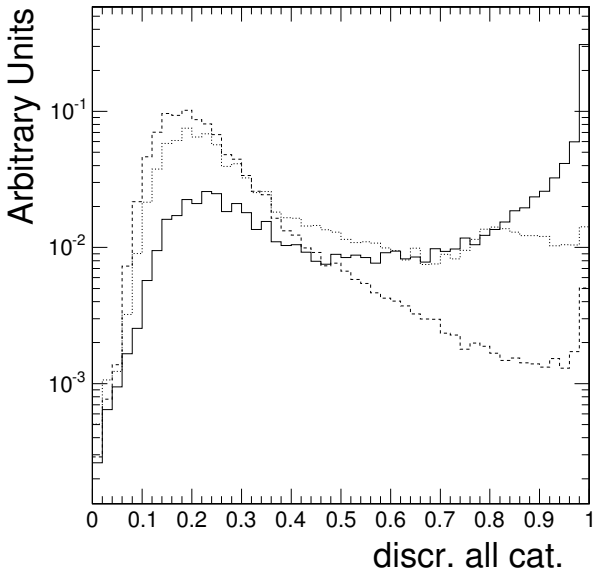


Figure 5.3: The distributions of the discriminator d variable for b-jets (solid), c-jets (dotted) and uds-jets (dashed). The distribution is shown for a sample of QCD events for jets with underlying moment between 50 and 80 Gev .

definitions of the true jet flavour. The first one is called "physics" definition and the second one "algorithmic" definition. In the "physics" definition the reconstructed jets is matched to the initial parton from the primary process (e.q. for a sample of $t\bar{t}2j$ for all primary partons from top quarks, W and the two *extra* partons). In this case if a jet comes from a gluon that split in a couple of $b\bar{b}$, or $c\bar{c}$, it is still considered a *gluon*-jets. In the "algorithmic" definition instead if a jet originating from a gluon that has split into a $b\bar{b}$ pair would get the flavour of the b-quark assigned, because from the algorithm point of view the jet looks in many aspects as a genuine b-jet. As shown in Fig. 5.4 top left and right, the main difference in these definitions is for gluon jets, where the non-negligible splitting rates into $b\bar{b}$ and $c\bar{c}$ pairs impose a serious limitation for b-tagging. In a $t\bar{t}2j$ sample about 65% of the two extra jets are originated by gluons and the splitting rate for these events is typically at the level of a few percent. As conclusion a summary of the b-tagging algorithm performance, using algorithmic definition, for three different values of b-tagging efficiency is given in Tab. 5.1 for the most dangerous background $t\bar{t}2j$. The results reported in Tab. 5.1 shows that with a more realistic detector simulation, different from the "fast-simultation" used in the past [60, 61, 62], the b-tagging performance are worst respect to previous studies. As we will see in the next chapter this has an important effect on

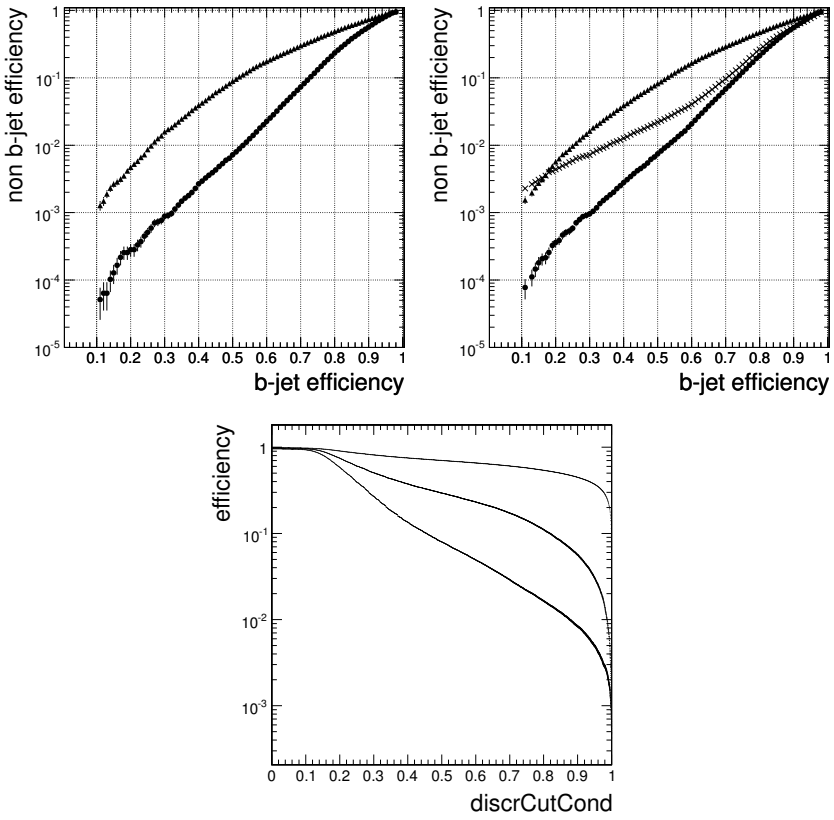


Figure 5.4: On the top left: Non-b-jet mistagging efficiency versus b-jet tagging efficiency for c-jets (triangles), and uds-jets (stars) for the $t\bar{t}H$ sample with $m_H = 120$ GeV/c^2 and jets with a minimum transverse momentum of 20 GeV/c. For this plot the "physics definition" of the original jet flavour has been used. In this definition, jets originating from the splitting of gluons into heavy flavour pairs are labelled as gluons leading to a significantly reduced rejection power for gluon jets. There are no original gluon jets in the $t\bar{t}H$ sample. On the top right: The corresponding plot for the $t\bar{t}2j$ sample, where gluon jets are represented by crosses. On the bottom: Tagging efficiency for b-jets (topmost curve), and Mis-tagging efficiencies for c-jets (middle) and uds-jets (bottom) as a function of the discriminator cut for the $t\bar{t}2j$ sample.

background rejection. A possible improvement on b-tagging performance can

b-tag ϵ	uds-tag ϵ	g-tag ϵ	c-tag ϵ
50%	0.8%	2%	9%
60%	2%	4%	18%
70%	7%	10%	30%

Table 5.1: Summary of the secondary vertex tag perfomances for a sample of $t\bar{t}2j$ events.

be achivied including lepton information in the combined secondary vertex tag. A separate algorithm in CMS, the *Soft Lepton* b-tagging [65], exploits the presence of leptons in b-jets that arise in semileptonic decays of B hadrons. This is achieved by means of a discriminator that is constructed for jets with non-isolated leptons. Combining the secondary vertex tag with the *Soft Lepton* b-tagging could have various potential benefits including a relative improvement that, in a preliminay study, is estimate to be about 10% for the rejection of light flavour jets. However this improvement in tagging were not used in the analysis presented in this work.

Another important consideration for the b-tagging algorithm is that the b-tagging performance falls relative quickly as the momentum scale of jets does. This is mainly due to a poor impact parameter resolution due to multiple scattering for low p_T tracks. This is highlighted in Fig 5.5 where the misidentification efficiency for c, uds and g jets, for a fixed b-tagging efficiency of 50%, is given as a function of jet p_T.

Since the $t\bar{t}H$ channel provides a large number of low p_T jets, high b-tagging performances are not easy to obtain.

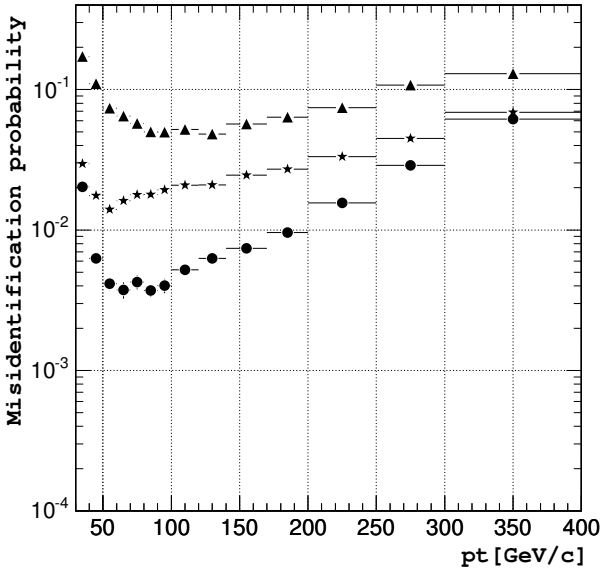


Figure 5.5: The non-b jet mistagging probabilities for a fixed b-jet tagging effciency of 50% as a function of the jet transverse momentum (jets with $|\eta| < 2.4$ were considered) for c-jets (triangles), uds-jets (circles) and $gluon$ jets (stars) obtained for jets in an event sample of QCD events for the combined secondary vertex algorithm.

Chapter 6

$t\bar{t}H \rightarrow b\bar{b}b\bar{b}qq'l\nu_l$: Analysis Method

The $t\bar{t}H$ channel, as discussed in chapter 1, could be a very useful alternative to the $H \rightarrow \gamma\gamma$ channel for the Higgs boson discovery in the mass range $m_H < 135$ GeV. In this work, for the first time in CMS, the Higgs discovery potential, using the associated production with a $t\bar{t}$ pair, will be discussed using a full detector simulation [63]. This means that the simulation of the particle interaction with the detector material and the digitisation of the electronic response is described in high details using the best tools available in the official CMS framework. The more realistic description of the detector response, leads to a more pessimistic picture of the jet resolution and b-tagging performances even if more sophisticated tools have been used with respect to the past. This has a direct effect on both the Higgs invariant mass reconstruction and on the signal significance with respect to the resonant backgrounds. In this chapter the reconstruction method and the signal-background separation will be presented in detail.

6.1 Event generation and simulation

Because the signal identification relies upon the presence of top quark decay products, it comes as no surprise that the most significant backgrounds are those associated with $t\bar{t}$ events themselves. The main backgrounds turn out to be $t\bar{t}$ plus light jets, $t\bar{t}b\bar{b}$ and $t\bar{t}Z$ with $Z \rightarrow b\bar{b}$.

For the $t\bar{t}H$ signal and the irreducible $t\bar{t}b\bar{b}$ background generation compHEP [53] plus PYTHIA [48] are used. To produce $t\bar{t}$ plus n jets background events correctly is necessary to match a matrix element generator with the parton showering of PYTHIA. This is not trivial because the soft QCD ef-

fects represented by the PYTHIA parton shower program are not completely distinct from the higher order perturbative diagrams. While there are jet energies for which the two are clearly distinct, they nevertheless represent two extremes in a continuum and so one is forced to artificially place a boundary in the phase space. A given $(N+1)$-jet event can be obtained in two ways: from the collinear/soft-radiation evolution of an appropriate $(N+1)$-parton final state or from an N-parton configuration where hard, large-angle emission during its evolution leads to the extra jet. A factorisation prescription, in this case often called "matching scheme" [49], defines, on an event-by-event basis, the paths from which the extra jet have to come from. The primary goal of a matching scheme is therefore to avoid double counting by preventing some events to appear twice.

In the present analysis, the $t\bar{t}$ plus n jets is produced using ALPGEN [50] + PYTHIA for the matrix elements and parton showering, respectively. The matching procedure used by ALPGEN to avoid double counting is called MLM method and a detailed description can be found in Ref. [49]. The phase space, between parton generation and shower evolution is selected by two main parameters. Parton-level events are defined by a minimum E_T threshold, E_T^{min} for partons and a minimum separation among them, $\Delta R_{p-p} > R_{min}$. After evolution, a jet cone algorithm with cone size of R_{min} and a minimum transverse energy E_T^{min} is applied to the final state. The event is considered fully matched if each parton produced at matrix element has a matched jet. In the exclusive samples only events with n jets equal to n partons are allowed, while in the inclusive sample it is possible to have a number of jets greater than the number of final state partons. The parameters chosen for the generated samples in this analysis are: $E_T^{min} = 20$ GeV and $R_{min} = 0.7$. It has to be noted that the choice of the matching parameters could introduce systematics in the cross section; yet a more important uncertainty comes from the cross section at matrix element level, because for these events next-to-leading order calculations are not yet available.

The primary Monte Carlo data samples used in this analysis are summarised in Table 6.1 along with some of their generation information. The next-to-leading order (NLO) signal cross-sections for different Higgs mass hypotheses are listed in Table 6.2 together with the branching ratios of $\mathrm{H} \rightarrow \mathrm{b\bar{b}}$ [68].

A generator level filter, described below, was applied to the $\mathrm{t\bar{t}b\bar{b}}$ and $\mathrm{t\bar{t}Z}$ events. In order to apply requirements on jet kinematics, jet reconstruction was performed at generator level with the PYCELL routine in PYTHIA. The filter consists of the following cuts: $\sum_{jet} E_T^{jet}$ greater than 300 GeV, E_T of the two most energetic jets in the event must be above 40 GeV and the E_T of the third and fourth leading jets must be above 30 GeV. Background events

Channel	Generator	PDF
Semi-leptonic $t\bar{t}H$	CompHEP + PYTHIA 6.215	CTEQ4L
$t\bar{t}b\bar{b}$	CompHEP + PYTHIA 6.215	CTEQ4L
$t\bar{t}Z$	CompHEP + PYTHIA 6.215	CTEQ4L
$t\bar{t}Nj$	ALPGEN 2 + PYTHIA 6.325	CTEQ5L

Table 6.1: Generator and parton density functions (PDFs) used for producing signal and background datasets.

m_H	115 GeV	120 GeV	130 GeV
σ_{NLO} (pb)	0.747	0.664	0.532
$BR(H \to b\bar{b})$	0.731%	0.677%	0.525%

Table 6.2: NLO signal cross-sections and $H \to b\bar{b}$ branching ratios for different Higgs mass hypotheses.

were passed on to the full simulation only if they pass the above filter or if they have an isolated lepton with p_T >10 GeV in the pseudorapidity range $|\eta| < 3$, and the summed p_T of particles inside a cone with $\Delta R = 0.2$ around the lepton direction is below 1 GeV. The leading order CompHEP cross-sections of background processes together with the effective cross-sections after the generator filters are listed in Table 6.3.

	$t\bar{t}b\bar{b}$	$t\bar{t}Z$
σ_{LO} (pb)	3.28	0.65
$\sigma_{LO} \times \epsilon$ (pb)	2.82	0.565

Table 6.3: LO CompHEP cross-sections and effective cross-sections after the generator filters of the considered background processes.

For the ALPGEN sample, $t\bar{t}$ +1j exclusive, (i.e one and only one extra jet), $t\bar{t}$ +2j exclusive, $t\bar{t}$ +3j exclusive, and $t\bar{t}$ +4j inclusive, (i.e. with at least 4 extra jets) no kinematic filter has been used at generator level. The only filter applied to the $t\bar{t}$ plus n jets events is a veto in the additional jets originate from b quarks. They are expected to be uncorrelated with events from the $t\bar{t}b\bar{b}$ event sample, although a small amount of double counting could be present due to the gluon splitting into $b\bar{b}$. The ALPGEN cross sections for the different jet multiplicity processes are listed in Table 6.4.

The particle detector interaction are simulated using a CMS program

	exclusive $t\bar{t}$+1j	exclusive $t\bar{t}$+2j	exclusive $t\bar{t}$+3j	inclusive $t\bar{t}$+4j
σ_{LO} (pb)	170	100	40	61

Table 6.4: LO ALPGEN cross-sections for the different jet multiplicity samples.

called CMSIM, which provides a full GEANT3 [64] (for the CompHEP + PYTHIA samples) or GEANT4 [66] (ALPGEN samples) simulation of the CMS detector. The detector response was digitised with low luminosity pile-up, $\mathcal{L} = 2 \times 10^{33}$ cm^{-2} s^{-1}, using the official CMS reconstruction software (ORCA) [63].

6.2 Level 1 and high level trigger selections

A dedicated $t\bar{t}$H trigger was not available and therefore could not be implemented in the analysis. As a result, it is assumed in what follows that the signal is recorded by the CMS Level 1 (L1) and High Level Triggers (HLT) as described in Ref. [42, 43]. The semi-leptonic channels thus use the single muon or single electron trigger, using a p_T threshold of 19 GeV or 29 GeV, respectively. As described in section 2.3, the muon trigger uses the isolation tool while the electron trigger does not use the isolation in HLT level because still not available at the time when this study has been performed. Especially for the electron, the possibility to have a trigger with a lower p_T threshold will help in recover signal efficiency.

The efficiencies quoted are determined by counting the numbers of accepted events relative to the total number of events in these samples. It has to be noted that the trigger efficiencies are defined with respect to exclusive signal samples (pure semi-leptonic (μ) or (e) events) and inclusive background samples (all W decays allowed).

As a final comment, it should be noted that during data taking, more sophisticated triggers will certainly be available and could implemented for this search. They would combine such elements of the signatures of these events as the presence of leptons, missing transverse energy (E_T^{Miss}), multiple jets, and possibly even some loose b-tagging information. As such, the efficiency for collecting these events on tape will very likely be much higher than implied by the results shown here.

Efficiency (%) per channel	Single μ	Single e
$H \to b\bar{b}$ with $m_H = 120$GeV	63.5	52.4
$t\bar{t}b\bar{b}$	19.0	16.1
$t\bar{t}1j$	13.9	11.3
$t\bar{t}2j$	14.0	11.1
$t\bar{t}3j$	14.0	11.1
$t\bar{t}4j$	13.4	11.1
$t\bar{t}Z$	20.4	18.8

Table 6.5: Signal and background efficiencies of the Level 1 and High Level Triggers.

6.3 Event selection and reconstruction

The strategy for selecting $t\bar{t}H$ events with one isolated muon or electron in the final state can be summarised in the following three steps [31]:

- preselection
- choice of jet pairing
- selection

The preselection requires, besides the HLT stream for a single muon or a single electron, isolated leptons using the likelihood method described in chapter 3 and a jet selection based on kinematic and b-tagging requirements. Then, using the identified lepton with the highest probability to come from the W decay and the missing transverse energy, the longitudinal component of the neutrino momentum p_z is computed from four-momentum conservation in the W boson decay.

In the choice of jet pairing step we determine out of six or more jets in an event, those maximising the probability of assignment to a top quark or Higgs decay.

After the jet assignment a tighter selection is applied to maximise the discovery potential for a period of data taking corresponding to an integrated luminosity of 60 fb^{-1}.

In the following the $t\bar{t}H \to b\bar{b}b\bar{b}qq'\mu\nu_\mu$ channel will be named single muon channel, while the $t\bar{t}H \to b\bar{b}b\bar{b}qq'e\nu_e$ single electron channel.

6.3.1 Preselection

After the HLT, the lepton identification is performed with the likelihood selection. The likelihood is used not only to identify the lepton with the

highest probability to come from the W decay but also to reject background events with dileptonic W decays. The likelihood selection is performed requiring at least a muon with $-Log(L_\mu) < 1.4$ or at least an electron with $-Log(L_e) < 1.2$. The likelihood veto to reject possible dileptonic events is used in this way:

- Single muon channel: electron veto if an electron with $-Log(L_e) < 1.2$ is found in the event; muon veto if a second muon candidate with $-Log(L_\mu) < 1.4$ is found in the event.
- Single electron channel: muon veto if a muon with $-Log(L_\mu) < 1.4$ is found in the event; electron veto if a second electron candidate with $-Log(L_e) < 1.2$ is found in the event.

In the case of single muon channel, the distribution of the $-Log(L_\mu)$ for the second muon candidate and the distribution of $-Log(L_e)$ for the first electron candidate are shown in Fig. 6.1 for a pure semi-leptonic (μ) sample (solid line) and the inclusive $t\bar{t}$2j background (dashed line). The peak near to zero for the inclusive $t\bar{t}$2j backgrounds shows the presence of event candidates with both Ws decaying leptonically. For the single muon channel the application of the muon and electron vetoes results in a lowering of the signal efficiency by only 2%, while the total background rejection is increased by about 13%. Moreover in the single electron channel, the misidentification of the jet with the isolated electron has been excluded by imposing a veto on the jet if the electron lies inside a jet cone radius of 0.1.

The jet selection is based on cuts both on the number of jets and on the jet transverse energy. Jets are reconstructed using the *iterative cone* algorithm with a cone radius of $\Delta R = 0.5$ and the jet transverse energy is corrected using the official CMS calibrations described in section 4.3. The jet E_T selection plays an important role here because it modifies the jet number distributions and in general the signal efficiency. The E_T of the sixth energetic jet in the event, after the lepton selection, is shown in Fig 6.2. The jets are selected either if the E_T of the sixth jet is greater than 20 GeV in the pseudorapidity region $|\eta| < 3.0$ or if jets have a transverse energy with 10 GeV $< E_t <$ 20 GeV but with at least two tracks pointing to the signal primary vertex within a distance along the z (beam) axis of $|z_{PV} - z_{track}| < 1$ mm. The latter condition is used to help in rejecting jets with low E_T coming from pile-up interactions. As shown in Fig 6.2 there is not any gain in cutting harder on the jet E_T.

In Fig 6.3 we show the number of jets distribution after the lepton selection and the jet E_T cuts described above. Since the request of at least six jets is necessary to reconstruct all the invariant masses, it has also been found

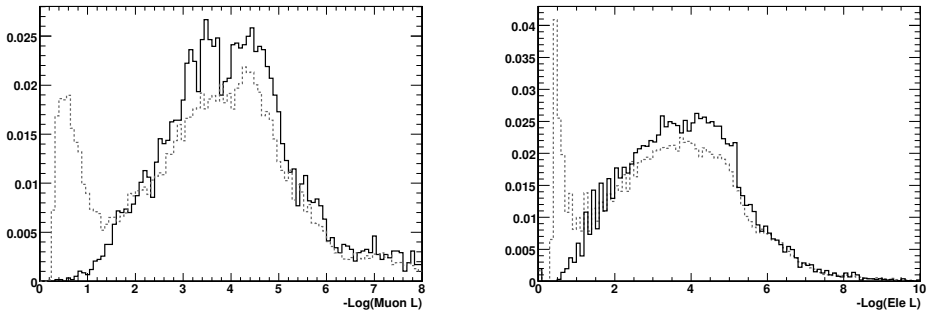


Figure 6.1: Avoiding the appearance of a second lepton in the single μ analysis. On the left: Muon likelihood distributions. On the right: Electron likelihood distributions for a potential second lepton candidate in these events. In both cases solid lines refer to the pure semi-leptonic (μ) sample while the dashed lines refer to the inclusive $t\bar{t}2j$ background. The peaks near to zero clearly highlight the presence of a second muon (left) or an electron (right) in the background samples which allow inclusive decays of the Ws. The distributions are normalised to Unity.

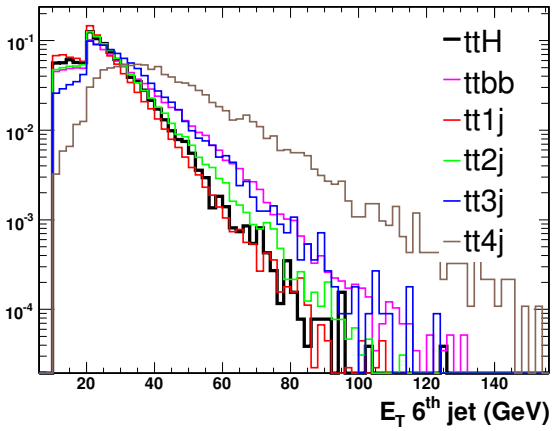


Figure 6.2: E_T distribution of the 6th jet for signal (black) and backgrounds (coloured).

advantageous to put an upper limit of seven jets allowed in the event. This helps both in rejecting the $t\bar{t}$3j and the $t\bar{t}$4j backgrounds and events with multiple hard radiation emission. After the jet selection, to further reject

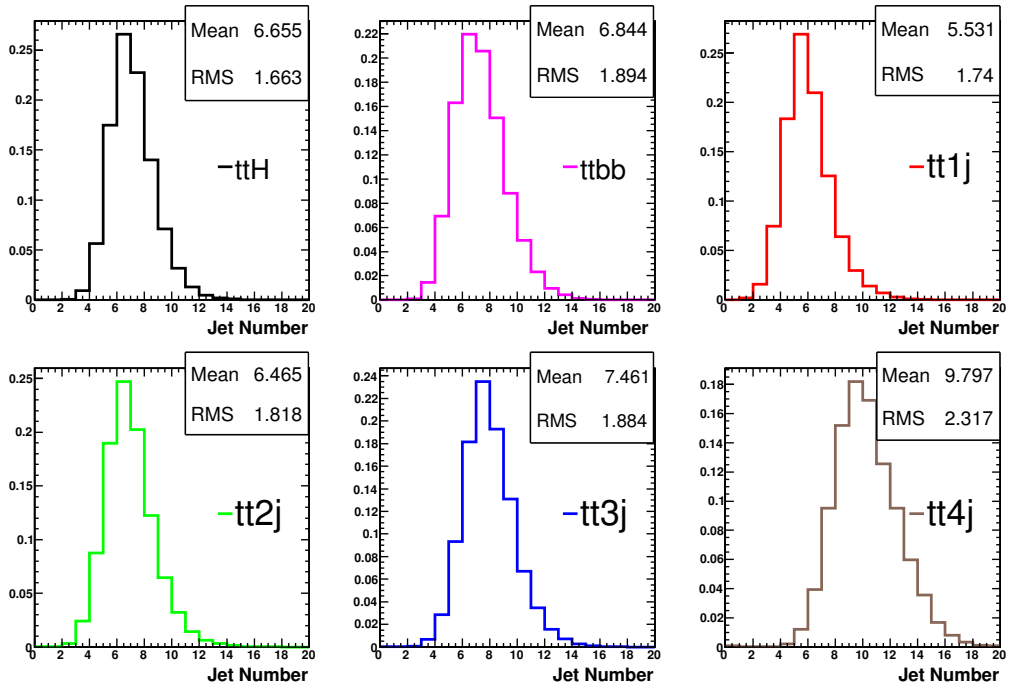


Figure 6.3: Jet number distribution for signal and backgrounds.

the $ttNj$ backgrounds, at least 4 jets are required to be tagged as b-jets with a minimal discriminator value corresponding to about 70% of b-efficiency.

The efficiencies of the preselection cuts described above are summarised in Tab. 6.6 and in Tab. 6.7 for the single muon and single electron channel ($m_H = 120$ GeV), respectively. It has been verified that for backgrounds like $tt1j$ and $tt4j$, that have a different signature with respect to the signal, these cuts are more effective.

6.3.2 Choice of the jet pairing

In order to perform a complete reconstruction of the event, the longitudinal component of the neutrino momentum has to be computed from four-momentum conservation in the W boson decay: $m_W^2 = (E^\mu + E^\nu)^2 - (\vec{p}^\mu + \vec{p}^\nu)^2$. This equation gives 2 real solutions for p_z^ν in 66% of the cases. In the remaining 34%, the neutrino is taken to be collinear with the lepton: $p_z^\nu = p_z^l$. The resolution of p_z^ν before and after the collinear approximation is shown

muon channel				
	HLT+ Isolated μ	+6, 7 jets	+Lepton Veto	+ 4Btag (70% b-ϵ)
$t\bar{t}H$	62.67	30.88	30.24	5.64
$t\bar{t}b\bar{b}$	15.87	6.61	5.61	0.78
$t\bar{t}1j$	13.70	4.28	3.82	0.087
$t\bar{t}2j$	13.57	5.89	5.10	0.152
$t\bar{t}3j$	13.27	5.63	4.61	0.154
$t\bar{t}4j$	12.72	1.90	1.44	0.052
$Zt\bar{t}$	17.43	6.77	4.68	0.230

Table 6.6: Preselection efficiency for the single muon channel ($m_H = 120$ GeV).

electron channel				
	HLT+ Isolated e	+6, 7 jets	+Lepton Veto	+ 4Btag (70% b-ϵ)
$t\bar{t}H$	49.13	24.26	23.65	4.17
$t\bar{t}b\bar{b}$	12.17	5.01	4.77	0.56
$t\bar{t}1j$	10.83	3.13	2.83	0.076
$t\bar{t}2j$	10.72	4.55	3.99	0.124
$t\bar{t}3j$	10.80	4.68	3.88	0.136
$t\bar{t}4j$	10.55	1.72	1.28	0.040
$Zt\bar{t}$	14.06	5.45	3.92	0.200

Table 6.7: Preselection efficiency for the single electron channel ($m_H = 120$ GeV).

in Figure 6.4. A small degradation in the longitudinal resolution is observed but the reconstruction efficiency of the leptonic W boson decay is increased to 100%.

If the invariant mass of the W decaying leptonically is imposed by the four-momentum conservation to extract the neutrino p_z, the invariant masses of the W decaying hadronically and of the 2 top quarks have to be extracted from the combination of the 6 or 7 jets accepted in the analysis. The Higgs invariant mass is then reconstructed with the jets having the highest b-tagging probability and that have not been used to reconstruct the intermediate W and top quark masses. For this purpose, different methods have been tested: a simple χ^2, a likelihood method using mass distribution functions for right and bad jet combinations, a kinematic fit with mass constraints and a combined method that uses kinematic information, b-tagging probability and kinematic fit with mass constraints. With these methods the maximum pairing efficiency to find the right b-jets coming from the expected b-quarks of the Higgs decay, is about 30%. Thus, because 70% of the events have an

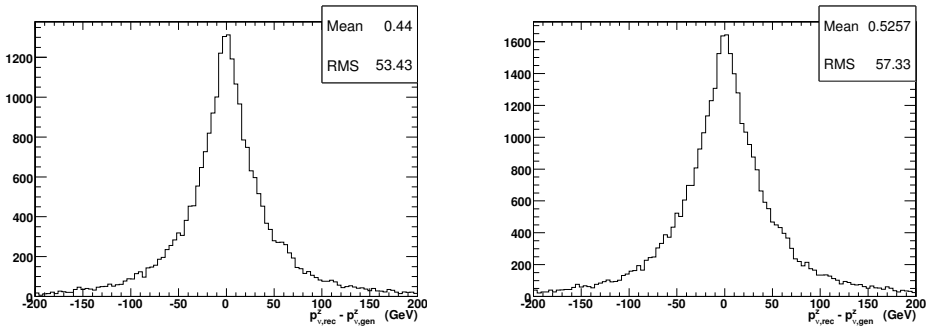


Figure 6.4: Longitudinal resolution of the neutrino. On the left: Only those cases in which there are real solutions obtained for the W boson's four-momentum constraint. On the right: Real plus collinear solutions as described in the text.

Higgs boson invariant mass that follows the distribution of the combinatorial background, the Higgs invariant mass results in a broad peak.

The method that we will present in this work for the complete reconstruction of semi-leptonic $t\bar{t}H$ events is similar to that used in Reference [29, 67]. In order to choose the optimal jet combination for the top quark mass reconstruction a likelihood (L_{Event}) is defined by using masses, b-tagging and kinematic information from the whole event:

$$L_{Event} = L_{Mass} \times L_{bTag} \times L_{Kine}. \tag{6.1}$$

To increase the speed of the algorithm while also enhancing the rejection of events that are badly reconstructed, only those combinations with mass resonances within 3σ of expectation are considered in the jet assignments. The mean values and widths of the invariant masses used in this procedure are taken from the values obtained with jet-parton matching. The corresponding distributions are shown in Figure 6.5.

The mass information considered in the likelihood L_{Mass} is the probability of the kinematic fit with invariant mass constraints (top quarks and hadronically decaying W) as described in details in Reference [69]. A general overview of the method and the improvements with respect to a simple χ^2 method are summarised in Appendix A. In $t\bar{t}H$ events, a mass constraint on the hadronically decaying W improves the probability to correctly identify the b-jet coming from the top quark decay and hence also improve the chance of correctly finding the b-jets originating from the Higgs boson decay. However, cause the relative low improvement in the jet assignment, the kinematic fit with mass constraints is found to have a little effect on the overall result of the analysis.

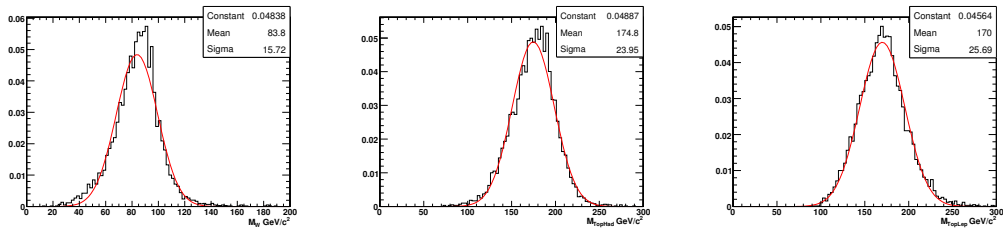


Figure 6.5: Invariant masses of the hadronically decaying W boson and the hadronically and leptonically decaying Top quarks using jet-parton matching with $\Delta R_{j-p} < 0.3$.

The b-tagging function L_{bTag} is defined as the product of the b-tag discriminators:

$$L_{bTag} = D_{Top_{Had}} \times D_{Top_{Lep}} \times D_{H_1} \times D_{H_1} \times (1 - D_{W1}) \times (1 - D_{W2}) \quad (6.2)$$

where Top_{Had} and Top_{Lep} are expected to be the two b-jets from the hadronic and leptonic top quarks, respectively, while H_1 and H_2 are expected to be the two b-jets coming from Higgs and W_1 and W_2 are of course the two jets from the hadronically decaying W boson. As expected, high values of this variable correspond to combinations where b-jets are assigned to the top quarks and Higgs, and light quark jets are assigned to the W.

The kinematic function takes into account the observation that the b-jets coming from top quarks tend to be slightly more energetic than b-jets coming from the Higgs boson. It is defined as:

$$L_{Kine} = \arctan\left(\frac{4(E_{Top_{Had}} + E_{Top_{Lep}} - E_{H_1} - E_{H_2})}{(E_{Top_{Had}} + E_{Top_{Lep}} + E_{b1H} + E_{b2H})}\right) \times \frac{1}{\pi} + \frac{1}{2} \quad (6.3)$$

Among all possible combinations of jet-parton assignments, the one with the highest value of L_{Event} is chosen for use in the final reconstruction of the top quarks. To understand the difficulty to find the right b-jets from the Higgs decay, it has to be noted that the number of combination for each event is:

$$N = \begin{pmatrix} n \\ 4 \end{pmatrix} \times 4! \times \frac{1}{2} \times 2 \quad (6.4)$$

that is equal to 360 for $n = 6$ jets and 840 for $n = 7$ jets in the final state. The factor of $\frac{1}{2}$ is for the two jets assumed to come from the hadronic W (exchanging the jets gives the same invariant mass), while the factor of 2 is for the two solutions of neutrino p_Z. Of course in the case of linear approximation for the neutrino, the factor of 2 must not be applied. Finally, the invariant mass of the two remaining jets with highest b-tagging discriminator values is used to reconstruct the Higgs mass.

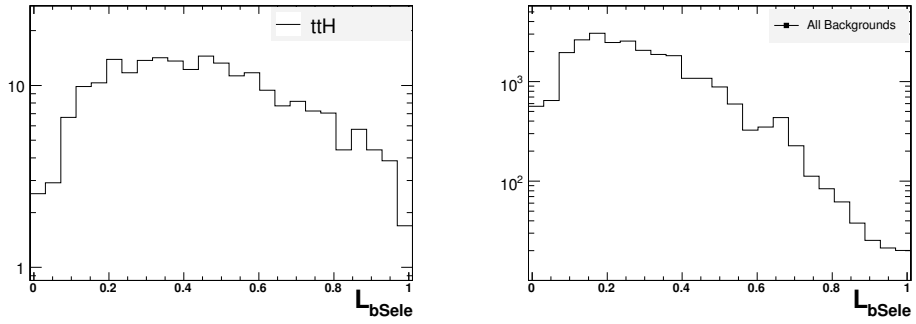


Figure 6.6: L_{bSele} selection cut for t$\bar{\text{t}}$H with $W \rightarrow qq', W \rightarrow \mu\nu$ (on the left) and for background (on the right).

6.3.3 Selection

After the jet assignment just described is complete, criteria are applied to further reject background while pure $t\bar{t}H$ events are selected. In particular, a stronger b-tag requirement is applied on the event variable L_{bSele} given by the combination of b-tag discriminators: $L_{bSele} = D_{Top_{Had}} \times D_{Top_{Lep}} \times D_{H_1} \times D_{H_2}$ where $D_{Top_{Lep}}$, $D_{Top_{Had}}$, D_{H_1} and D_{H_2} are the discriminators associated to the Higgs and top quark jets found by the choice of the jet pairing of the previous step of the analysis.

The variable L_{bSele} is shown in Figure 6.6 for the single muon channel as well as for all backgrounds combined. The distributions are normalised to the number of expected events in 60 fb^{-1}. To optimise the background rejection, the signal significance and the signal to background ratio as a function of the selection cut L_{bSele} has been studied. The result are shown in Fig. 6.7 for the single muon channel and Fig. 6.8 for the single electron channel. In these figures, the backgrounds that give 0 remaining events are not taken into account. The behaviour of the signal significance versus different values of L_{bSele} is shown to be fairly stable in the L_{bSele} cut range between 0.45 and 0.75. Cutting hard on L_{bSele} helps to increase S/B but is achieved at the cost of a significant reduction in signal statistics. For a cut on L_{bSele} of 0.55, the reconstructed hadronic W boson mass and the leptonic and hadronic top quark masses are shown in Fig 6.9 for the semileptonic (μ) $t\bar{t}H$ channel. The mean values of the invariant masses are very similar to that one reconstructed using parton matching criteria.

The distributions of reconstructed mass of the Higgs boson for final selected events are shown in Figures 6.10 and 6.11 for signal only (left) and for the combination of the different backgrounds (right) for the muon and electron channels, respectively. The grey histogram on the left distribution

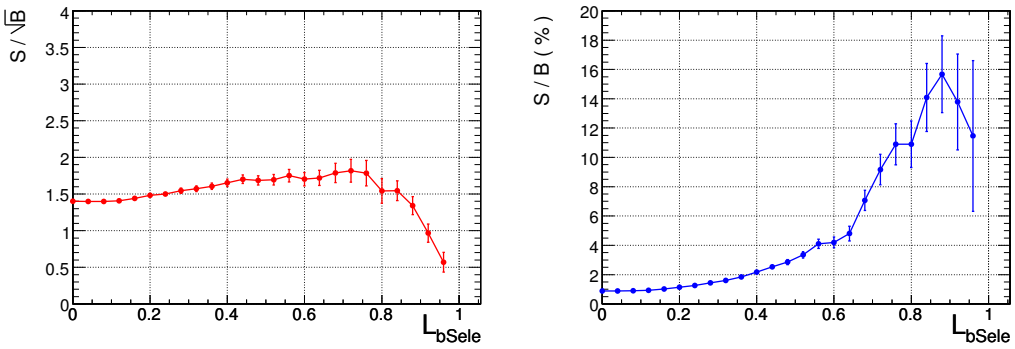


Figure 6.7: $t\bar{t}H$ ($W \to qq', W \to \mu\nu$): Signal Significance (left) and Signal to Background ratio (right) as function of the cut on L_{bSele}.

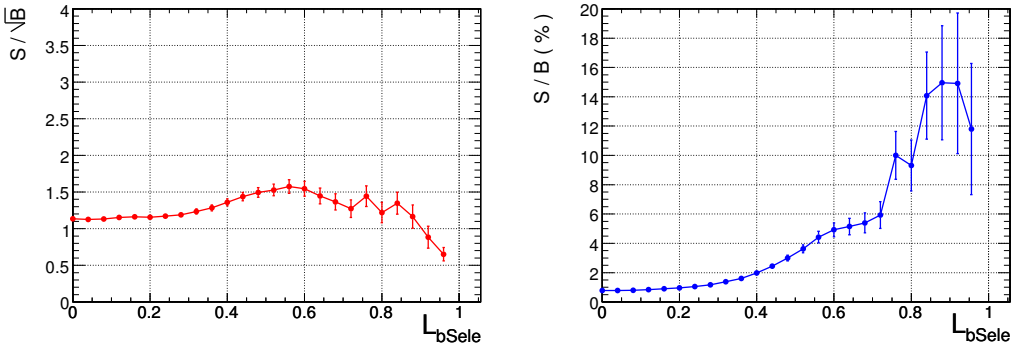


Figure 6.8: $t\bar{t}H$ ($W \to qq', W \to e\nu$): Signal Significance (left) and Signal to Background ratio (right) as function of the cut on L_{bSele}.

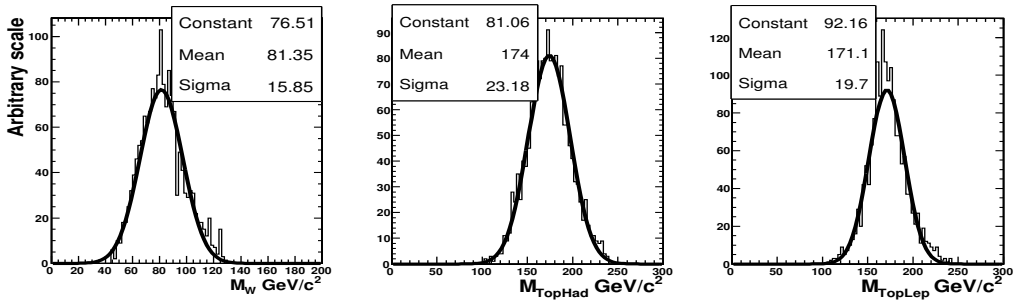


Figure 6.9: Reconstructed hadronic W boson mass and the leptonic and hadronic top quark masses for a cut on L_{bSele} of 0.55 in the case of semileptonic (μ) $t\bar{t}H$ channel.

of Figure 6.10 represents the events where at least one jet has been wrongly assigned to the Higgs boson. The fraction of signal events where the two b-jets are correctly assigned to the Higgs boson (i.e. the pairing efficiency), for the L_{bSele} cut of 0.55, is roughly 31% in the muon channel and about 29% for the electron channel.

6.3.4 Results

The preselection and selection efficiencies with the corresponding numbers of expected events and signal significances are reported in Table 6.8 and Table 6.9 for the channel with a muon or an electron in the final state. The number of expected events is computed for an integrated luminosity of 60 fb^{-1} in the Standard Model Higgs mass range from 115 to 130 GeV. In this table, the results are presented for two different working points. The tight cut for $L_{bSele} > 0.75$ gives better results for the final signal background separation, but it suffers from a very small number of remaining background events, due to the limited Monte Carlo Statistics. For the determination of the shape of the invariant mass (Fig. 6.10) as well as for the systematic uncertainties evaluation (Section 6.4), a looser cut at $L_{bSele} > 0.55$ has been chosen. In conclusion in Tab. 6.10 the signal discovery potential for the single muon and single electron are combined together. The combined results can be straightforward obtained from Tab. 6.8 and Tab. 6.9 simply adding the signal and background events. This is possible because the lepton selection and veto create distinct samples for the single muon and electron channel.

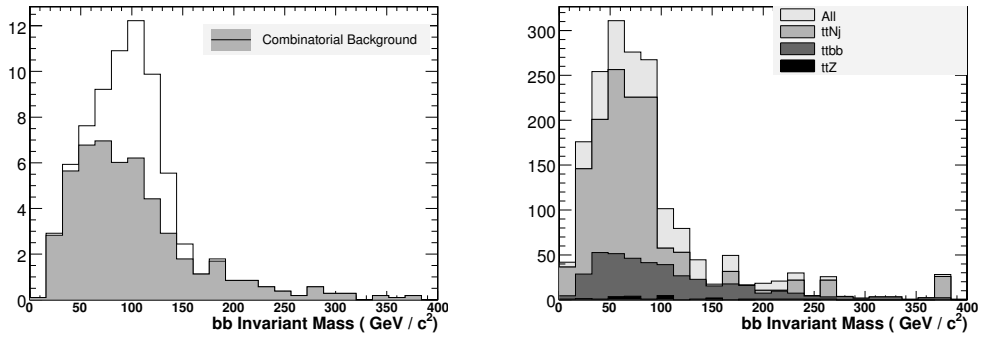


Figure 6.10: $t\bar{t}H$ ($W \to qq', W \to \mu\nu$). Left: Invariant $b\bar{b}$ mass for signal only (combinatorial background is shaded grey). Right: The sum of the reconstructed $m_{b\bar{b}}$ spectra for backgrounds with a value of $L_{bSele} > 0.55$. The distributions are normalised to an integrated luminosity of 60fb^{-1} .

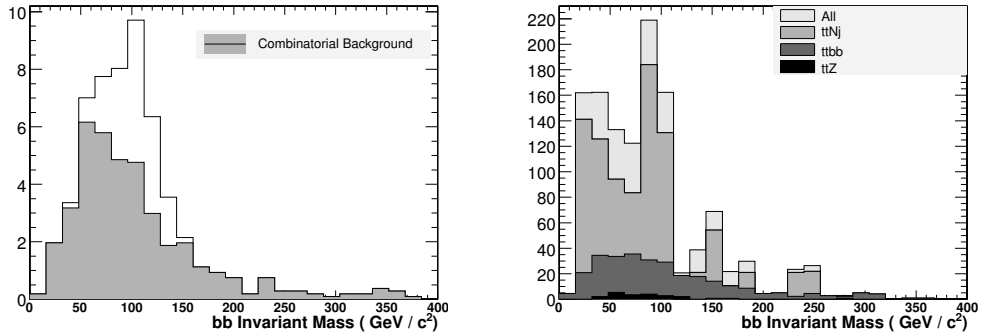


Figure 6.11: $t\bar{t}H$ ($W \to qq', W \to e\nu$). Left: Invariant $b\bar{b}$ mass for signal only (combinatorial background is shaded grey). Right: The sum of the reconstructed $m_{b\bar{b}}$ spectra for backgrounds with a value of $L_{bSele} > 0.55$. The distributions are normalised to an integrated luminosity of 60fb^{-1}. The large fluctuations in the $t\bar{t}$Nj background are due to a very limited number of remaining events that have to be scaled by large factors.

muon channel					
	Analyzed Ev.	ϵ_{loose} (%)	N^{ev}_{loose} 60fb^{-1}	ϵ_{tight} (%)	N^{ev}_{tight} 60fb^{-1}
$t\bar{t}H$ (115)	27768	2.00 ± 0.08	96 ± 4	0.80 ± 0.05	38 ± 3
$t\bar{t}H$ (120)	41929	1.90 ± 0.07	75 ± 3	0.74 ± 0.04	29 ± 2
$t\bar{t}H$ (130)	19466	2.23 ± 0.11	55 ± 3	0.84 ± 0.07	21 ± 2
$t\bar{t}b\bar{b}$	372737	0.247 ± 0.008	419 ± 14	0.0877 ± 0.0048	148 ± 8
$t\bar{t}1j$	393000	0.0051 ± 0.0011	520 ± 120	0.00076 ± 0.00044	78 ± 45
$t\bar{t}2j$	568999	0.0105 ± 0.0014	633 ± 82	0.00070 ± 0.00035	42 ± 21
$t\bar{t}3j$	101000	0.0050 ± 0.0022	119 ± 53	0.	$< 27(68\% C.L)$
$t\bar{t}4j$	86697	0.0035 ± 0.0020	126 ± 73	0.	$< 48(68\% C.L.)$
$Zt\bar{t}$	50000	0.068 ± 0.012	23 ± 4	0.026 ± 0.007	9 ± 2
Total Background			1840		< 352
$S/\sqrt{B}$ (115)			2.2		2.0
S/B (115)			5.1%		10.8%
$S/\sqrt{B}$ (120)			1.8		1.6
S/B (120)			4.1%		8.2%
$S/\sqrt{B}$ (130)			1.3		1.1
S/B (130)			3.0%		6.0%

Table 6.8: Selection efficiency for $L_{bSele} > 0.55$ (ϵ_{loose}) and for $L_{bSele} > 0.75$ (ϵ_{tight}), number of expected events and signal significance in 60 fb^{-1} for the muon and electron $t\bar{t}H$ channels. The signal datasets are labeled by the generated Higgs mass in GeV (parentheses). Also quoted are binomial errors arising from the finite sizes of processed datasets. All numbers refer to the complete Higgs mass range.

electron channel					
	Analyzed Ev.	ϵ_{loose} (%)	N^{ev}_{loose} 60fb^{-1}	ϵ_{tight}(%)	N^{ev}_{tight} 60fb^{-1}
$t\bar{t}H$ (115)	27692	1.39 ± 0.07	66 ± 3	0.52 ± 0.04	25 ± 2
$t\bar{t}H$ (120)	42228	1.42 ± 0.06	56 ± 2	0.53 ± 0.04	21 ± 1
$t\bar{t}H$ (130)	19127	1.57 ± 0.09	39 ± 2	0.61 ± 0.06	15 ± 1
$t\bar{t}b\bar{b}$	372737	0.176 ± 0.007	297 ± 12	0.0641 ± 0.0041	109 ± 7
$t\bar{t}1j$	393000	0.0038 ± 0.0010	390 ± 100	0.00025 ± 0.00025	26 ± 26
$t\bar{t}2j$	568999	0.0067 ± 0.0011	401 ± 65	0.00123 ± 0.00046	74 ± 28
$t\bar{t}3j$	101000	0.0040 ± 0.0020	95 ± 48	0.	$< 27(68\% C.L)$
$t\bar{t}4j$	86697	0.0023 ± 0.0016	84 ± 60	0.	$< 48(68\% C.L.)$
$Zt\bar{t}$	50000	0.064 ± 0.011	22 ± 4	0.022 ± 0.007	7 ± 2
Total Background			1289		< 291
$S/\sqrt{B}$ (115)			1.8		1.5
S/B (115)			5.1%		8.6%
$S/\sqrt{B}$ (120)			1.6		1.2
S/B (120)			4.4%		7.2%
$S/\sqrt{B}$ (130)			1.1		0.9
S/B (130)			3.0%		5.2%

Table 6.9: Selection efficiency for $L_{bSele} > 0.55$ (ϵ_{loose}) and for $L_{bSele} > 0.75$ (ϵ_{tight}), number of expected events and signal significance in 60 fb^{-1} for the muon and electron $t\bar{t}H$ channels. The signal datasets are labeled by the generated Higgs mass in GeV (parentheses). Also quoted are binomial errors arising from the finite sizes of processed datasets. All numbers refer to the complete Higgs mass range.

	N^{ev}_{loose} 60fb^{-1}	N^{ev}_{tight} 60fb^{-1}
$t\bar{t}H$ (115)	162	63
$t\bar{t}H$ (120)	131	50
$t\bar{t}H$ (130)	94	36
Total Background	3129	< 643
$S/\sqrt{B}$ (115)	2.9	2.5
S/B (115)	5.2%	9.8%
$S/\sqrt{B}$ (120)	2.3	2.0
S/B (120)	4.2%	7.8%
$S/\sqrt{B}$ (130)	1.7	1.4
S/B (130)	3.0%	5.6%

Table 6.10: Number of expected events and signal significance in 60 fb^{-1} for the combined muon and electron $t\bar{t}H$ channels. The signal datasets are labeled by the generated Higgs mass in GeV (parentheses). All numbers refer to the complete Higgs mass range.

6.4 Discussion of systematic uncertainties

The uncertainties in various quantities are estimated based on the expected knowledge of the CMS experiment when 60 fb^{-1} of data have been collected. In according to other CMS analyses, the following "standard" sources of systematic error are considered:

- Jet energy scale (JES)
- Jet resolution
- b-jet and c-jet tagging efficiencies
- uds-jet tagging efficiencies
- Luminosity

The systematics listed above are assumed to be uncorrelated. Each is varied independently according to procedures detailed below. Each variation produces a change in the selection efficiency $\Delta\epsilon$ and the corresponding change in expected event yields ΔN_X ($X = t\bar{t}H, t\bar{t}1j, ...$) for the signal and background. In cases where the "up" and "down" variation produces an asymmetric range for $\Delta\epsilon$, the range is symmetrized by taking the maximum of the two. The change in yields for all backgrounds are then added in quadrature to obtain the total expected change in background yields ΔN due to all systematics.

Following commonly agreed prescriptions for CMS analyses, the JES uncertainty is taken into account by shifting the calibrated energies of each jet up (down) by a relative 10% for jets with transverse momentum less than 20 GeV. For jets with transverse momentum between 20 GeV and 50 GeV, the relative variation falls linearly with increasing transverse momentum from 10% to a final 3%. Jets with transverse momentum greater than 50 GeV are assumed to have a flat 3% JES uncertainty [70].

The jet resolution should be smeared by an overall 10% [70]. This is achieved by shifting the energy of each jet by a random number drawn from a Gaussian distribution of mean zero and width σ_{jet}:

$$E_T^{jet} \rightarrow E_T^{jet} + Gaus(0, \sigma_{jet}) \tag{6.5}$$

σ_{jet} is taken to be 10% of the jet energy, $\sigma_{jet} = 0.1\ E_T^{jet}$.

For purposes of the b-tagging systematic, all jets in the event are categorised by flavor according to Monte Carlo truth. When the true flavor of the jet has been determined, the following relative uncertainties in the tagging efficiencies of jets of the various flavors are considered [71]:

- 4% for b- and c-jets
- 10% for u-, d-, and s-jets

As decided within group, the tagging efficiencies are varied simultaneously for b-jets and c-jets, taking into account that each uncertainty is likely to be fully correlated between them. Similarly but in a separate run, the u-, d-, and s-jet tagging efficiencies are varied together by the stated amount. The procedure for affecting a relative uncertainty of $\delta\varepsilon$ in tagging efficiency is to throw a flat random number for each jet (in the category of interest), and untagging it a fraction $\delta\varepsilon$ of the time. Similarly, the tagging efficiency can be varied upwards by tagging untagged jets a fraction $\delta\varepsilon \times \varepsilon_{tag}/(1 - \varepsilon_{tag})$ of the time, where ε_{tag} is the tagging efficiency for that category. This factor is required to properly scale the un-tagging probability so as to affect an increase that is a $\delta\varepsilon$ relative to ε_{tag}. The above procedure assumes that any correlation between the b, c-tagging and uds-tagging systematics is small.

Lastly, the systematic uncertainty in luminosity is assumed to be 3%, from which a corresponding change in yield is calculated.

The systematic effects are evaluated for the single muon channel only. Since none of the sources of uncertainty considered involve lepton identification, the results are assumed to be applicable to the single electron channel as well. The effect of the various systematic uncertainties on the selection efficiencies for the muon analysis are shown in Tab. 6.11. In Tables 6.12

and 6.13, the systematic uncertainties are propagated to the expected signal significance.

All these uncertainties have been calculated for the loose working points of the analyses, since the tight working points do not give enough remaining background events to reasonably quantify a systematic uncertainty.

semi-leptonic	JES+jet res	bc-tag	uds-tag	Σ	# events ϵ_{loose}	# events ϵ_{tight}
$t\bar{t}H$ (115)	20.1	10.7	1.0	23.0	22	9
$t\bar{t}H$ (120)	18.9	10.9	0.8	22.0	17	6
$t\bar{t}H$ (130)	17.8	9.4	0.7	20.4	11	4
$t\bar{t}b\bar{b}$	16.6	9.5	0.8	19.4	81	29
$t\bar{t}$1j	33.9	14.3	5.0	37.3	193	29
$t\bar{t}$2j	22.5	7.2	7.0	24.8	157	10
$t\bar{t}$3j	15.4	5.0	5.0	17.2	21	5
$t\bar{t}$4j	23.2	5.3	6.6	24.9	31	12
muon channel						
backgrounds ϵ_{loose}	23.7	9.4	4.8	25.9	476	
backgrounds ϵ_{tight}	21.5	9.1	3.6	23.6		83
electron channel						
backgrounds ϵ_{loose}	23.7	9.47	4.67	26.4	339	
backgrounds ϵ_{tight}	20.2	8.01	4.08	22.6		66

Table 6.11: Systematic uncertainties relative to selection efficiencies (in percent) for the semi-leptonic $t\bar{t}H$ channels. Σ is the quadrature sum of all changes in the given row. Σ also includes the 3% uncertainty of the luminosity which is the same for all the samples. The last two colums show the abolute uncertainty (in number of events) at the two different working points ϵ_{loose} and ϵ_{tight}.

muon	S	S/B	$S/\sqrt{B}$	$S/\sqrt{B+dB^2}$
$L_{bSele} > 0.55$ (ϵ_{loose})				
$t\bar{t}H$ (m_H=115 GeV)	96	0.052	2.2	0.20
$t\bar{t}H$ (m_H=120 GeV)	75	0.041	1.8	0.15
$t\bar{t}H$ (m_H=130 GeV)	55	0.030	1.3	0.11
$L_{bSele} > 0.75$ (ϵ_{tight})				
$t\bar{t}H$ (m_H=115 GeV)	38	0.108	2.0	0.44
$t\bar{t}H$ (m_H=120 GeV)	29	0.082	1.6	0.34
$t\bar{t}H$ (m_H=130 GeV)	21	0.060	1.1	0.24

Table 6.12: Significance of muon channel before and after taking into account the uncertainty dB in the total number of background events due to systematics. The result is shown for the two different cuts on L_{bSele}, but assuming the same systematic uncertainties (as computed at the first working point) for both.

electron	S	S/B	$S/\sqrt{B}$	$S/\sqrt{B+dB^2}$
$L_{bSele} > 0.55$ (ϵ_{loose})				
$t\bar{t}H$ (m_H=115 GeV)	66	0.051	1.8	0.20
$t\bar{t}H$ (m_H=120 GeV)	56	0.044	1.6	0.17
$t\bar{t}H$ (m_H=130 GeV)	39	0.030	1.1	0.12
$L_{bSele} > 0.75$ (ϵ_{tight})				
$t\bar{t}H$ (m_H=115 GeV)	25	0.086	1.5	0.37
$t\bar{t}H$ (m_H=120 GeV)	21	0.072	1.2	0.31
$t\bar{t}H$ (m_H=130 GeV)	15	0.052	0.9	0.22

Table 6.13: Significance of electron channel before and after taking into account the uncertainty dB in the total number of background events due to systematics. The result is shown for the two different cuts on L_{bSele}, but assuming the same systematic uncertainties (as computed at the first working point) for both.

Chapter 7

Combined Results

The possibility to observe the Higgs signal in the associated production with a $t\bar{t}$ pair has been explored by the CMS working group, dividing the research in three main final state following the way in which the two W bosons decay. Roughly in 46% of the events both Ws decay fully hadronically, in 29% one W decay hadronically while the other W decay in μ or e lepton and neutrino, finally order 5% contain two oppositely charged leptons (either can be an electron or muon), i.e. the dilepton decay of the W bosons. The remaining 20% or so of events correspond to those cases where one or both of the W bosons decay to a tau lepton and neutrino and are not easily distinguished cause of the rich decay repertoire of the tau meson.

A detailed description of the event selection for each of the main event topologies is given in Ref. [31]. In this chapter we will give a brief summary of the fully hadronic and dileptonic final state analysis methods and then we will show the combined final results including systematic and cross section uncertainties.

7.1 Di-lepton channel: $t\bar{t}H \rightarrow b\bar{b}b\bar{b}l'\mu'_l l\nu_l$

Di-lepton $t\bar{t}H$ events are selected by requiring two reconstructed leptons (e, μ) accompanied by significant missing transverse energy and at least four but no more than seven jets, at least three of which have been b-tagged according to the combined secondary vertex tag algorithm described in chapter 5. Lepton identification is performed using the muon and electron likelihoods described in section 3.1.2 and 3.2.2, respectively. In the semi-leptonic analysis, events with more than one identified lepton are vetoed, but in the di-lepton analysis those events are retained. The likelihood acceptance cuts used for leptons in the di-lepton channel are therefore chosen to be the same

as the second-lepton veto cuts for both semi-leptonic channels. In this way, the sample of events for the di-lepton $t\bar{t}H$ analysis is by construction strictly complementary to those used in the semi-leptonic channels.

The cuts of the di-lepton $t\bar{t}H$ selection are summarised below [31]:

- 2 oppositely charged leptons (e,μ) passing identification criteria ($-Log(L_\mu) <$ 1.4 for muons, $-Log(L_e) < 1.2$ for electrons)
- corrected $E_T^{miss} > 40$ GeV
- 4 to 7 jets with calibrated $E_T > 20$ GeV and $|\eta| < 2.5$, reconstructed with the *iterative cone* algorith with a jet cone of $\Delta R = 0.5$.
- ≥ 3 selected jets b-tagged with discriminator $D > 0.7$

The above is termed the "loose" working point because there is evidence that it is possible to increase the purity (S/B), of the selection by way of the following more stringent "tight" criteria:

- lepton selection as above
- E_T^{miss} requirement as above
- 4 to 6 jets with calibrated $E_T > 20$ GeV and $|\eta| < 2.5$, reconstructed with the *iterative cone* algorithm with a jet cone of $\Delta R = 0.5$
- ≥ 4 selected jets b-tagged with discriminator $D > 0.7$

Although the significance $S/\sqrt{(B)}$ decreases, the cleaner selection is plagued less by systematic uncertainties which dominate the more realistic significance $S/\sqrt{B + dB^2}$. However, the numbers quoted for the "tight" working point are currently insufficiently precise because of limited dataset sizes. It has to be noted that at present, the di-lepton analysis is a counting experiment, thus no effort has been made to assign the missing transverse energy to the two neutrinos from the hard event and thus to reconstruct the invariant masses of the top quarks and Higgs boson.

The selection efficiencies for the two working points, with the corresponding number of expected events and the signal significance, are summarised in Tab. 7.1.

	Analysed ev.	$\epsilon_{loose}(\%)$	N^{ev}_{loose}	$\epsilon_{tight}(\%)$	N^{ev}_{tight}
$t\bar{t}H$ (115)	27900	0.511 ± 0.025	168 ± 8	0.088 ± 0.010	29 ± 3
$t\bar{t}H$ (120)	26141	0.490 ± 0.025	132 ± 7	0.070 ± 0.009	19 ± 3
$t\bar{t}H$ (130)	25911	0.490 ± 0.025	82 ± 4	0.072 ± 0.010	12 ± 2
$t\bar{t}b\bar{b}$	313894	0.637 ± 0.014	1080 ± 24	0.094 ± 0.007	159 ± 12
$t\bar{t}1j$	280385	0.0125 ± 0.0021	1270 ± 220	0	< 42 (68% C.L.)
$t\bar{t}2j$	276917	0.0448 ± 0.0040	2690 ± 240	0.00144 ± 0.00072	87 ± 43
$t\bar{t}3j$	90367	0.0553 ± 0.0078	1330 ± 190	0	< 31 (68% C.L.)
$t\bar{t}4j$	120042	0.0716 ± 0.0077	2620 ± 280	0.0025 ± 0.0014	92 ± 53
$t\bar{t}Z$	110156	0.304 ± 0.017	103 ± 6	0.0363 ± 0.0057	12 ± 2
all backgrounds			9090		< 422
$S/\sqrt{B}$ (115)			1.8		1.4
S/B (115)			1.8 (%)		6.9 (%)
$S/\sqrt{B}$ (120)			1.4		0.9
S/B (120)			1.5 (%)		4.5 (%)
$S/\sqrt{B}$ (130)			0.9		0.6
S/B (130)			0.9 (%)		2.9 (%)

Table 7.1: Selection efficiency ϵ_{loose} (including branching fraction where applicable) and resulting number of expected events N_{loose} in 60 fb^{-1}, for the di-lepton $t\bar{t}H$ channel. For a glimpse of possible improvements, the same for a tighter set of cuts is provided (ϵ_{tight}, N_{tight}). The signal datasets are labelled by the generated Higgs mass in GeV/c^2 (parentheses). Also quoted are binomial errors arising from the finite sizes of processed datasets. All numbers refer to the full mass range.

7.2 All-hadron channel: $t\bar{t}H \rightarrow b\bar{b}b\bar{b}qq'q''q'''$

The selection of the all-hadron $t\bar{t}H$ relies on kinematic and b-tagging jet variables [31]. Moreover, in order to combine the results from the three different decay sub-channels, a veto on leptons has been applied using the complementary cut developed within the semi- and di-leptonic analyses: the event is discarded if $-Log(L_\mu) < 1.4$ or $-Log(L_e) < 1.2$. For this final state two additional backgrounds of QCD events were generated with PYTHIA in the $\hat{p_t}$ ranges from 120 to 170 GeV and greater than 170 GeV, to check the rejection power on samples without top quark events. For these backgrounds a generator level filter was applied to reduce the amount of events to be generated. The filter used is the same as for the $t\bar{t}b\bar{b}$ and $t\bar{t}Z$ backgrounds (Sec. 6.1). The leading order CompHEP cross-sections of these background processes together with the effective cross-sections after the generator filters are listed in Table 7.2.

	QCD $\hat{p_T}$=120-170 GeV	QCD $\hat{p_T}$ >170 GeV
σ_{LO} (pb)	$3.82 \cdot 10^5$	$1.05 \cdot 10^5$
$\sigma_{LO} \times \epsilon$ (pb)	76.4	336.0

Table 7.2: LO CompHEP cross-sections and effective cross-sections after the generator filters of the QCD background processes.

As described in section 4.3, a dedicated $t\bar{t}H$ calibration is used to calibrate jet transverse energy for the all-hadron final state [55]. In this case a different set of calibration parameters are applied to b and *non-b*-jets; b-jets are identify using the combined secondary vertex tag with a discriminator value cut of 0.4. Moreover a detailed study has been done in order to optimise the jet reconstruction cone size. For this purpose a proto analysis is applied to the signal and the three most dangerous background samples ($t\bar{t}2j$, $t\bar{t}b\bar{b}$, qcd with $\hat{p_T} > 170$ GeV) varying the cone radius size between 0.35 and 0.50 in step of 0.5. The top quarks are reconstructed using a χ^2 method applied to the 8 most energetic jets in the tracker acceptance ($|\eta| < 2.7$). For each possible combination of jet-parton assignments for the 8 most energetic jets, the following invariant mass χ^2 is computed:

$$\chi^2_{mass} = \left(\frac{m_{W^+} - m_{jj}}{\sigma(m_W)}\right)^2 + \left(\frac{m_{W^-} - m_{jj}}{\sigma(m_W)}\right)^2 + \left(\frac{m_t - m_{jjj}}{\sigma(m_t)}\right)^2 + \left(\frac{m_{\bar{t}} - m_{jjj}}{\sigma(m_t)}\right)^2 \tag{7.1}$$

The jet-pairing combination with the lowest χ^2_{mass} that satisfies these requirements is chosen and the event is selected. Figure 7.1 plots significance

with respect to the S/B ratio for a range of b-tag discriminator values for each of the several cone sizes indicated. Lower discriminator values yield

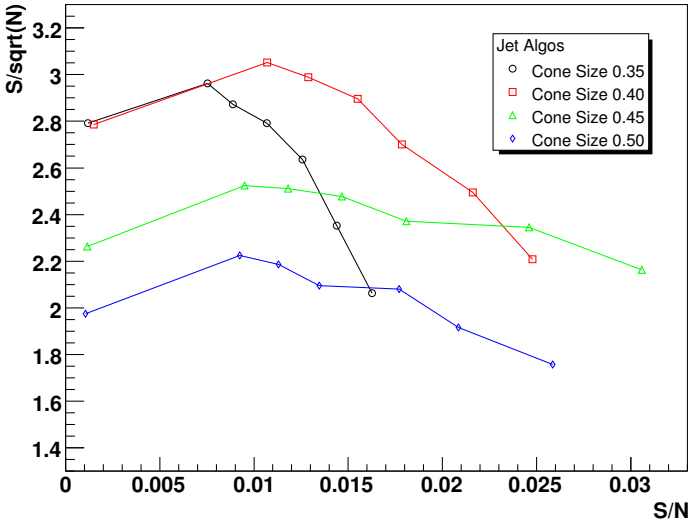


Figure 7.1: Change in significance and S/B resulting from variations in the bTag Discriminator Variable for the various cone sizes indicated in the legend.

higher significance but only at the cost of low S/B while, on the contrary, higher discriminator values give lower significance but higher S/B. A good compromise is in the middle range of each of the curves where neither S/B nor significance are unreasonably low. With this in mind, the best choice for the jet cone is seen to be $\Delta R = 0.40$.

The selection of the all-hadron channel with respect to the resonant backgrounds relies on this set of variables:

- Jet Transverse Energies of the 8 most energetic jets in the tracker acceptance
- Combined b-Tag discriminator value for each jet
- Centrality of the event defined as $\sum_{i=0}^{8} E_T^i / E^i$
- Centrality of the Higgs defined similarly, with the sum restricted to the 2 jets paired to the Higgs

The choice of the jet pairing is performed with the χ^2 method described above. For the 8 jets in the final state the number of combinations to assign

the jet to the top quarks and hadronic Ws is:

$$N = \left(\begin{array}{c} n \\ 6 \end{array} \right) \times 6! \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \tag{7.2}$$

where the terms $\frac{1}{2}$ coming from the possibility to exchange the two jets in both Ws and the two top quarks (both top quarks are decaying hadronically and can therefore not be distinguished in this kind of reconstruction). For 8 jets the number of combination is already huge, 2520. All possible combinations in jet-parton assignments for the 8 selected jets (ordered from the lowest to the highest χ^2) are considered starting from the lowest χ^2. Two working points have been chosen: the first uses loose cuts on b-tagging to get higher statistical significance (but lower S/B), while the second uses a tighter cut on b-tagging to obtain a higher S/B (but lower significance).

For the "loose" working point an event is selected if the following conditions are satisfied:

- $E_T^{7th} > 30$ GeV and $E_T^{8th} > 20$ GeV for the E_T ordered jets
- the χ^2 for each of the 2 W's and 2 top quarks are within 3 sigma of their expected values
- the third highest combined b-Tag discriminator for the 4 jets associated to the b-partons satisfies $D_3 > 0.80$
- Higgs centrality higher than 0.55 and no cut on event centrality

For the "tight" working point an event is selected if the following conditions are satisfied:

- $E_T^{7th} > 30$ GeV and $E_T^{8th} > 20$ GeV for the E_T ordered jets
- the χ^2 for each of the 2 W's and 2 top quarks are within 3 sigma of their expected values
- the 3 highest combined b-Tag discriminator for the 4 jets associated to the b-partons satisfy $D_3 > 0.85$ while the fourth satisfies $D_4 > 0.70$
- Event and Higgs centrality, respectively, in excess of 0.55 and 0.80

The number of analysed events, selection efficiencies with the corresponding number of expected events and the signal significance are reported in Tables 7.3 for the all-hadron decay channel. The number of expected events is computed for an integrated luminosity of 60 fb^{-1}. Both working points are considered.

	Analysed Ev.	$\epsilon_{loose}(\%)$	N^{ev}_{loose} 60fb^{-1}	$\epsilon_{tight}(\%)$	N^{ev}_{tight} 60fb^{-1}
$t\bar{t}H$ (115)	49636	2.32 ± 0.07	347 ± 10	0.294 ± 0.015	44 ± 4
$t\bar{t}H$ (120)	163494	2.55 ± 0.03	314 ± 5	0.366 ± 0.024	45 ± 2
$t\bar{t}H$ (130)	43254	2.80 ± 0.08	214 ± 6	0.358 ± 0.029	27 ± 2
$t\bar{t}b\bar{b}$	203135	0.702 ± 0.019	1190 ± 31	0.0645 ± 0.0056	109 ± 9
$t\bar{t}1j$	1031551	0.0084 ± 0.0009	860 ± 92	0.0005 ± 0.0002	49 ± 22
$t\bar{t}2j$	559111	0.0333 ± 0.0024	2000 ± 150	0.0009 ± 0.0004	54 ± 24
$t\bar{t}3j$	68015	0.079 ± 0.011	1910 ± 260	0.0015 ± 0.0015	35 ± 35
$t\bar{t}4j$	97334	0.182 ± 0.014	6660 ± 500	0.0021 ± 0.0015	75 ± 53
$Zt\bar{t}$	80226	0.358 ± 0.021	121 ± 7	0.0312 ± 0.0062	11 ± 2
qcd170	264310	0.0238 ± 0.0030	4810 ± 610	0.0004 ± 0.0004	76 ± 76
qcd120	55128	0.0018 ± 0.0018	83 ± 83	0 ± 0	$<95(68\%C.L.)$
Total Background			17600		< 505
$S/\sqrt{B}$ (115)			2.6		2.0
S/B (115)			2.0%		8.7%
$S/\sqrt{B}$ (120)			2.4		2.0
S/B (120)			1.8%		8.9%
$S/\sqrt{B}$ (130)			1.6		1.2
S/B (130)			1.2%		5.4%

Table 7.3: Analysed events, selection efficiency, number of expected events and signal significance in 60 fb^{-1} for the all-hadron $t\bar{t}H$ channel for the two working points ϵ_{loose} and ϵ_{tight}. The signal datasets are labelled by the generated Higgs mass in GeV/c^2 (parentheses). Also quoted are binomial errors arising from the finite sizes of processed datasets. All numbers refer to the full mass range.

7.3 Combined significance

Since the event samples for the channels studied are strictly disjoint for the lepton selections and vetoes, the results can be combined by simply adding the individual signal yields (background yields) to obtain a summed S (B). For each of the considered systematics, the resultant error in background yields are added for all four channels (single muon, single electron, dilepton and all-hadronic), since they are by definition fully correlated. The summed errors are then added by quadratures to get a combined systematic uncertainty dB. One then calculates the significance, inclusive of systematic uncertainties in the background yield, according to the formula $S/\sqrt{B+dB^2}$.

The systematics for the all-hadron and dilepton channel are treated in the same ways as described for the semi-leptonic channel in section 6.4. The overall results, including systematics, for the dilepton and all-hadron channels are summarised in Tab 7.4 [31].

di-lepton	S/B	$S/\sqrt{B}$	$S/\sqrt{B+dB^2}$
4-7 jets, 3-4 b-tagged (ϵ_{loose})			
$t\bar{t}H$ (115)	0.018	1.8	0.10
$t\bar{t}H$ (120)	0.015	1.4	0.08
$t\bar{t}H$ (130)	0.009	0.9	0.05
4-6 jets, 4-6 b-tagged (ϵ_{tight})			
$t\bar{t}H$ (115)	0.069	1.4	0.42
$t\bar{t}H$ (120)	0.045	0.9	0.27
$t\bar{t}H$ (130)	0.029	0.6	0.18
hadron	S/B	$S/\sqrt{B}$	$S/\sqrt{B+dB^2}$
Working Point ϵ_{loose}			
$t\bar{t}H$ (115)	0.020	2.6	0.07
$t\bar{t}H$ (120)	0.018	2.4	0.07
$t\bar{t}H$ (130)	0.012	1.6	0.05
Working Point ϵ_{tight}			
$t\bar{t}H$ (115)	0.087	2.0	0.22
$t\bar{t}H$ (120)	0.089	2.0	0.22
$t\bar{t}H$ (130)	0.054	1.2	0.13

Table 7.4: Significance before and after taking into account the uncertainty dB in the total number of background events due to systematics for the dilepton and all-hadron channel.

It is of interest to see how much better the results have the potential to be at tighter working points for the various analyses. Since the systematic uncertainties are not well quantified at these "tight" working points, because

of a lack in Monte Carlo Statistics, the same uncertainties as for the "loose" working points are used to reduce spurious statistical effects. Since these uncertainties can only be evaluated approximately, the results for the "tight" cuts should be understood to be indicative of what could be obtained with a possibly re-optimised analysis. This exercise indicates that further studies using a larger amount of Monte Carlo statistics might be able to achieve a higher sensitivity in this channel. Even so, the overall expectation for discovery potential for an integrated luminosity of 60 fb^{-1} is not qualitatively improved.

Figure 7.2 shows the full range of obtainable significances, with the dot marking the currently estimated value with no cross-section uncertainty as described for the semi-leptonic channel in section 6.4. The star corresponds to what one would obtain for 1% and 4% uncertainties on the $\text{t}\bar{\text{t}}\text{Nj}$ and $\text{t}\bar{\text{t}}\text{b}\bar{\text{b}}$ backgrounds, respectively, an arbitrarily chosen reference. It is interesting to note that it does not quite yield a substantial significance, even though background uncertainties of 1% and 4% for $\text{t}\bar{\text{t}}\text{Nj}$ and $\text{t}\bar{\text{t}}\text{b}\bar{\text{b}}$ are probably substantially better than what will be accessible in reality. Moreover this highlights the challenge that is faced in observing $\text{t}\bar{\text{t}}\text{H}$ and that the analyses have to be pushed to a "tight" working point to have less modification in results when systematics are included.

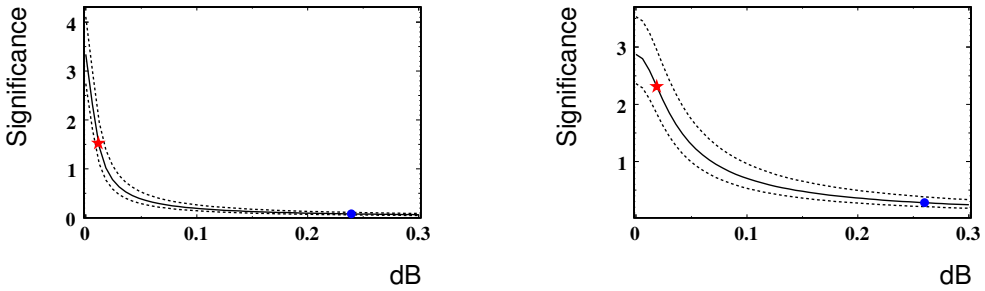


Figure 7.2: Expected range of combined significance (di-lepton + semi-leptonic + all-hadron) versus the total systematic uncertainty in background as a fraction of total background. Left: Results for the "loose" working points. Right: Results for the "tight" working points.

Even with the above prescriptions, however, there will remain uncertainties in the cross-sections depending on how well the procedures can be carried out. It is difficult to predict at this time exactly what will be the level to which the backgrounds can be understood, because the tools required are not yet in existence and because this understanding requires real data. In view

of this, it is interesting to consider how the combined significance of the measurements presented in this note would vary as a function of the fractional uncertainty in background cross-sections, i.e. dB_{xsec}/B. In principle these might have been included in the total systematic uncertainty dB_{sys}/B presented in Tab. 6.12, Tab. 6.13 and Tab. 7.4. However, since the uncertainties in these cross sections at the time of LHC operation are difficult to predict now, it was deemed to be more useful to factor them out and present them separately as described below. The solid central line in Figure 7.3 shows how the combined significance $S/\sqrt{B + (dB_{sys} + dB_{xsec})^2}$ degrades as a function of dB_{xsec}/B. The signal and background yields for the "tight" working points are used in the right side of Figure 7.3, because these give the best results after inclusion of systematics.

Other than this "fundamental" cross-section uncertainty, there are also the "correctible" errors in the cross-sections used at the time of writing, which can be compensated for once data has been collected. The upper and lower dashed curves in Figure 7.3 show the maximum and minimum allowed excursions, should the signal and background cross-sections be off by 10% and 20% respectively. Thus the upper (lower) dashed line corresponds to the signal cross-section scaled up (down) by 10% while at the same time the background cross-section is scaled down (up) by 20%.

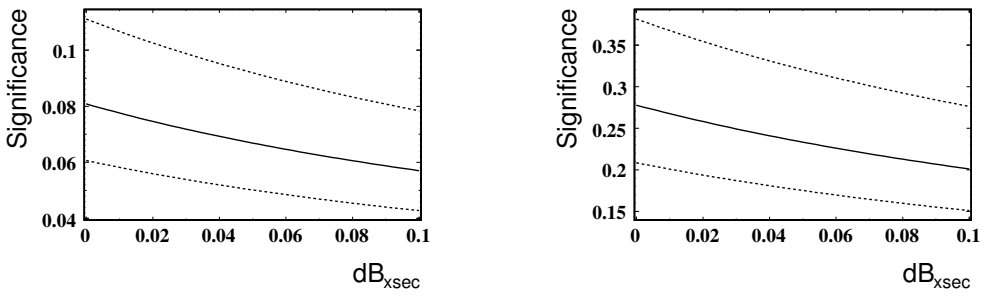


Figure 7.3: Expected range of combined significance (di-lepton + semi-leptonic + all-hadron, and includes the systematic uncertainties estimated in Section 6.4) versus an additional systematic uncertainty on the background cross-section as a fraction of total background. Left: Results for the "loose" working points. Right: Results for the "tight" working points.

The general pessimistic picture of the discovery potential for the $t\bar{t}H$ channel, when systematics are included, is also due the impossibility to extract a narrow Higgs mass peak above a large background. This would help in select regions of invariant mass (side band method) where the background

could be estimated from data, making the $t\bar{t}H$ analysis not only a simple counting experiment. A better jet energy resolution, that will improve the invariant mass resolution too, could be obtained measuring the charged energy component in jets using the tracker detector instead of the calorimeter. Detailed study are already in place in the CMS collaboration but they are not still available at the time of writing.

Conclusions

A full simulation study of all decay channels of the $t\bar{t}H$ with $H \rightarrow b\bar{b}$ channel has been performed and the discovery potential after an integrated luminosity of 60 fb^{-1} has been evaluated.

A likelihood method has been developed to identify leptons coming from W decays (*signal* leptons) with respect to the other possible leptons coming mainly from b, c, K and π decays (*background* leptons). This method finds in 90% of the cases the *signal* lepton with a fake efficiency of only about 1% for the muon case and 3.7% for electron case. Three jet clustering algorithms (the *iterative Cone* algorithm, the inclusive k_T algorithm and the *MidPoint Cone* algorithm) are compared in terms of to some pre-defined performance criteria based on reconstruction efficiency of the kinematics of the primary quarks. A scan of the most important algorithm parameters is performed in order to determine the optimal jet clustering from an analysis point of view, i.e. to maximise the reconstruction efficiency. A secondary vertex b-tagging algorithm has been developed to optimise the selection of b jets; several topological and kinematical secondary vertex related variables as well as information from track impact parameters are combined into a single tagging variable to discriminate between jets originating from b-quark and those from other sources. A likelihood method has been developed to determine, out of six or seven jets in an event, those maximising the probability of assignment to a top quark or Higgs decay. After the jet assignment a tighter selection, cutting on the b-tagging variable, is applied to maximise the discovery potential for a period of data taking corresponding to an integrated luminosity of 60 fb^{-1}.

In contrast to earlier studies, the global picture for this analysis is substantially more pessimistic. This change is due to mainly to the greater degree of realism that was made possible for this study by use of more advanced tools for event generation, detector simulation and physics reconstruction, which were not available to previous (fast-) simulation studies. Mistagging of light flavour jets in $t\bar{t}$Nj events, for example, proved to be a substantially more serious problem than had been foreseen in studies that made use of parametrised

b-tagging. This effect is particularly difficult to estimate without a full detector simulation based upon a relatively detailed material description of the apparatus, coupled with an equally detailed track reconstruction program.

On the other hand, it has been noted that the availability of large control samples of top events will enable b-tagging of high energy jets to be very well understood in CMS. This will probably enable some further suppression of light quark and charm jet tagging relative to b-tagging. Similarly, experience with real data will most likely improve jet reconstruction and energy measurements. This, in turn, will enhance the efficiency of many of the techniques described in this thesis.

Appendix A

Kinematic fit

A general overview of the kinematic fit with non-linear constraints is given in the following. When a constraint f depends in a non-linear way on the parameters which are fitted, the χ^2 minimisation can only be solved iteratively. Let us assume we have to solve a problem with n measured parameters $\vec{y}$, p unmeasured parameters $\vec{a}$ and m constraints $\vec{f}$ as defined in eq. A.1. These requirements will be fulfilled for the true parameters $\bar{a}$ and $\bar{y}$.

$$\begin{aligned} f_1(\bar{a_1}, \bar{a_2}, ..., \bar{a_p}, \bar{y_1}, \bar{y_2}, ..., \bar{y_n}) &= 0 \\ f_2(\bar{a_1}, \bar{a_2}, ..., \bar{a_p}, \bar{y_1}, \bar{y_2}, ..., \bar{y_n}) &= 0 \\ &\vdots \\ f_m(\bar{a_1}, \bar{a_2}, ..., \bar{a_p}, \bar{y_1}, \bar{y_2}, ..., \bar{y_n}) &= 0 \end{aligned} \tag{A.1}$$

In general the measured values $\vec{y}$ will not solve the constraints, so that one has to calculate corrections $\Delta\vec{y}$, where the sum $\vec{y'} = \vec{y} + \Delta\vec{y}$ fulfils the constraints. In the same step, the weighted sum

$$S(\vec{y}) = \Delta\vec{y^T} V^{-1} \Delta\vec{y} \tag{A.2}$$

should be minimal with V being the covariance matrix of the measured parameters. A general method to determine local extrema of non-linear functions of many variables is the definition of *Lagrange Multipliers*:

$$L(\vec{y}, \vec{a}, \vec{\lambda}) = S(\vec{y}) + 2 \sum_{k=1}^{m} \lambda_k f_k(\vec{y}, \vec{a}) \tag{A.3}$$

where $\vec{\lambda}$ are the Lagrange Multipliers. With linear constraints the solution can be found in one step, otherwise it has to be found iteratively.

The linearised constraints are given in the following equation:

$$f_k(\vec{y'},\vec{a'}) \approx f(\vec{y^*},\vec{a^*}) + \sum_{j=1}^{p} \frac{\partial f_k}{\partial a_j} \cdot (\Delta a_j - \Delta a_j^*) + \sum_{i=1}^{n} \frac{\partial f_k}{\partial y_i} \cdot (\Delta y_i - \Delta y_i^*) \quad \text{(A.4)}$$

where $\vec{y}/\vec{a}$ are the start values of measured/ unmeasured parameters, $\vec{y^*}/\vec{a^*}$ are the values of the measured/ unmeasured parameters after last iteration, $\vec{y'}/\vec{a'}$ are the values of meas./unmeas. parameters after current iteration, $\Delta\vec{y^*}/\Delta\vec{a^*}$ is $\vec{y^*} - \vec{y}/\vec{a^*} - \vec{a}$ and $\Delta\vec{y}/\Delta\vec{a}$ is $\vec{y'} - \vec{y}/\vec{a'} - \vec{a}$. The linearization is repeated until defined convergence criteria are fulfilled. These should guarantee firstly that the χ^2-expression S can only change by a given value ϵ_S from one iteration to the next and secondly that the constraints are fulfilled better than a given value ϵ_F.

$$\frac{S(n-1) - S(n)}{ndf} < \epsilon_S \quad \text{and} \quad \sum_{k=1}^{m} f_k^{(n)}(\vec{y},\vec{a}) < \epsilon_F \quad \text{(A.5)}$$

where ndf is just the difference between the constraints and the amount of unmeasured quantities, $ndf = m - p$. If the convergence criteria are not fulfilled the procedure is repeated with the corrected parameters of previous iteration. A fixed limit is set on the total amount of times that the convergence criteria can be applied in every fit.

In the program several parametrisations are currently implemented.

Cartesian four-vector parametrisation: the fitted momentum vector is described by three parameters x, y and z with

$$\vec{p_f} = x\vec{u_x} + y\vec{u_y} + z\vec{u_z} \quad \text{(A.6)}$$

Spherical four-vector parametrisation: the class implements vector parameterisations in spherical coordinates with three parameters r, θ and ϕ parameterising the momentum vector:

$$\vec{p_f} = (r\cos\phi\sin\theta, r\sin\phi\sin\theta, r\cos\theta) \quad \text{(A.7)}$$

Momentum deviation four-vector parametrisation: the class implements a vector parametrisation describing the deviation from the measured momentum in a local coordinate system

$$\vec{u_1} = \frac{\vec{p_m}}{|\vec{p_m}|} \quad , \quad \vec{u_2} = \frac{\vec{u_3} \times \vec{u_1}}{|\vec{u_3} \times \vec{u_1}|} \quad , \quad \vec{u_3} = \frac{\vec{u_z} \times \vec{u_1}}{|\vec{u_z} \times \vec{u_1}|} \quad \text{(A.8)}$$

The fitted momentum is parametrised with three parameters a, b and c:

$$\vec{p_f} = a|\vec{p_m}|\vec{u_1} + b\vec{u_2} + c\vec{u_3} \quad \text{(A.9)}$$

In all cases presented above the fourth component of the four-vector can be parametrised in three ways; fixed mass m ($E_f = \sqrt{|\vec{p_f}|^2 + m^2}$), free floating mass ($E_f = \sqrt{|\vec{p_f}|^2 + d^2 m^2}$) and free floating energy ($E_f = dE_m$).

A detailed description of the kinematic fit with mass constraints and some applications can be found in Ref. [69]. Here we report the case of the $t\bar{t}H$ channel.

Kinematic fit with mass constraints in $t\bar{t}H$ events

In this section the kinematic fit with mass constraint has been compared with a χ^2 method. The comparison has been done using a simpler method to assign the jet to the W and to the tops with respect to the likelihood method described in section 6.3.2.

In this case, the W mass is reconstructed with a jet combination of the two non-b tagged jets and the event is accepted if the W invariant mass is within two standard deviation from the expected value shown in Fig. 6.5. The b-tagged jets, selected with a b-tagging probability of about 45%, are used to reconstruct the leptonic and hadronic top quarks. If the χ^2 method is used, for each possible combination of jet-parton assignments, the following invariant mass χ^2 is computed:

$$\left(\frac{m_t - m_{j\mu\nu_\mu}}{\sigma(m_t)}\right)^2 + \left(\frac{m_t - m_{jjj}}{\sigma(m_t)}\right)^2 . \tag{A.10}$$

If the kinematic fit is used, three mass constraints are imposed in the kinematic fit, the hadronic W boson mass and the leptonic and hadronic top quark. The two b-jets which provide the largest value of the kinematic fit probability are taken to reconstruct the two top quarks and the remaining b-jet pair is assigned to the Higgs boson.

The four-vector parametrisation of the jets is in spherical coordinates with three parameters r, θ and ϕ and a free floating energy $E_f = dE_m$ parametrised with an additional parameter d. For the muon and the neutrino a cartesian four-vector with a fixed mass m parametrisation is used. The covariance matrices are assumed to be diagonal. The jet resolutions are parametrised as a function of the reconstructed jet energy, while fixed resolutions are used for the muon and neutrino momenta.

The improvement obtained using the kinematic fit with mass constraints is checked by computing how many times both methods find the right pair of b-jets coming from the two b-quarks of the Higgs boson decay. The invariant mass distributions for all combinations (white) and wrong combinations (grey), for the b-jet pair Higgs candidates found with the χ^2 method and

with the kinematic fit, are shown in Fig. A.1. A combination is considered as wrong if only one or no one of the b-jets are matched with the b-quarks from the Higgs decay. A jets is considered matched to a quark if the distance in the η, ϕ metric ($\Delta R_{jet-par} = \sqrt{(\eta_{jet} - \eta_{par})^2 + (\phi_{jet} - \phi_{par})^2}$) is less than 0.5.

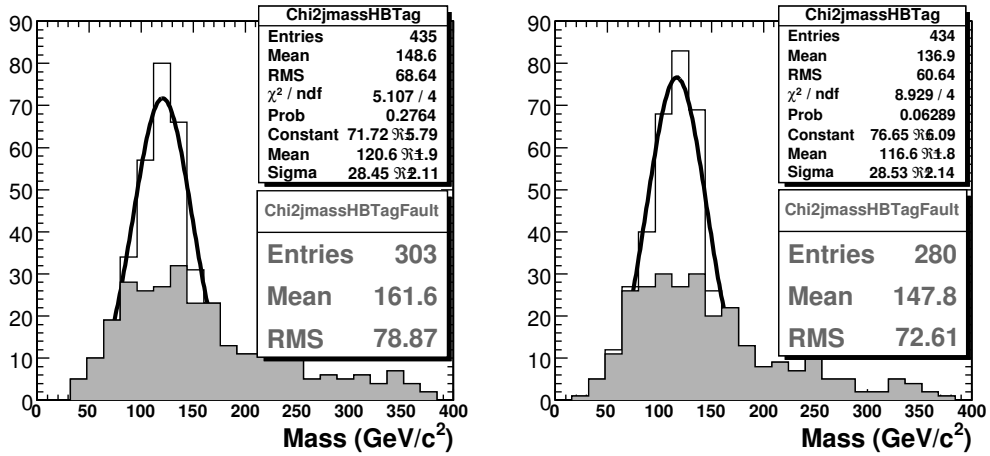


Figure A.1: Invariant mass distributions for all combination (white) and wrong (grey) combination, for the b-jet pair Higgs candidates found with χ^2 method (left) and the kinematic fit method (right).

The pairing efficiency with kinematic fit is 35.5% to be compared to 30% using the χ^2 method: this corresponds to an increase of 5.5%. The mass constraint on the hadronically decaying W improves the probability to correctly identify the b-jets coming from the top quark decay and hence also the probability of correctly finding the b-jets originating from the Higgs boson decay. It has to be noticed that no further constraints can be applied to improve the measurement of the jets coming from the Higgs decay, since the presence of a neutrino does not allow to use a constraint on the transverse energy of the events.

Bibliography

[1] M. Gell-Mann, Phys. Lett. **8** (1964) 214

[2] C. N. Yang and R. L. Mills, Phys. Rev. **96** (1954) 191.

[3] S. Glashow, Nucl. Phs. **22** (1961) 579.

[4] S. Weinberg, Phys. Rev. Lett. **19** (1967) 1264.

[5] A. Salam, *Elementary Particle Theory*, Ed.N.Svartholm (1968)

[6] P.W. Higgs, Phys. Lett. **12** (1964) 132

[7] The Gargamelle Neutrino Collaboration, F.J. HAsert *et al.*, Phys. Lett. **B46** (1973).

[8] The UA1 Collaboration, G. Arnison *et al.*, Phys. Lett. **B122** (1983).

[9] F. Mandl, G. Shaw, *Quantum Field Theory*, John Wiley & Sons (1984)

[10] N. Cabibbo, L. Maiani, G. Parisi, R. Petronzio, Nucl. Phys. **B158** (1979) 295

[11] M. Lindner, Z. Phys. **C31** (1986) 295

[12] M. Sher, Phys. Rep. **179** (1989) 273

[13] G. Altarelli, G. Isidori, Phys. Lett. **B337** 141

[14] W. Hollik, G. Duckeck, *Electroweak Precision Tests at LEP* (2000) Springer Tracts in Modern

[15] F. M. Renard, *Basics of Electron Positron Collisions* (1981) Editions Frontires, Gif sur Yvette

[16] M. Veltman, Acta Phys. Polon. **B8** (1977) 475

[17] H. Spiesberger, M. Spira and P. M. Zerwas, *The standard model: Physical basis and scattering experiments*, arXiv:hep-ph/0011255.

[18] E. Tournefier, *Electroweak results from LEP and SLC and tests of the standard model*, arXiv:hep-ex/0105091.

[19] P. B. Renton, *Precision electroweak tests of the standard model*, Rept. Prog. Phys. **65** (2002) 1271

[20] D. Schaile, *proceedings of the XXVII'th ICHEP*, Glashow, Scotland, July 1994

[21] E. Brubaker *et al.* [Tevatron Electroweak Working Group], *Combination of CDF and D0 results on the mass of the top quark*, arXiv:hep-ex/0608032.

[22] R. Barate *et al.* [LEP Working Group for Higgs boson searches], *Search for the standard model Higgs boson at LEP*, Phys. Lett. B **565** (2003) 61 [arXiv:hep-ex/0306033].

[23] LHC, *The Large Hadron Collider Conceptual Design*, CERN/AC/**95-05** (1995).

[24] CMS collaboration *CMS Physics* CERN/LHCC **06-021** (2006)

[25] G. Davatz, M. Dittmar and A. S. Giolo-Nicollerat, *Standard model Higgs discovery potential of CMS in the $H \to WW \to l\nu l\nu$ channel*, CERN-CMS-NOTE-**2006-047**

[26] S. Baffioni *et al.*, *Discovery potential for the SM Higgs boson in the $H \to ZZ(*) \to e^+e^-e^+e^-$ decay channel*, CERN-CMS-NOTE-**2006-115**

[27] S. Abdullin *et al.* [CMS Collaboration], *Search strategy for the standard model Higgs boson in the $H \to ZZ(*) \to 4\mu$ decay channel using M(4mu)-dependent cuts*, CERN-CMS-NOTE-**2006-122**

[28] M. Pieri, S. Bhattacharya, I. Fisk, J. Letts, V. Litvin and J. G. Branson, *Inclusive search for the Higgs boson in the $H \to \gamma\gamma$ channel*, CERN-CMS-NOTE-**2006-112**

[29] V. Drollinger, T. Muller and D. Denegri, *Searching for Higgs bosons in association with top quark pairs in the $H \to b\bar{b}$decay mode*, arXiv:hep-ph/0111312.

[30] J. Cammin, M. Schumacher, *The ATLAS discovery potential for the channel $t\bar{t}H$, $H \to b\bar{b}$*, ALT-PHYS-**2003-024**

[31] D. Benedetti *et al.*, *Search for $H \to b\bar{b}$ in association with a $t\bar{t}$ pair at CMS*, CERN-CMS-NOTE-**2006-119**

[32] CMS collaboration: *Technical Proposal*, CERN/LHCC **94-38** LHCC/P1 (1994)

[33] M. Lamont, *Estimates of Annual Proton Doses in the LHC*, LHC Project Note **375**

[34] R. Bailey and P. Collier, *Overall Strategy for Early Luminosity Operation with Protons* CERN EDMS Document **LHC-OP-BCP-0001** (2004)

[35] CMS collaboration *The Magnet Project* TDR CERN/LHCC **97-10** (1997)

[36] CMS collaboration *The Tracker Project*, TDR CERN/LHCC **98-06** (1998)

[37] L. Borrello, A. Messineo, E. Focardi and A. Macchiolo, *Sensor design for the CMS Silicon Strip Tracker*, CERN-CMS-NOTE-**2003-20**

[38] CMS collaboration *Detector Perfomance and Software* TDR CERN/LHCC **06-001** (2006)

[39] CMS collaboration *The Electromagnetic Calorimeter* TDR CERN/LHCC **97-33** (1997)

[40] CMS collaboration *The Hadron Calorimeter* TDR CERN/LHCC **97-31** (1997)

[41] CMS collaboration *The Muon Project* TDR CERN/LHCC **97-32** (1997)

[42] CMS collaboration *The Level-1 Trigger*, TDR CERN/LHCC **00-38** (2000)

[43] CMS collaboration *Data Acquisition & High-Level Trigger*, TDR CERN/LHCC **02-26** (2002)

[44] E. James, Y. Maravin, M. Mulders, N. Neumeister *Muon identification in CMS*, CERN-CMS-NOTE-**2006-010**

[45] S. Baffioni *et al. Electron reconstruction in CMS* CERN-CMS-NOTE-**2006-040**

[46] R. Frühwirth, *Application of Kalman Filtering to Track and Vertex Fitting*, Nucl.Instrumand Methods **A262** (1987) 444

[47] L. Agostino, M. Pieri, *HLT Selection for Electrons and Photons*, CERN-CMS-NOTE-**2006-078**

[48] T. Sjostrand, L. Lonnblad and S. Mrenna, PYTHIA 6.2: Physics and manual, arXiv:hep-ph/0108264.

[49] M.L. Mangano, M. Moretti and R. Pittau, Nucl.Phys.**B632** (2002) 343, hep-ph/0108069

[50] M.L. Mangano, M. Moretti, F. Piccini, R. Pittau and A.D. Polosa, JHEP **07** (2002)001, hep-ph/0206293

[51] C. Buttar *et al.*, *Les Houches physics at TeV colliders 2005, standard model, QCD, EW, and Higgs working group: Summary report*, arXiv:hep-ph/0604120.

[52] G. C. Blazey *et al.*, *Run II jet physics*, arXiv:hep-ex/0005012.

[53] A. Pukhov *et al.*, *CompHEP: A package for evaluation of Feynman diagrams and integration over multi-particle phase space. User's manual for version 33*, arXiv:hep-ph/9908288.

[54] A. Heister, O. Kodolova, V. Konoplyanikov, S. Petrushanko, J. Rohlf, C. Tully and A. Ulyanov, *Measurement of jets with the CMS detector at the LHC*, CERN-CMS-NOTE-**2006-036**

[55] A. Santocchia, *Optimization of jet reconstruction settings and parton-level corrections for the t anti-t H channel*, CERN-CMS-NOTE-**2006-059**

[56] H. F. Pi, P. Avery, D. Green, J. Rohlf and C. Tully, *Measurement of missing transverse energy with the CMS detector at the LHC*, CERN-CMS-NOTE-**2006-035**

[57] A. Nikitenko, S. Kunori and R. Kinnunen, *Missing Tranverse Energy Measurement with Jet Energy Corrections* CERN-CMS-NOTE-**2001/040**

[58] C. Weiser, *A combined secondary vertex based B-tagging algorithm in CMS*, CERN-CMS-NOTE-**2006-014**

[59] T. Speer, K. Prokofiev, R. Fruhwirth, W. Waltenberger and P. Vanlaer, *Vertex fitting in the CMS tracker*, CERN-CMS-NOTE-**2006-032**

[60] The Cmsjet Home Page - http://cmsdoc.cern.ch/ abdullin/cmsjet.html

[61] S. Abdullin, A. Khanov, N. Stepanov, *CMSJET* CMS TN/94-180 (2002)

[62] V. Drollinger, *Upgrade of Fast Tracker Response Simulation, the FAT-SIM utility*, CMS IN **2000/034**

[63] ORCA reference guide, available at: http://cmsdoc.cern.ch/orca/

[64] The CMSIM Home Page - http://cmsdoc.cern.ch/cmsim/cmsim.html

[65] P. Demin, S. De Visscher, A. Bocci and R. Ranieri, *Tagging b jets with electrons and muons at CMS*, CERN-CMS-NOTE-**2006-043**

[66] The OSCAR Home Page - http://cmsdoc.cern.ch/OSCAR/

[67] V. Drollinger, *Reconstruction and Analysis Methods for Searches of Higgs Bosons in the Decay Mode $H^0 \to b\bar{b}$ at Hadron Colliders*, Phd Thesis (2001)

[68] W. Beenakker, S. Dittmaier, M. Kramer, B. Plumper, M. Spira and P. M. Zerwas, *NLO QCD corrections to t anti-t H production in hadron collisions*, Nucl. Phys. B **653** (2003) 151.

[69] J. D'Hondt *et al.*, *Fitting of event topologies with external kinematic constraints in CMS*, CERN-CMS-NOTE-**2006-023**

[70] J. Rohlf, C. Tully, *Recommendations for Jet and Missing Transverse Energy Reconstruction Settings and Systematics Treatment*, CMS-IN-**2006-025**

[71] S. Lowette, J. D'Hondt, J. Heyninck and P. Vanlaer, *Offline calibration of b-jet identification efficiencies*, CERN-CMS-NOTE-**2006-013**