UNIVERSITÉ DE STRASBOURG

THÈSE

pour l'obtention du diplôme de

DOCTEUR DE L'UNIVERSITÉ DE STRASBOURG

spécialité : physique des particules

présentée par

Cristina Ferro

Mesure de la section efficace de production de paires de quarks top dans le canal tau+jets dans l'expérience CMS auprès du LHC

présentée le 14 mai 2012 devant la commission d'examen composée de

M. Daniel Bloch: directeur de thèse
M. Daniel Froidevaux: rapporteur
M. Eric Kajfasz: rapporteur
Mme Anne-Catherine Le Bihan: co-encadrante
Mme Christelle Roy: examinatrice
M. Roberto Tenchini: examinateur

Grazie!

Ebbene si...sono già passati tre anni. Guardando indietro mi sembra ieri che sono partita, con poca voglia e tanta nostalgia per amici, fidnazato e familiri che lasciavo a casa... Beh! devo dire che dinuovo la vita mi ha sorpeso in positivo! Davvero non immaginavo che nella fredda Strasburgo, avrei passato 3 anni davvero splendidi...da tutti i punti di vista. Per tutto questo le persone che devo ringraziare sono tantissime. Provo a fare una sorta di lista che sono sicura non sarà esauriente! Per cominciare un grazie di cuore va a chi, prima mi ha scelto lavorativamente, poi accolto in laobratorio, e per finire, mi ha seguito con dedizione e pazienza infinita durante tutti questi 3 anni: Daniel Bloch. Un grazie davvero speciale per tutto quello che mi hai insegnato! Un semplice grazie sicuramente non basta per la "mia" splendida Cate! Grazie per tutti i momenti belli e brutti passati insieme, "gioie e dolori" condivisi per finire in tempo quella che possiamo definire l'analisi dell'anno...nessuno prima d'ora aveva mai calcolato una "pull" con tanta precisione e dedizione, per non parlare del fondamentale sistematico sulla PDF!!! Ma a parte il lavoro devo ringraziare Cate per l'amiciza nata e cresciuta in questi anni che ha valso ben più di qualsiasi risultato. Un grazie infinito anche a chi questa tesi l'ha letta, corretta e valutata: Daniel Froidevaux, Eric Kajfasz, Christelle Roy e Roberto Tenchini.

Ancora grazie a tutti i colleghi/amici di Strasburgo: Jeremy , Eric Chabert, Eric Conte, Pierre, Caroline, Denis, Marco, Isabelle, Jerome B., David, Florian, Marie, Antonin e per finire Cecile e Anselmo. Un vero e proprio monumento andrebbe fatto all'assistenza tecnica costante di Yannick e Jerome, nonche' all' impeccabile organizzazione di Leila! Grazie davvero a tutti per il supporto morale e materiale che mi avete dato, e pure per i tanti caffe' relax presi insieme!

Un grazie speciale lo devo anche ai tanti colleghi che dal CERN mi hanno aiutato! Primi tra tutti i miei conveener: Roberto Tenchini e Roberto Chierici per il costante supporto per l'analisi, e Simone Gennai per il continuo e costruttivo confronto su tirgger e tau. Grazie davvero per il vostro competente sostegno. Non certo meno importanti i tanti amici "cernioti": Giulio, Salvatore, Andrea B., Matteo, Andrea L., Stefano, Ciccio, Riccardo, Andrea Lucaroni, Sandrillo, Giacomino e ultime ma

in realtà prime in classifica Maddy e Robe!!

A questa lista già di per sè lunga...devo ancora aggiungere tutti gli amici fiorentini, che nonstante la distanza mi sono stati sempre super vicini. È proprio vero che la distanza fa capire quali sono i rapporti che valgono lasciando indietro quelli che di valore ne avevano ben poco! Allora continuiamo con la lista: Matteo, Francesco R., Francesco Da Vela, Eli, Ceci, Marta, Lore G., Lore M., Luchino e ultimi ma in realtà primi: Lore T. e Paolino! Mi sentirei di estendere la lista di amici alle sorelle "acquisite": Laura, Bea, Michy e la mia miticissima Dam! Grazie di esserci e di esserci sempre state!

Direi che la lista debba continuare con chi di questi anni ha costituito la colonna portante: Antonio. Beh...in questo caso la parola grazie davvero non rende l'idea.

Per finire, anche se senza di loro nulla avrebbe avuto inizio, un grazie lungo 29 anni, va ai miei genitori, alle mie super sisters (Fuddy e Mary) a Maurizio (meglio conosciuto come Cogni), agli splendidi sorrisi di Marcella, Luigi e Roberto, e infine alla dolcissima e in realtà prima su tutti Matilde....da due anni e mezzo ad oggi nulla piu' di lei e del suo sorriso rallegra ogni giornata!

Contents

CONTENTS

Introduction

Between 1960 and 1970, a new theory was born in order to explain the interactions between the elementary particles. It is the so called Standard Model (SM) theory, well supported by many experimental observations like the discovery of the W and Z bosons in 1983 at CERN with the S$p\bar{p}$S collider, and the observation of the top quark in 1995 at the Tevatron ($p\bar{p}$) collider. Even though the SM theory found many experimental confirmations, one particle is still not observed (at the date of the thesis defense): the Higgs boson, the particle that according to the SM theory is responsible of the mass of all other SM particles. The discovery of the Higgs boson is one of the main goals pursued at the LHC (Large Hadron Collider).

On the other hand, we already know that, even if the SM Higgs boson exists, the SM theory is only a limit, of a more general still unknown theory, valid at the energy accessible until now. Furthermore the SM theory is not able to provide an answer to relevant results stemming mainly from astrophysics and cosmological observations. For example in the SM there is no particle candidate to explain the so called *dark matter*. For these reasons several theories alternative to the SM were proposed. At the LHC many studies to test these alternative theories, like Supersymmetry (SUSY), are ongoing.

After two decades of development and construction, the LHC started its operation in September 2009. It is the most powerful proton-proton (p-p) collider ever built and is designed to reach a luminosity of 10^{34} $cm^{-2}s^{-1}$. During the 2011 data taking the LHC reached $5 \cdot 10^{33}$ $cm^{-2}s^{-1}$ for an energy of 7 TeV in the p-p center of mass. Along the 27 km LHC circumference are located 4 main experiments: ALICE, ATLAS, LHCb and CMS in which my thesis work is developed. The 2011 data taking was extremely rich for CMS that was able to collect about 5.0 fb^{-1} of data, (figure **1**). In this thesis we will present results obtained by analyzing the largest part of the 2011 dataset.

The first CMS measurement of the top anti-top production cross section in the hadronic tau jets final state will be presented. This work is structured as follows. First an introduction to the SM theory and to the top quark physics is given. Then a summary of the main features of the CMS detector and its trigger system is presented. Hence we shall describe the main features of the design of the trigger that we developed specifically for our analysis. Then we shall see how the reconstruction
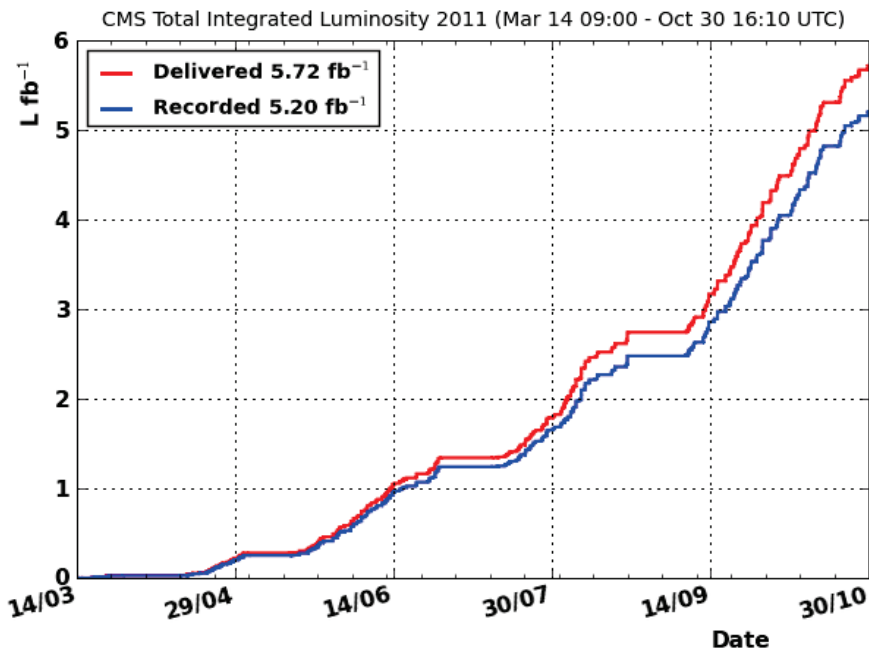
Figure 1: Total integrated luminosity recorded by the CMS experiment during the 2011 data taking.

and the identification of the particles is done in CMS, with a particular attention to the b-jet[1] identification. In the last chapter we shall see how the measure of the top anti-top cross section in the hadronic tau + jets final state is achieved and we shall describe the obtained results. The treatment of the systematic uncertainties will be also presented.

---

[1]We call b-jets the jets stemming from the b quark hadronization.

Chapter 1

The Top quark physics

1.1 The *Standard Model* of the Elementary Particles

1.1.1 Elementary Particles

The Standard Model (SM) is a quantum field theory describing the elementary particles and their interactions. According to the SM our matter is composed of six spin-$\frac{1}{2}$ particles called fermions and divided in two categories: **leptons** and **quarks** with their respective anti-particles. The anti-particles have the same mass than the particles but an opposit electric charge. In the case of neutral anti-particles, they have the same mass but an opposit intrinsic magnetic moment with respect to the corresponding particles.

While leptons are individually observed, quarks exist only in bound states named hadrons. They are divided in two categories: mesons, composed of a quark and an anti-quark, and baryons composed of three quarks.

In the SM, leptons and quarks interact with each other via three fundamental interactions which are mediated by spin-1 particles called bosons. The three forces of the Standard Model and the respective force carriers are:

- **Strong Interaction (8 massless gluons g)**: only between quarks and/or gluons.

- **Weak Interaction ($\mathbf{W^+}$, $\mathbf{W^-}$ and $\mathbf{Z^0}$ massive bosons)**: between leptons and/or $W^\pm$, Z^0 bosons as well as quarks.

- **Electromagnetic Interaction (1 massless photon γ)**: between all electrically charged particles.

The only interaction not described by the SM is the gravity force, which is in any case negligible for typical distances and masses involved in the ordinary space-time particle physics.

A summary of the SM particles and of their classification is given in figure **1.1**. The matter particles are classified in three generations ordered according to their increasing masses.

Figure 1.1: Schematic view of the Standard Model elementary particles.

All described particles have been observed and we don't have any experimental evidence for the existence of a fourth generation of particles.

The first left column of the table **1.1** contains the component of the ordinary matter, while the other particles are mainly produced in cosmic rays and in accelerators. The corresponding anti-particles, instead, are the constituents of the anti-matter. For the current knowledge of the visible universe, we know that its largest part is built of ordinary matter, and today is still unclear the reason of this asymmetry.

1.1.2 The Standard Model Theory

The Standard Model is a Quantum Field Theory based on a $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ local gauge symmetry that introduces the fundamental interactions. For this reason we usually refer to the force carriers as **gauge bosons**.

*A gauge theory is a physics theory based on the idea that the lagrangian has to be globally and locally invariant if we apply a given transformation (symmetry). It means that the system has to be invariant for a given transformation even if the transformation is applied to a specific region of the spacetime. This idea can be seen as a generalization of the **principle of equivalence** introduced by the **general relativity theory**. Thanks to the mathematic framework of the symmetry group, we are able to describe three of the four fundamental interactions (the strong, the weak and electromagnetic forces) by the symmetry $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$, where SU(n) is the group of the unitary matrix n×n, with determinant=1, and U(1) is a transformation of a complex phase.*

The main features of the electromagnetic and weak interactions can be understood using as example the behavior of leptons (1). In particular, for the lepton sector these two interactions are expressed by the following lagrangian, globally invariant under $SU(2)_L \otimes U(1)_Y$ gauge symmetry:

$$\mathcal{L}_{\text{EW}}^{\text{lepton}} = i \sum_{k=1}^{3} \left(\bar{L}_L^k \gamma^\mu \partial_\mu L_L^k + \bar{L}_R^k \gamma^\mu \partial_\mu L_R^k \right). \tag{1.1}$$

In eq.(1.1) the γ^μ with μ=0,1,2,3 are the four Dirac matrices, while $L_{L,R}$ are the left/right handed component of the lepton spinor L.

The $L_{L,R}$ are defined as follows:

$$L_L = \frac{1}{2}(1 - \gamma^5)\binom{\nu_l}{l} \tag{1.2}$$

$$L_R = \frac{1}{2}(1 + \gamma^5)(l) \tag{1.3}$$

where $\frac{1}{2}(1 \pm \gamma^5)$ are the projectors on the chirality states, with $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$. Due to the parity transformation properties of the weak interaction (2), (3), only leptonic states with definite chirality can be involved in the interaction. It can be demonstrated that for zero mass particles the definitions of chirality and helicity concide, this is why only left handed neutrinos can intervene (4). In fact, even if the neutrino mass is not exactly null it is usually much smaller than its momentum, so the equivalence between helicity and chirality can be cosidered as valid with excellent approximation.

In order to switch from global to local gauge invariance we need to move to the so called "covariant derivatives", $(\mathcal{D}_\mu)$, instead of the ordinary derivatives presented in eq.(1.1). Covariant derivatives are built to transform linearly under $SU(2)_L \otimes U(1)_Y$ symmetry and have the following form:

$$\mathcal{D}_\mu = \partial_\mu - \frac{ig}{2}\vec{\tau}\vec{W} - ig'Y B_\mu \tag{1.4}$$

where W_μ is a vector of three gauge fields needed to satisfy the local SU(2) symmetry, while B_μ is the gauge field that guaranties the U(1) local symmetry; g and g$'$ are the coupling constants for the gauge fields; Y is the weak hypercharge and $\vec{\tau}$ are the Pauli matrices. To the four gauge fields of the SU(2)$_L \otimes$U(1)$_Y$ symmetry correspond the four gauge bosons W$^+$, W$^-$ and Z^0 for the weak interaction and γ for the electromagnetic interaction as introduced by S.Glashow, A.Salam and S.Weinberg. Using the eq.(1.4) in eq.(1.1) we see that the lagrangian describes two forces, the weak and electromagnetic forces, which are actually two manifestations of the same fundamental interaction called **Electroweak** interaction. However, the electroweak interaction lagrangian does not contain mass terms for the fermions and for the gauge bosons W$^+$, W$^-$ and Z^0, in obvious contrast with what is experimentally observed; hence we need to include artificially mass terms in the lagrangian, which in principle could break the SU(2)$_L \otimes$U(1)$_Y$ gauge symmetry. This problem can be solved adding one more field, known as the **Higgs fields**.

1.1.3 Higgs field and electroweak interaction mediators

The Higgs field is organized in doublets whose potential is invariant under SU(2)$_L \otimes$U(1)$_Y$ local gauge transformations. The Higgs field lagrangian, in its global gauge invariant form, is written as:

$$\mathcal{L}_{Higgs} = (\partial_\mu \phi^\dagger)(\partial^\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda(\phi^\dagger \phi)^2 \tag{1.5}$$

The lagrangian described in eq.(1.5) is not invariant under SU(2)$_L \otimes$U(1)$_Y$ local transformation because the derivatives do not transform linearly under a transformation depending on a specific point of the spacetime x^μ. In order to achieve the local SU(2)$_L \otimes$U(1)$_Y$ gauge invariance we have to replace the ordinary derivatives with the covariant one (D_μ) described in eq.(1.4). Using the eq.(1.4) in the eq.(1.5) and expanding it around the ground state (writing $\phi(x) = v + h(x)$), the Higgs lagrangian becomes:

$$
\begin{aligned}
\mathcal{L}_{Higgs} = {} & \frac{1}{2}(\partial_\mu h \partial^\mu h + \mu^2 h^2) \\
& -\frac{1}{4} A_{\mu\nu} A^{\mu\nu} \\
& -\frac{1}{4}(W^{+\dagger}_{\mu\nu} W^{\mu\nu+} + W^{-\dagger}_{\mu\nu} W^{\mu\nu-}) + \frac{g^2 v^2}{4}(W^{+\dagger}_\mu W^{\mu+} + W^{-\dagger}_\mu W^{\mu-}) \\
& -\frac{1}{4}(Z_{\mu\nu} Z^{\mu\nu}) + (\frac{g^2 v^2}{4\cos^2\theta_W} Z_\mu Z^\mu) \\
& + \text{interaction-terms}
\end{aligned}
\tag{1.6}
$$

where $A_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, $Z_{\mu\nu} = \partial_\mu Z_\nu - \partial_\nu Z_\mu$ and so on. The first line of eq.(1.6) represents the Higgs boson scalar field, with mass $\sqrt{-2\mu^2}$; the second line represents

a massless field identified with the electromagnetic field; the third line represents $W^\pm$ with a mass of $\frac{g \cdot v}{2}$; finally the fourth line represents the Z field with a mass of $\frac{M_W}{cos\theta_W}$ with $v = \frac{1}{\sqrt{\sqrt{2}G_F}} = 246$ GeV. The Z_μ, $W_\mu^\pm$, B_μ and A_μ fields are connected to each other by the following expressions:

$$B_\mu = \frac{-g' Z_\mu + g A_\mu}{\sqrt{g^2 + g'^2}} \tag{1.7}$$

$$W_\mu^3 = \frac{g Z_\mu + g' A_\mu}{\sqrt{g^2 + g'^2}} \tag{1.8}$$

$$W_\mu^\pm = \frac{W_\mu^2 \mp i W_\mu^2}{\sqrt{2}} \tag{1.9}$$

The Higgs boson is the only particle still not experimentally observed.

1.2 The Strong interaction

Until now we have introduced the electromagnetic and weak interactions. In the same way we can describe the strong interaction which involves only quarks.
Historically the quark's model was some what hampered because no experimental evidence of the existence of single quarks was found. In addition some particles, like the Δ^{++}, were found to be formed by three u quarks apparently identical to each other. Hence this particle seemed to violate the Pauli's exclusion principle. This is why, in 1964, Greenberg introduced the idea of *color charge* (5) . In practice quarks can exist in three different color states, conventionally called green, red, and blue, i.e. they have an additional quantum number named *color*. Besides that, only quark bound states carrying $no - color$ can exist. This effect is knows as *confinement* and justifies the non-observation of free quarks as well as the apparently violation of the Pauli's exclusion principle for the Δ^{++} existence (the three quarks, are not identical but differ from each other by their *color charge*).
From a mathematical point of view, we can describe the strong interaction, and include it in the electroweak lagrangian, extending the local gauge symmetry SU(2)$_L \otimes$U(1)$_Y$ to SU(3)$_C \otimes$SU(2)$_L \otimes$U(1)$_Y$, where SU(3)$_C$ represents the so called *color symmetry*.
  The strong interaction lagrangian is written as:

$$\mathcal{L}^{QCD} = \mathcal{L}_{\text{invar}} + \mathcal{L}_{\text{gauge fix}} + \mathcal{L}_{\text{ghost}} \tag{1.10}$$

where the terms $\mathcal{L}_{\text{gauge fix}}$ and $\mathcal{L}_{\text{ghost}}$ are needed for technical reasons due to the used strategy to normalize the QCD lagrangian (6). The $\mathcal{L}_{\text{invar}}$ term, instead, is invariant under local $SU(3)_c$ transformations and reads:

$$\mathcal{L}_{\text{invar}} = \sum_f \bar{\psi}_f \left(i\gamma_\mu \mathcal{D}^\mu - m_f \right)\psi_f - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \tag{1.11}$$

where f runs over the six quark fields, $\mathcal{D}^{\mu}$ is the covariant derivative:

$$\mathcal{D}_{\mu} = \partial_{\mu} + ig_s A_{\mu a} T_a \tag{1.12}$$

and

$$F_{\mu\nu} = \partial_{\mu} A_{\nu a} - \partial_{\nu} A_{\mu a} - g_s C_{abc} A_{\mu b} A_{\nu c} \tag{1.13}$$

where $A_{\mu a}$ are the fields of the eight colored gluons, T_a are the eight generators of $SU(3)$, C_{abc} are the structure constants that define the commutation rules of the $SU(3)$ generators.

Even if the strong and electroweak interactions are well described by the SM, it is important to focus our attention on a crucial difference between the strong and the electroweak interactions. It is well known, that if we increase the distance between two electric charges, the strength of their interaction quickly decreases. It is in fact, inversly proportional to the square of the distance between the two *electric* charges. On the contrary, the strong interaction exponentially increases with the distance between the *color* charges.

To understand the practical effect of this feature of the strong interaction we can consider the quarks produced at LHC. They are emitted with high energy, so they try to get away with the result that the interaction between each other increases. It means that from an energetic point of view they will arrive to a distance where it will be more convenient to create new pairs of quark and anti-quark than to allow the original quarks to continue to move away. This is why we don't see single quarks but the so called *jets* of grouped hadrons with neutral color. This process is known as *hadronization* and it is still one of the less understood processes of particle physics.

1.3   The Top quark

The discovery of the bottom quark in 1977 suggested the existence of the top quark. The top quark, according with the Standard Model theory is the weak-isospin partner of the bottom quark, characterized by a charge of $+2/3$ and transforms as a color triplet under the SU(3) group of the strong interaction.

Even if the top quark quantum numbers have not been directly measured, we have many indirect evidences supporting these assignments. For example the analysis of EW observables in Z decay requires the existence of a $T_3 = 1/2$, charge -2/3 fermion, with a mass in the range of 170 GeV, consistent with the Tevatron ($p\bar{p}$ collider located in Chicago) measurements. Also the Tevatron top cross section measurement, and its comparison with the theoretical expectation, is consistent with the production of a color-triplet and spin-1/2 particle. Using the 2011 data, CMS set a constraint on the charge of the top quarks excluding a charge of +4/3 (7). The LHC will also provide the first direct measurements of the top quark quantum numbers.

The top quark was finally discovered in 1995 (8) at the Fermilab Tevatron. The present result from Tevatron of the top quark mass is: $M_{top} = (173.2 \pm 0.9) GeV$ (9). The recent measure given by the combination of the ATLAS and CMS results is $M_{top} = (175.4 \pm (1.0)_{stat} \pm (2.2)_{sys}) GeV$ (10). This value of the top mass makes the top quark the heaviest standard model particle with the most precised quark mass.

Because of its high mass and its close value to the electroweak symmetry breaking energy scale, the top quark can be also seen as a key to understand how the particle masses are generated by the Higgs mechanism[1]. The masses of the top quark, Higgs and W bosons are in fact closely related with the higher order corrections of various physics process, and a precise knowledge of the top mass can be used to predict, or at least to constrain the Higgs boson mass. Indeed the radiative corrections of the propagators of the W and Z bosons are proportional to the square of the top quark mass (figure **1.2**: left), and the Higgs boson contributes logarithmically to such corrections (figure **1.2**: right).
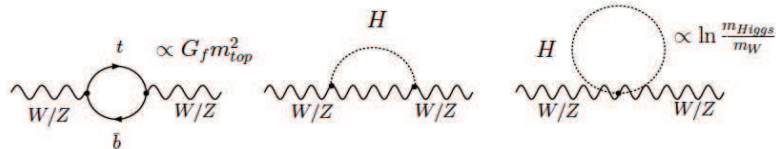
Figure 1.2: Left: contribution of the top quark mass to the radiative correction of the W and Z bosons propagators. Center, right: radiative correction of the Higgs boson to the W and Z bosons propagators

In the figure **1.3** the limits imposed to the Higgs boson mass starting from a precise measure of the W and top quark masses are presented. In this figure the yellow band represents the excluded region in CMS.

1.4 Top quark production at LHC

The top quarks can be produced in pairs ($t\bar{t}$) via strong interaction, or via weak interaction that gives rise to only one top quark, the so called *single top*. At LHC the top pair production has the largest cross section, with about 10^5 $t\bar{t}$ pairs produced per year in *pp* collisions at $\sqrt{s}$=7 TeV for the nominal luminosity of $10^{34}s^{-1}cm^{-2}$.

While at Tevatron the production of the $t\bar{t}$ pairs mainly happened via annihilation of quark-antiquark pairs, at LHC the top pair production is due to gluon fusion

---

[1]The Yukawa coupling between the Higgs boson and the top quark is really large and close to 1.
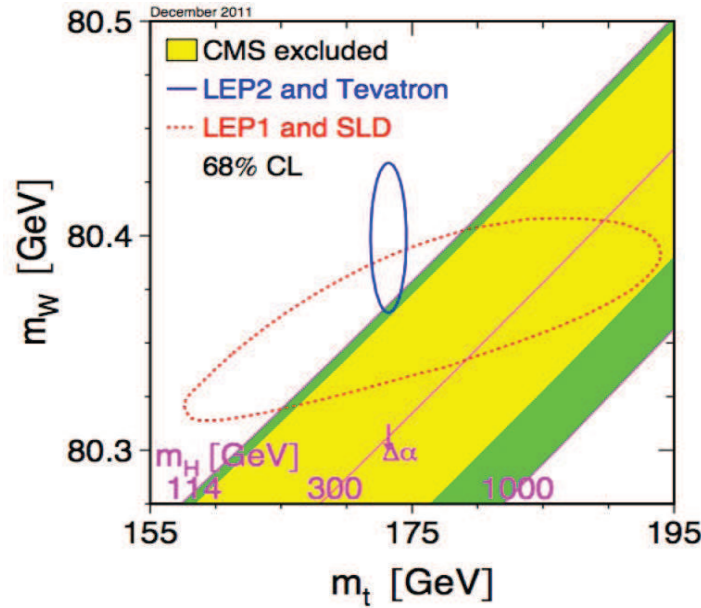
Figure 1.3: The figure shows the relation between the top and W boson masses and the mass of a SM Higgs boson. The top and W masses are represented on the x and y axis respectively. The blue (plain) ellipse shows the region delimited by the W and top masses directly measured at LEP2 and Tevatron. In the red (dashed) ellipse shows the indirect measurement from LEP1 and from SLD experiment. The yellow band is the region excluded for the Higgs mass by the CMS experiment ($127.5 < M_H < 600$ GeV at 95% C.L), while the green band represents the region not yet excluded by the CMS collaboration (11).

in 85% of the cases, and only 15% of the pairs are produced via quark annihilation processes. In figure **1.4** the Feynman diagrams relative to the quark-antiquark annihilation as well as to the gluon fusion processes are presented.
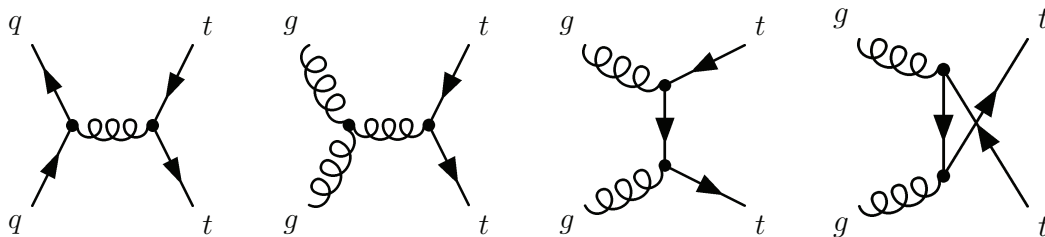
Figure 1.4: Feynman diagrams of the leading order processes for the top pair production at LHC. Quark-antiquark annihilation ($q\bar{q} \rightarrow t\bar{t}$) (left), and gluon fusion ($gg \rightarrow t\bar{t}$) (right).

At Tevatron the total $t\bar{t}$ production cross section is 6.97 pb (at NLO) for $\sqrt{s}$=1.96 TeV $p\bar{p}$ collisions, whereas it is significantly larger at LHC. The actual cross section is estimated to be of 164±10 pb (at NNLL) (12) for pp collisions at 7 TeV, and is

expected to increase to 833 pb for pp collisions at 14 TeV.

As already mentioned, the top quark can be produced alone via weak interaction. The single tops are produced mainly in three different ways described in the Feynman diagrams of figure **1.5**. One needs to notice that for the s and t channels the top and anti-top cross sections are different in pp collision due to the Parton Distribution Function (PDF) of the proton.
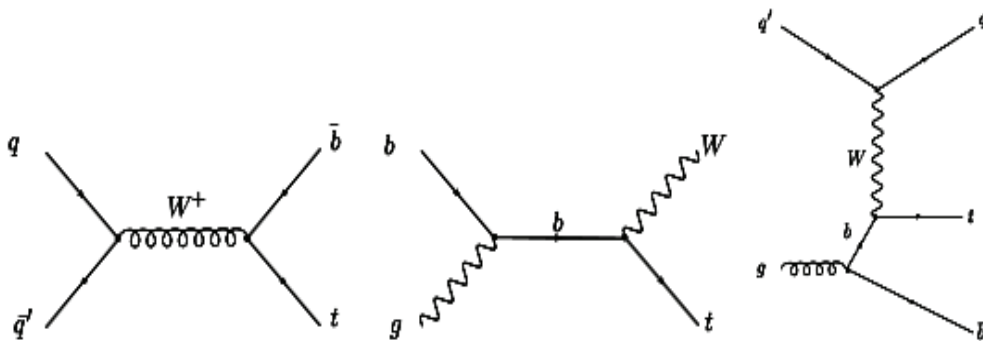
Figure 1.5: Production modes for the single top: (left) s channel, (center) tW channel, (right) t channel.

1.5 Top quark decay

An important feature of the top quark is its lifetime $\sim 10^{-24}$ps, about 10 times shorter than the characteristic formation time of hadrons (13). It means that the top quark decays before it hadronizes. The lifetime of a $t\bar{t}$ pair is in any case too small to allow a proper definition of a bound state with sharp binding energy.
Even if the t and $\bar{t}$ quarks are likely to interact with the remnant proton partons which don't take part in the hard interaction to form light hadrons[1] (*early fragmentation* process), this effect is negligible at the LHC energy. Hence we can assume that the top quarks are produced and decay like free quarks. The decay modes $t \rightarrow d + W^+$ and $t \rightarrow s + W^+$ are CKM suppressed relatively to $t \rightarrow b + W^+$ by factors of $|V_{td}|^2$ and $|V_{ts}|^2$. Since these values are small compared to $|V_{tb}|^2$ ($0.9990 < |V_{tb}|^2 < 0.9992$) only the decay $t \rightarrow b + W^+$ is considered, as shown in figure **1.6**. Focusing our attention on the $t\bar{t}$ pair production, several decay channels can be investigated. They correspond to the decay mode of the W boson (hadronic or leptonic) originating from the top decay.

- di-lepton channel ($t\bar{t} \rightarrow W^+b, W^-\bar{b} \rightarrow l_1\nu_1 b, l_2\nu_2\bar{b}$): in this case both W^+ and W^- bosons stemming from the t and $\bar{t}$ quarks decay in a lepton (e, μ, τ) plus

---

[1]See section1.6 to understand the behavior of the pp collision.
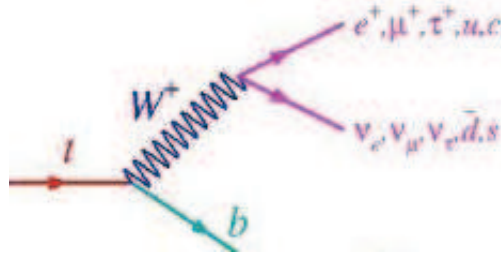
Figure 1.6: Main decay mode of the top quark.

the corresponding neutrinos. The final state contains two isolated leptons with high p_T, two jets coming from the b quark hadronization (b-jets), and high missing transverse energy due to the neutrinos presence. The branching ratio for this channel is about 10%.

- lepton + jets channel ($t\bar{t} \to W^+b, W^-\bar{b} \to q_1q_2b, l_1\nu_1\bar{b}$): it corresponds to the case in which one of the two W bosons decays into a lepton and the other one into a quark pair. So the final state contains one isolated high energy lepton, some missing transverse energy due to the presence of one neutrino, and 4 jets (two of them are b-jets). Each channel, electron+jets, muon+jets and tau+jets, has a branching ratio of about 15%. In this thesis we will present the first measurement of the $t\bar{t}$ cross section in the hadronic tau+jets channel at CMS. This specific channel has a branching ratio of 9.8%. The remnant 5.2% of the tau+jets branching ratio is due to the leptonic decay of the tau.

- fully hadronic channel ($t\bar{t} \to W^+b, W^-\bar{b} \to q_1q_2b, q_3q_4\bar{b}$): in this case both W bosons decay into a quark pair. So the final state consists of 6 jets, two of them being b-jets. Even if this channel is the one with the largest branching ratio, 45%, it is challenging to investigate due to the important multijet background.

A summary of the branching ratios for the different channels is presented in the figure **1.7**.

1.6 Cross section calculation

One of the main feature of the hadronic collisions is that the hadrons are not elementary particles, but are made of *partons*. In particular the protons are built out of three valence quarks (uud).

In general, when two protons collide, the interaction involves a parton pair as shown in figure **1.8**. Possible inelastic interactions are divided in two groups:

- collision between partons *head on* characterized by an high transverse momentum transfer. In this case it is possible to create particles with high mass as the top quark.

12
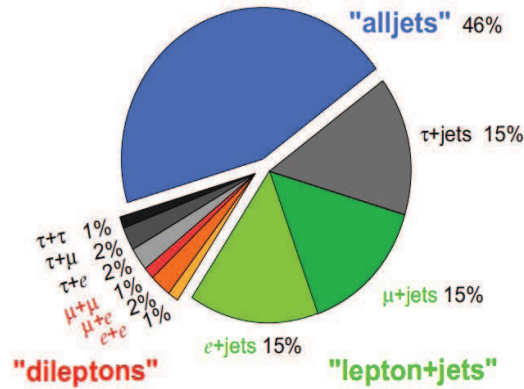
Figure 1.7: Branching ratios of the $t\bar{t}$ decay channels.
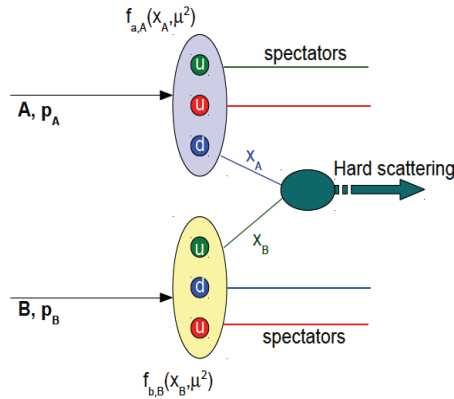
Figure 1.8: Illustration of a *pp* collision. The two protons, (A,B), are each made out of three partons (uud). Only a pair of the constituent partons are involved in the hard scattering, and they carry a fraction of the momentum of the incident proton (A,B) equal to x_a and x_b respectively.

- interaction with low transverse momentum transfer and small scattering angle with respect to the beam direction. These collision are usually referred to as *minimum bias*.

The *head on* interactions are rarer than the *minimum bias* events.

Another important aspect of the proton-proton collisions is that we can't determine event by event the available energy in the center-of-mass, since only a fraction $0 < x_a, x_b < 1$ of the energy of the incident protons, (A,B), is used in the partons interaction. Moreover, if $\sqrt{s}$ is the energy of the proton-proton center-of-mass, the energy of the parton interaction is given by $\sqrt{s'} = \sqrt{x_a x_b (p_A + p_B)^2}$ where p_A, p_B are the quadri-momentum of the incident protons. In a symmetric collider as LHC $p_A = p_B = p$, so $\sqrt{s'} = \sqrt{x_a x_b s}$ changes for each event since x_a and x_b are generally different from each other. This is why the center of mass of the parton system is

boosted along the beam direction. Even if it represents an additional experimental complication, it is maybe one of the main advantages of the pp collisions since we can explore concurrently a wide energy spectrum.

For the calculation of the $t\bar{t}$ cross section, hence, we use the parton model. The cross section calculation is based on the factorization theorem (14) according to which we can compute the cross section as the convolution of parton distribution functions (PDF) $f_i(x_i, \mu^2)$ for the colliding protons (A, B) and the hard parton-parton cross section σ_{ij}:

$$\sigma(AB \to t\bar{t}) = \sum_{i,j} dx_i dx_j f_{i,A}(x_i, \mu^2) f_{j,B}(x_j, \mu^2) \cdot \sigma_{ij}(ij \to t\bar{t}; \hat{s}, \mu^2) \qquad (1.14)$$

The parton distribution function $f_{i,A}(x_i, \mu^2)$ describes the probability density for finding a parton i inside the proton A carrying a longitudinal momentum fraction x_i. These are extracted from data and they exist in several different parametrizations. As an example in figure **1.9** the PDFs of the MSTW2008(15) parametrization are shown. The parton distribution functions and the parton-parton cross section
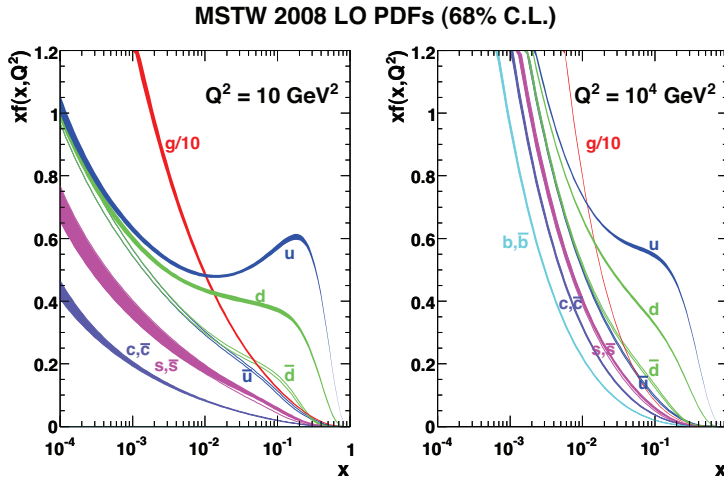
Figure 1.9: PDFs of the proton, in the MSTW2008 parametrization, at different scales: (left) $\mu^2 = Q^2 = 10$ GeV^2, (right) $\mu^2 = Q^2 = 10^4$ GeV^2.

$\hat{\sigma}_{ij}$ depend on the factorization and renormalization scale μ representing the momentum exchanged in the partons interaction. For calculating heavy quark production the commonly used scale is of the order of the heavy quark mass, so in our case $\mu = M_{top}$. Since in general the calculations are performed at finite order in the perturbation theory, cross-section predictions depend on the choice of μ. The effect of the μ-dependence is estimated by varying the scale between $\mu = M_{top}/2$ and $\mu = 2M_{top}$. The variations in the cross section are quoted as an indicative theoretical uncertainty of the prediction. The sum in (1.14) runs over all pairs of light partons (i, j) contributing to the process.

1.7   Reasons to study top physics

As already mentioned, the $t\bar{t}$ pairs are produced copiously at LHC. Hence they are used to calibrate the energy of the jets but also to estimate the efficiency of the b-tag algorithms (16). More generally since the top quark is the heaviest particle of the SM and its coupling with the Higgs boson is expected to be large, the top quark can be a special key to understand the mechanism of how particles aquire mass. In addition since the high top quark mass has the same order of magnitude than the mass of particles predicted by other beyond SM theories, the study of the top quark can be involved in the search of signatures of these alternative theories. More specifically the following reasons sustain the study of the top quark physics:

- a precise measure of the top mass can allow to constrain the Higgs boson mass. The huge number of $t\bar{t}$ pairs produced at LHC will allow to reduce the statistical error and also to better control the systematic uncertainty.

- As already explained the $t\bar{t}$ events are characterized by the presence of one or two high energy leptons, by missing transverse energy as well as by the presence of b-jets. These are similar ingredients to many new physics final states. The understanding of the $t\bar{t}$ events is thus crucial for the rejection of the background in many new physics direct searches.

- One more goal of the top physics program at CMS for the incoming year is the extraction of the parton distribution function (PDF) of the protons. These functions, describing the long-distance structure of the hadron, are assumed to be universal, i.e. process independent. The main goal of this study is to constraint the uncertainty of the PDFs that have a large impact of the LHC phenomenology, including top-quark predictions. The $t\bar{t}$ events should help to pin down the uncertainties of the gluon PDF at high value of carried momentum.

1.8   Top-antitop in the hadronic τ + jets channel

Several specific reasons support the choice to study $t\bar{t}$ events in the semileptonic hadronic τ channel. This channel, will be the largest one containing τ in the final state. Since the top is the heaviest quark and the τ the heaviest lepton, both belonging to the third generation of fermions, the study of this decay mode is of interest to rule out any mass or flavor dependent coupling of the top quark to tau final states. Besides that, this channel can be a significant background for the SUSY low mass charged Higgs boson search. In the next paragraph an introduction to the Supersymmetry (SUSY) theory is done.

The supersymmetry theory   As we know the Standard Model was extensively tested with high precision, but it is not able to explain a number of experimental observations like for example the nature of the *Dark Matter*. Several cosmological and astrophysical measurements, lead to think that 25% of the energy density in the Universe is composed by neutral and weakly interacting matter, (figure **1.10**). The SM has no possible candidates for such kind of matter (apart neutrinos, but thier mass is too low).
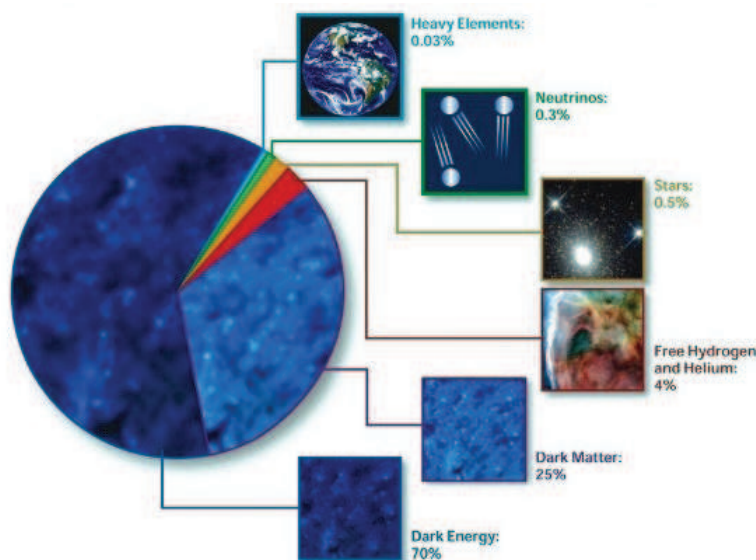
Figure 1.10: Composition of the universe according to the recent results (17).

It means that the Standard Model is an effective theory, that is valid at the currently accessible energy. It should represent an approximation at low energy of a more general theory whose effects will be visible only at higher energy. Hence, in order to find an explanation to the open issues, many alternative theories to the SM are developed. The most simplest possible extension of the SM is the two-Higgs-doublet model (18) which is the basis of the Supersymmetry theory.

The *Supersymmetry* (SUSY) is one of the most studied theories for the physics beyond the SM. There are mainly three arguments supporting SUSY:

- Hierarchy problem: the Higgs boson mass is stabilized against the radiative corrections (19).

- All gauge coupling constants that take part in the SUSY lagrangian naturally converge in only one value for an energy of the order of 10^{16} GeV (20). In other words, the theory provides an unification energy scale, (figure **1.11**).

- SUSY provides dark matter candidates (LSP).

16
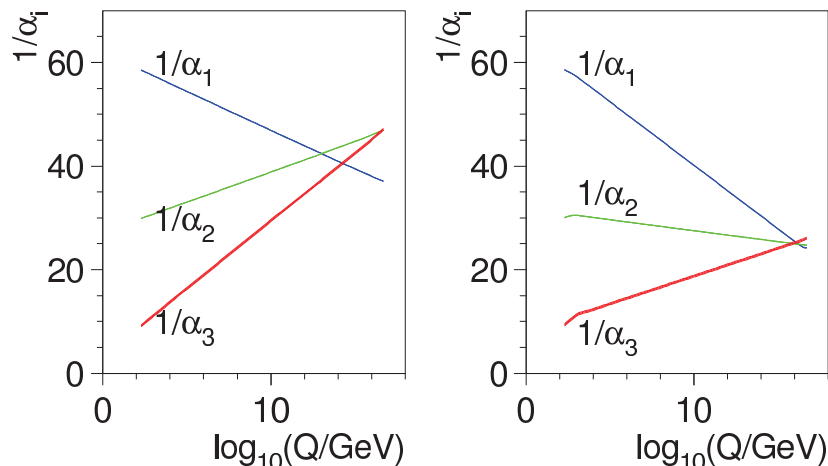
Figure 1.11: Evolution of the inverse of the gauge coupling constant as function of the logarithm of the energy for the Standard Model (left), and cMSSM SUSY model (right).

The SUSY theory relies on the existence of supersymmetric partners of the ordinary SM particles and two additional Higgs doublets. In order to introduce these supersymmetric partners two SUSY operators are defined, $(Q, \bar{Q})$. They turn a bosonic state into a fermionic one and vice versa:

$$\bar{Q}|Boson> = |Fermion>; Q|Fermion> = |Boson> . \tag{1.15}$$

This means that starting from a spin-n state, the SUSY operators produce a spin-$n \pm \frac{1}{2}$ state. The basic idea of SUSY is that the theory has to be invariant under a Q transformation. To realize this invariance we need for each SM boson (fermion) a corresponding superpartner boson (fermion), figure **1.12**.

In principle if SUSY is an exact symmetry, the superpartners should have the same mass as the SM counterparts, and they should have been already discovered by previous experiments. The absence of such observation could be due to the fact that it is a broken symmetry with SUSY masses large enough to escape the current experimental limits. The SUSY breaking mechanism is used to give a mass to the SUSY particles and is usually obtained by adding a term in the lagrangian involving only the SUSY particles, explicitly breaking the symmetry. This procedure introduces about 100 parameters which can be partly fixed if we assume that an underlying mechanism produces the SUSY breaking in a natural way, or in other words, if we assume that the SUSY breaking mechanism is spontaneous like the Higgs mechanism in the SM. Actually a form of spontaneous symmetry breaking mechanism is not possible in SUSY since none of the fields can have a non-zero vacuum expectation value that breaks SUSY without spoiling the gauge invariance. So most of the supersymmetric models obtain the SUSY breaking by introducing a so called hidden sector. In particular the SUSY breaking takes place at some unknown high energy in the hidden sector and it is mediated to the visible sector

Figure 1.12: A schematic summary of the SM particles with their super partners.

(containing the superparticles) by the exchange of weakly interacting *messengers*, (figure **1.13**).
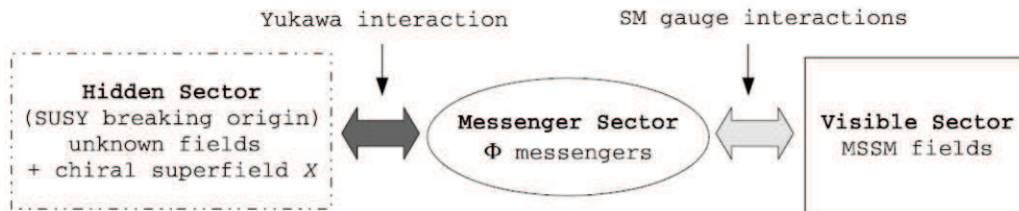
Figure 1.13: A schematic view of the Gauge-Mediated SUSY Breaking.

The simplest SUSY model is the cMSSM (constraint Minimal SuperSymmetric Model) in which, the mass, couplings and interactions between the superparticles contained in the visible sector are fully fixed by 4 parameters and one sign:

- m_0: the common mass of the scalar particles;

- $m_{1/2}$: the common mass of fermions;

- A_0: the common trilinear coupling constant;

- $tan\beta$: the ratio between the vacuum expectation values of the two charged Higgs bosons doublets;

- sign of μ that represents the Higgs mass mixing.

All of them are defined at the energy scale of the unification. It is interesting to notice that for $tan\beta > 20$ the charged Higgs decays into a τ. In particular if the mass of the H^+ is smaller than the top mass, we should see a new decay channel of the top, $t \to H^+ b$ where H^+ decays mainly in $\tau + \nu_\tau$. An excess in the cross section measurement in the $t\bar{t} \to \tau+$jets channel with respect to the standard model expectation could be seen.

1.9   Monte Carlo generators

In the last part of this chapter we would like to present a brief description of the events simulation in CMS. The Monte carlo production of hadronic collisions results from a long sequence of steps which profit from both analytical and numerical computations. We can summarize the simulation of events with the following steps (figure **1.14**):

- The cross section of the involved process is computed using the Matrix Element (ME) method (21). It means that the matrix element corresponding to the Feynmann diagram of the considered process is calculated. Partons and leptons in the final state are produced according to the differential cross sections computed in the previous step;

- Resonances produced in the hard event are then decayed;

- When two partons take part in the hard event, they emit gluons that are continuously emitted and reabsorbed. This effect is called Initial State Radiation (ISR). Also the final state partons can produce additional radiation known as Final State Radiation (FSR). In principle we can think to simulate ISR and FSR using again the ME, but this technique presents mainly two problems. The first one is that the exact computation of the matrix element of a diagrams with many gluons is too complex (22). The second one is due to the fact that, even if only one of the emitted gluons is collinear to the partons which emit, the corresponding ME diverges. So the ME technique is only usable in case of hard emission with large scattering angle. Hence, in order to cover the phase-space not solved by the ME we need to use another technique. It is known as Parton Shower (PS) (23). It starts by computing the vertex of the diagram at the Leading Order (LO) and proceeds with the simulation of the gluons distributed according with the momentum distribution obtained in the collinear limit of the ME. So the PS is an approximation of the ME in the collinear limit that allows to remove divergencies. The two techniques are therefore complementary, and a combination of them is needed in order to profit of their peculiarities in their respective validity ranges.
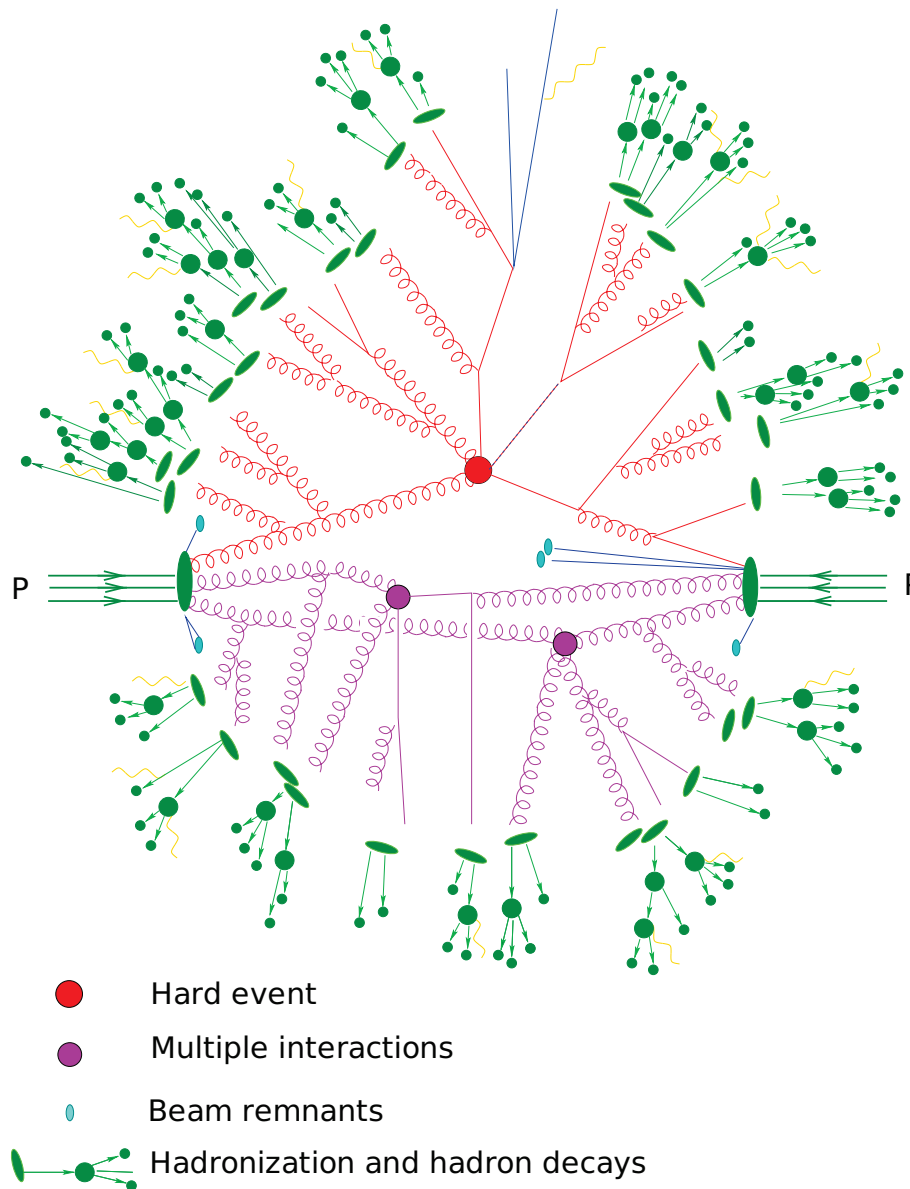
Figure 1.14: Sketch of the generation of simulated event in a typical event generator. The simulation of the hard scattering is followed by the softer multiple interactions. At the end the hadron remnants are treated. Quarks and gluons are turned into hadrons by hadronisation. Then the τ leptons and b-hadrons are decayed.

- Partons which don't take part to the hard scattering can give rise to interactions with smaller transfered momentum. These Multiple Parton Interactions (MPI) contribute to the so called underlying structure of the event that must be simulated and corrected for the initial and final state radiations.

- The calculations described above are done in the perturbative regime, but, since the produced partons move away from each other, the coupling constant increases more and more giving rise to the confinement effects. When the coupling constant is strong enough a new quark-antiquark pair is produced from the vacuum and the partons turn into hadrons. This is the hadronization step of the simulation.

- At the end, the generator deals with the decay of b-hadrons and τ leptons. For τ leptons the generator technique used is called TAUOLA and is described in (24). While particles with very short lifetimes are decayed by the generator itself, those that live long enough to arrive at the detector are left undecayed.

After the description of the theoretical aspects of the SM and more specifically of the top-quark physics, we shall describe the CMS detector in the next chapter.

Chapter 2

The CMS experiment at the LHC

2.1 Introduction

The Large Hadron Collider (LHC)(25) is the most powerful proton-proton collider ever built. The operation at LHC started in September 2009 and since March 2010 we have stable beams of protons which collide at 7 TeV in the center of mass. Thanks to the high energy and luminosity that the LHC can provide, we are able to study rare events produced with a tiny cross section, down to the femtobarn.

Why was it chosen to built an hadron collider instead of an electron collider as LEP? There are mainly two reasons: first of all the energy in the center of mass of electron collisions is limited by the synchroton energy emission, a factor $O(10^{12})$ higher than for the proton synchroton emission. Then, due to the composite nature of protons, with an hadron collider we can investigate a wider energy spectrum of collision products, which in addition can be explored simultaneously. This is an important feature for an experiment involved in the discovery of "unknown" new physics.

In this chapter we shall shortly describe the LHC complex and we shall focus our attention on the components of the CMS experiment (26).

2.2 The Large Hadron Collider at CERN

The LHC accelerator was built in the same tunnel that hosted the LEP the previous electron-positron collider. In figure **2.1** a schematic view of the different constituents of the LHC is shown.

In the LHC two proton beams are accelerated in a 27 km circumference ring. The final acceleration is provided in several steps. The proton injection starts at the duo-plasmatron, which is the proton source. A linear accelerator (LINAC) boosts the protons to an energy of 750 KeV using Radio Frequency Quadruples. Then in the Proton Synchrotron Booster (PSB) the energy of protons is increased up to 1.4 GeV. After that the LHC bunch train starts in the Proton Synchrotron (PS), in
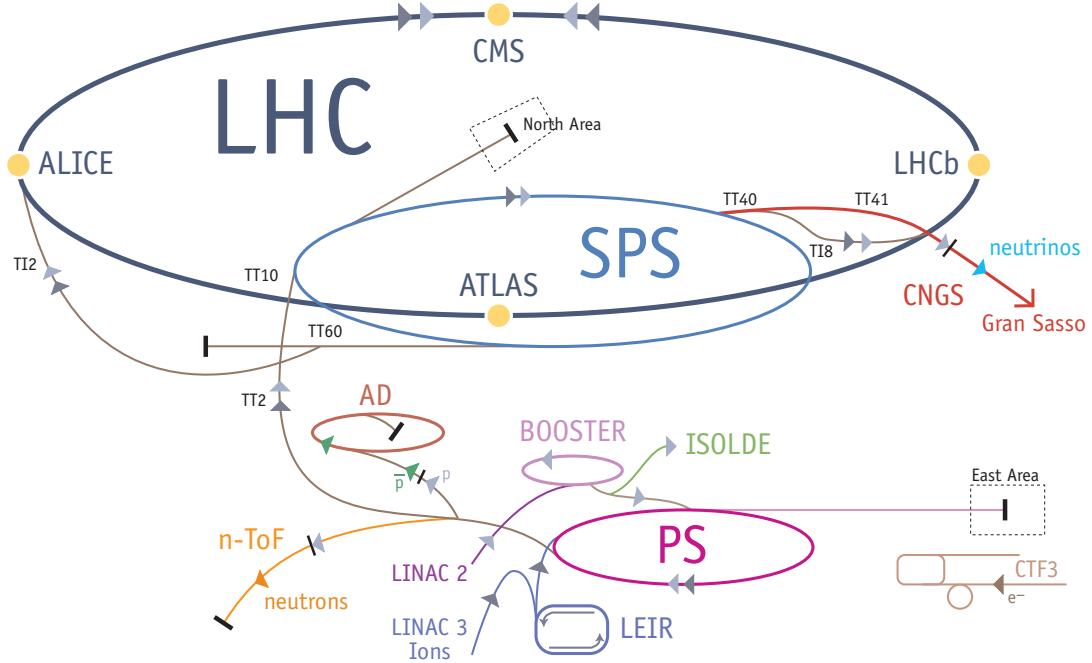
Figure 2.1: The LHC accelerator complex.

which the energy increases again up to 25 GeV. Then the protons are accelerated up to 450 GeV by the Super Proton Synchrotron (SPS), and finally they are injected into the LHC ring where the acceleration reaches the nominal proton energy of 3.5 TeV.

The two beams collide in four interaction points where the four main experiments are located. Two general purpose experiments, called ATLAS (27) and CMS, are dedicated to general Standard Model measurements and to the search for new physics; one experiment called LHCb (28) is dedicated to the b-hadron physics and more precisely to the measurements of CP violation; one experiment called ALICE (29) is built to investigate heavy ion physics. This collider is able to investigate mass scales from the order of a few GeV, as in the case of B mesons, up to a few TeV for the discovery of new vector bosons or quark compositeness. In order to extend the LHC capability to explore rare processes a very sophisticated magnet system was designed in order to keep the high momentum protons in the machine orbit. The classical formula that relates the bending radius with the charged particles momentum and the magnetic field is:

$$B[\text{T}] = \frac{p[\text{GeV}]}{0.3 \cdot \rho[\text{m}]} \tag{2.1}$$

where B is the magnetic field in Tesla, p the momentum in GeV and ρ the orbit radius

in meters. The LHC circumference is about 27 km, so the magnetic field needed to keep in orbit 3.5 TeV protons is about 2.7 T, a value close to the technological edge for superconducting magnets nowadays. A sophisticated superfluid helium cooling system is used in order to keep the dipole at a temperature below 1.9 K. As we can see from eq.(2.1) the beam energy is limited by the power of the magnetic field and by the circumference of the LHC. So the only way to increase the rate of rare and interesting events is to raise the *luminosity* $\mathcal{L}$. This variable is defined as:

$$\mathcal{L} = f \frac{n_1 n_2}{4\pi \sigma_x \sigma_y} \tag{2.2}$$

where n_1 and n_2 are the number of particles in beams 1 and 2 respectively, f is the collision frequency, and $\sigma_{x,y}$ are the transverse dimensions of the beam. The rate of events, n, for a given process with cross section σ is related to the luminosity as:

$$n = \mathcal{L}\sigma. \tag{2.3}$$

In figure **2.2** the production cross section of many Standard Model processes are shown as a function of the center-of-mass energy. In the hard proton proton collision, the energy in the center of mass $\sqrt{\hat{s}}$ of the fundamental interaction is smaller than the total centre of mass energy, $\sqrt{s}$:

$$\sqrt{\hat{s}} = \sqrt{x_1 x_2 s} \tag{2.4}$$

where x_1 and x_2 are the energy fractions of the two partons participating in the hard scattering (Bjorken factors). Since the center of mass of the two hardly interacting partons is boosted along the beam direction, the Lorentz boost invariant observables are very important to characterize the event. One example is given by the transverse momentum p_T, defined as the projection of the momentum vector on the plane perpendicular to the beam axis. Another useful observable is the rapidity y defined as:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) = \tanh^{-1}(p_z/E) \tag{2.5}$$

where E is the energy of the particle, and p_z the projection of particle momentum along the beam direction. If we apply to y a Lorentz boost along z with speed β, the rapidity y becomes $y - \tanh^{-1}(\beta)$. This is why the rapidity differences and the shape of the rapidity distribution are invariant.
It is also possible to show that in the ultra-relativistic approximation ($m \sim 0$ so $p_z \sim E$) the rapidity y is equal to the pseudorapidity η defined as

$$\eta = -\ln \left(\tan \left(\frac{\theta}{2} \right) \right) \tag{2.6}$$

where θ is the polar angle between the direction of the momentum of the particle and the beam direction. Since we usually work in the ultra-relativistic approximation, the pseudorapidity is often used to replace the rapidity because it depends only on the direction of the momentum of the particle.
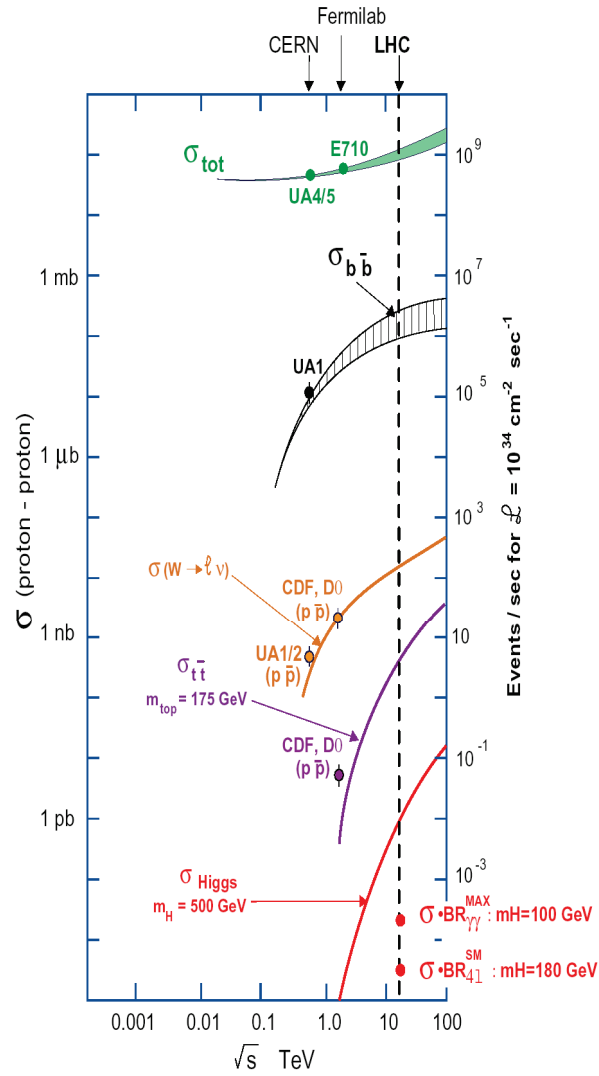
Figure 2.2: Expected cross section for several process as a function of the center-of-mass energy of pp collision.

The LHC schedule for 2010 and 2011. The four LHC experiments registered the first pp collisions at $\sqrt{s}$ =7 TeV on March 30th 2010 after only one month of beam commissioning. The 2012 data taking will start with collisions at 8 TeV.

Until summer 2011 the number of bunches per beam gradually increased. In September 2010 the number of bunches per beam reached 368, corresponding to a luminosity of 2.1×10^{32} cm^{-2}s^{-1}. The final integrated luminosity delivered by the LHC in 2010 run was 36 pb^{-1}. The 2011 run started on March 13th. The number of bunches per beam increased until 1380. In term of instantaneous luminosity, we arrived to 3×10^{33} cm^{-2}s^{-1} already in September. Thanks to the gradual increase of the istantaneous luminosity (up to 5×10^{33} in December 2011), CMS collected 5.0 fb^{-1} at the end of 2011.

2.3 The CMS Detector

The Compact Muon Solenoid (CMS)(26) is a general purpose experiment. It is characterized by the presence of a superconducting solenoidal magnet able to provide a magnetic field of 3.8 T that allows a compact design of the detector. Its total dimensions are: 21.6 m of length, a diameter of 15 m, and a total weight of 12500 tons.

The structure of CMS consists of several cylindrical detecting layers, coaxial with the beam direction in the *barrel* region. The barrel layers are closed at both ends with disks, referred to as the *endcap* regions. In figure **2.3** a schematic view of the CMS detector is presented. The strengths of the CMS design are a redundant muon tracking system, a very good electromagnetic calorimeter and a high quality tracking system.

The coordinate system used in CMS is a right-handed cartesian frame. The x axis points towards the LHC centre, the y axis is directed upward along the vertical and the z axis corresponds to the beam direction. Usually the cartesian coordinates are replaced by a pseudo-angular reference frame suggested by the cylindrical symmetry of CMS and by the invariant description of proton-proton collisions. The pseudo-angular reference frame is defined by the triplet (r, ϕ, η), where r is the distance in the transverse plane, ϕ is the azimuthal angle, measured starting from the x axis positive direction, and η is the pseudorapidity.

The CMS experiment contains four main subdetectors. The Silicon Tracker, the Electromagnetic Calorimeter (ECAL) and the Hadronic Calorimeter (HCAL) are located inside the magnetic field provided by the superconducting solenoid, while the Muon Chambers are located in the magnet return yoke. In the following we will describe in more details the main features of each subdetector.
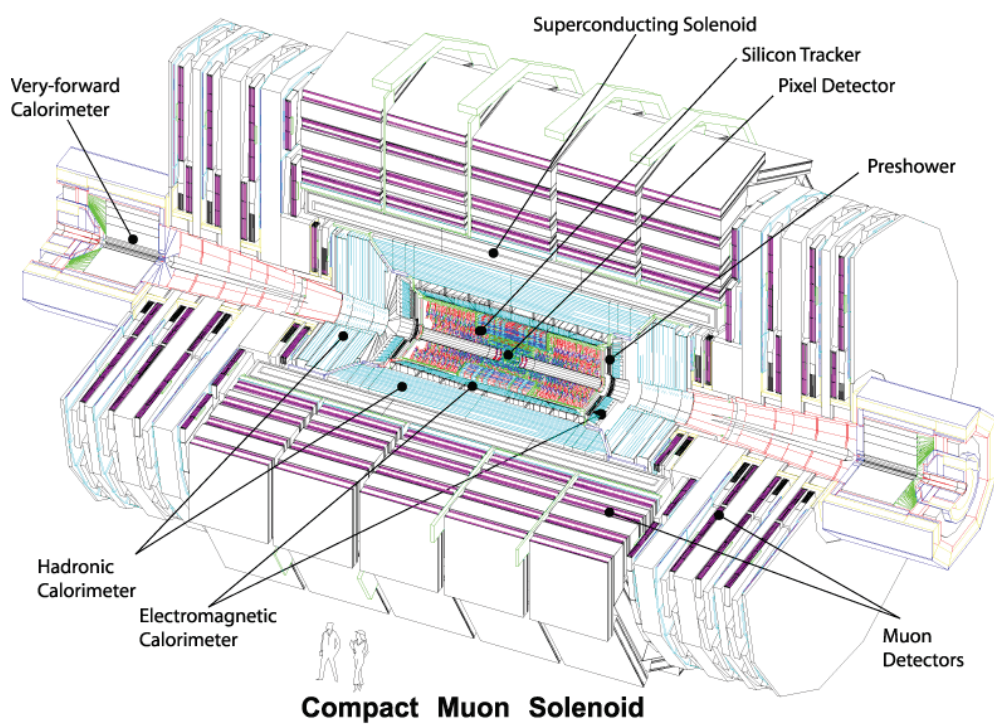
Figure 2.3: A view of the CMS detector. Each subdetector is labeled.

2.3.1   The Solenoid

The CMS magnet provides a magnetic field of 3.8 T and is the biggest supercon-ducting solenoid ever built. It is made of four layers of NbTi and its dimensions are 6 m in diameter and 12.5 m in length. The energy stored in the magnet is about 2.6 GJ at full current. The energy can be dumped to resistors in only 200 ms in case of a quench, when a magnet looses its superconducting property. The magnet return yoke of the barrel has 12-fold rotational symmetry and it consists of three sections along the z-axis; each section is splitted into 4 layers interspersed with the Muon Chamber. The residual magnetic field present in the yoke (it is about half of the field (1.8 T) in the central region of the detector) is used to curve the muons in the Muon Chamber in order to facilitate the muon momentum measurement. More information about the features of the CMS solenoid can be found in (30).

2.3.2   The Tracker

The Silicon Tracker is placed in the region closest to the beam pipe and it covers the region $|\eta| < 2.4$, $r < 120$ cm. The main goal of this part of the detector is to provide a precise measurement of the momentum of the charged particles and to allow a precise determination of the position of the secondary vertices, hence a good resolution on the point where the particles decay or interact with the detector.

In addition, since at LHC we produce complex events containing high multiplici-ties of charged particles, the recontruction of tracks lead to a complicated pattern recognition. To ease this problem we have only two possibilities: the first one is to keep low the occupancy of the detector, it means that a high granular detector is needed, the second one is to provide a large hit redundancy that is achieved with a large number of detecting layers. To do that the tracker consists of two main parts: a Silicon Pixel detector, and a Silicon Microstrip detector. In the following section these subdetectors will be described in more details.

Since the tracker detector is the closest one to the beam pipe, pixel micro-strips and readout electronic are subjected to a huge flux of radiation that can be the cause of important damages. This is why, the pixel detector, which is exposed to the highest flux per unit area, will be replaced at least once during the LHC lifetime. In order to limit the effect of radiation damage on the sensor performances the tracker operates at low temperature (-10 °C).

Another important aspect for the tracker is the material budget[1] since the elec-tron energy loss due to bremsstrahlung and the presence of nuclear interactions of hadrons need to be kept as low as possible to not spoil the tracking performances. The tracker depth in terms of radiation length X/X_0[2] and in terms of interaction

---

[1]With material budget expression, we usually refer to the amount of material needed to build the detector which is not active for the detection of a particle

[2]X_0 is the distance over which a high energy electron reduces its energy to a fraction $1/e$ of the initial energy by bremsstrahlung emission.

length λ/λ_0[1] as obtained from the full simulation of the tracker is shown in figure **2.4** as a function of η. The material budget is higher in the region $1< |\eta| <2$ that is the transition area between barrel and endcap.
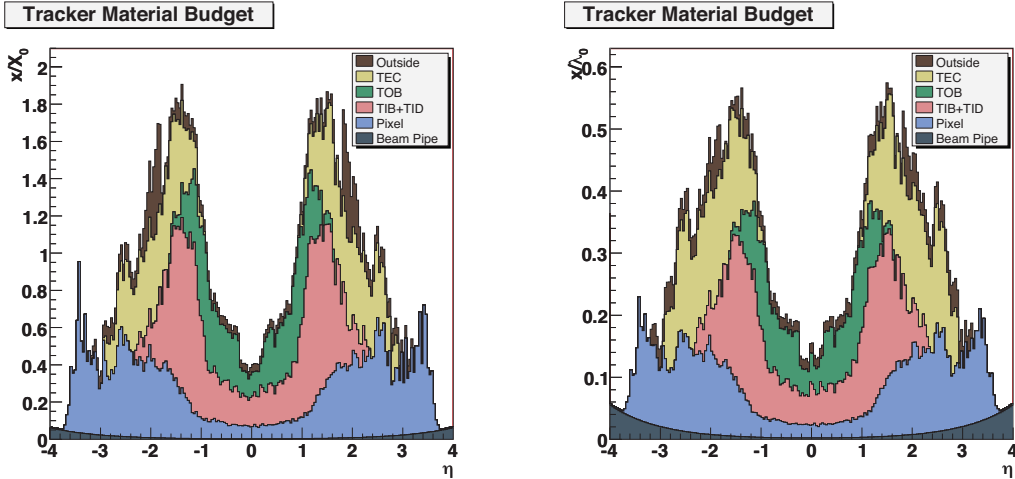
Figure 2.4: Radiation length (on the left) and interaction length (on the right) of the tracker as a function of η. Contributions from different components are put into evidence.

During the 2010 data taking, studies of the material budget with real data were performed. They were mainly based on the reconstruction of the secondary vertices coming from photon conversions and nuclear interaction within the material of the traker volume.

Up to 70% of photons traversing the tracker material, convert in e^+e^- pairs. In order to reconstruct these photons, a sophisticated tracking algorithm is used as described in references(31), (32). In figure **2.5** the conversion vertices reconstructed in data are shown in the (z,R) plane. The tracker structure emerges clearly.

About the nuclear interactions, from simulation we expect that about 5% of the neutral pions with $p_T \sim 5$ GeV/c interact with the tracker material. The reconstruction of the nuclear interactions is fully based on the tracker information (31) and in figure **2.6** the distribution of the transverse radius of the reconstructed nuclear interaction vertices for data and simulation is shown. A good agreement between data and simulation is observed.

Another crucial point is the alignment of the tracker modules which plays a crucial role to obtain high precision on the spatial resolution. Deviations are caused by assembly inaccuracies, deformations due to cooling and stress from the magnetic

---

[1]λ_0 is the mean free path of a hadron before having an interaction when traversing a material.
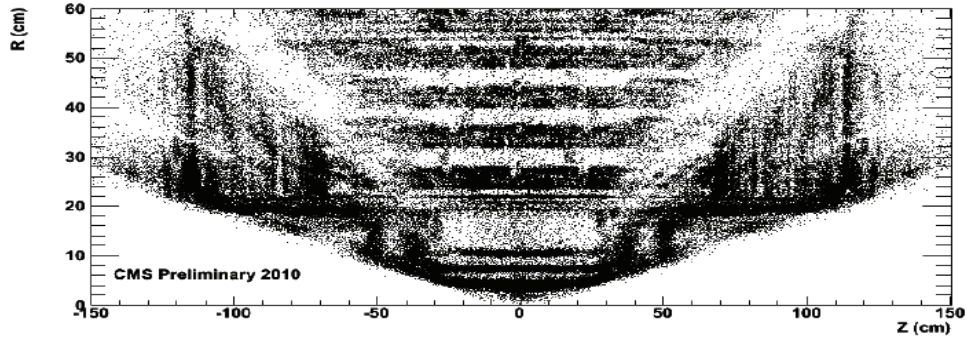
Figure 2.5: Conversion vertices reconstructed in data in the (z, R) plane.

Figure 2.6: Distribution of the transverse radius of the reconstructed nuclear inter-action vertices for data and simulation. In data the coordinates are computed with respect to the actual average Pixel barrel centre.

31

field. The geometry was determined during assembly to an accuracy of 80 to 150 μm. An infrared laser system is used for continuous monitoring of the position of selected tracker modules. The final alignment is done with tracks from well known physics processes, e.g. cosmic muons, or di-muons from Z boson decays.

The Pixel vertex detector
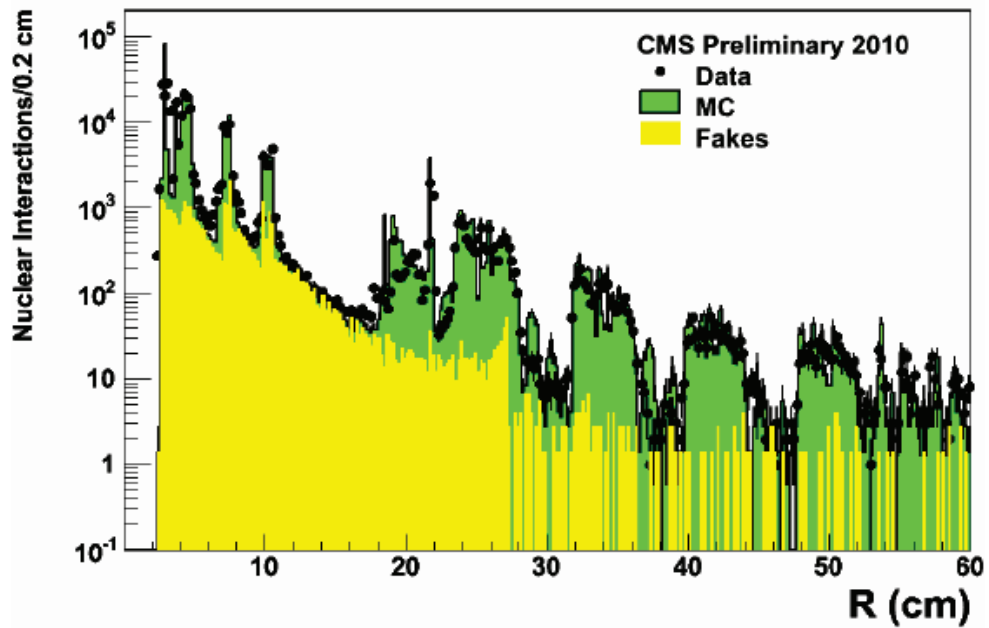
The pixel device (33) is the innermost part of the tracker. In figure **2.8** a schematic view of the CMS pixel detector is shown. It is built with three barrel layers (with a length of 53 cm each) positioned at $r = 4.4$, 7.3 and 10.2 cm, and two disks on each side, placed at $z = \pm 34.5$ cm and ± 46.5 cm to guarantee at least two crossed layers per track coming from the center of the detector within the fiducial acceptance $|\eta| < 2.5$. Each layer is composed with modular detector units, containing a 250 μm thin segmented sensor plate with highly integrated readout chips. The size of each pixel is 100×150 μm^2. The rectangular shape has been chosen in order to optimize the resolution of both coordinates, $r\phi$ and z. In particular we have a resolution of 10 μm for the $r\phi$ coordinate, while the resolution along the beam axis (z) is of 20 μm. A good resolution from the pixel detector is an important feature to provide a good measurement of the vertices.

To read the signal coming from each pixel, these are combined with analog signal readout to profit of the charge sharing effect among pixels and improve the position resolution by interpolation. The charge sharing between pixels is enhanced by the Lorentz drift of the charge carriers, which is about 25^0 in the barrel and 4^0 for electrons in a 4 T magnetic field at 100 V of bias voltage, three times wider than for the holes. Therefore initial n-type substrate sensors are chosen to collect electron signals on n^+ implants, which in turn are more radiation hard. In the barrel the pixels are tilted to induce significant charge sharing between neighbouring implants in the rϕ plane, improving the intrinsic hit[1] resolution down to 10-15 μm, much lower than 150 μm, the width of each n^+ implant. In figure **2.7** the mechanism of the charge sharing induced by Lorentz drift is presented.

Charge sharing is present also along z direction for inclined tracks leading to a similar resolution. The detectors placed on the disks are rotated with an angle of 20^0 around the central radial axis to benefit of charge sharing improved both in r and $r\phi$ directions by induced Lorentz effects.

The silicon micro-strips detector

The silicon micro-strips detector (34) is the outer part of the tracker and it is built with layers of silicon microstrip. The detector unit is made with one or two sensors

---

[1]The *hit* definition is not trivial, but at this level it can be defined as the trajectory point intersecting the detector layer plane.
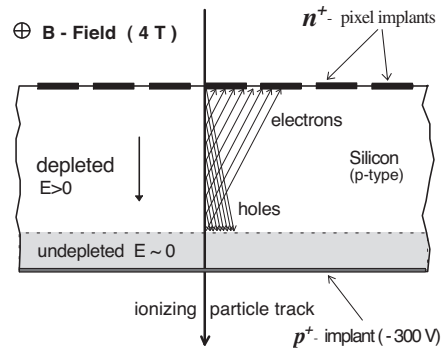
Figure 2.7: Charge sharing induced by Lorentz drift.
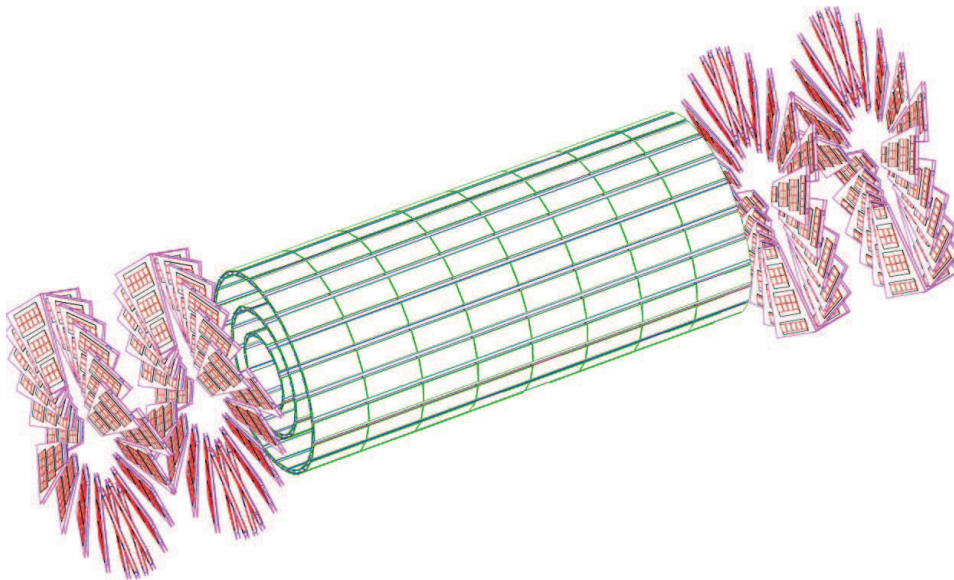
Figure 2.8: A schematic view of the CMS pixel detector.

glued on a carbon fibre mechanical support together with the read out electronics. The sensor is a n-type phosphorus doped substrate with p^+ implant strips. The p^+-n junction is reversely biased by applying a positive voltage (hundreds of Volts) to the n side keeping the strips at ground. In this way the region between the junction and the backplane is deprived of free charge carriers, and the only present charges are the thermally created ones. In particular, the passage of an ionizing particle through the silicon creates electron-hole pairs which drift in the electric field towards the backplane and the p^+ implants respectively. Since the mean energy required to create an electron-hole pair in silicon is 3.6 eV, a minimum ionizing particle (mip) passing through a 300 μm thick sensor with an average energy loss per path length of 390 eV/μm should create 32500 electron-hole pairs. The signal coming from each strip is then transmitted to ADCs located in the counting room via optical links. In order to decouple the readout electronics from the detector leakage current, insulating capacitor layers of dielectrics (SiO_2, Si_3N_4) are placed between the p^+ and the aluminum strips electrodes.

The silicon strip tracker is divided in four parts, as it is shown in figure **2.9**: TIB
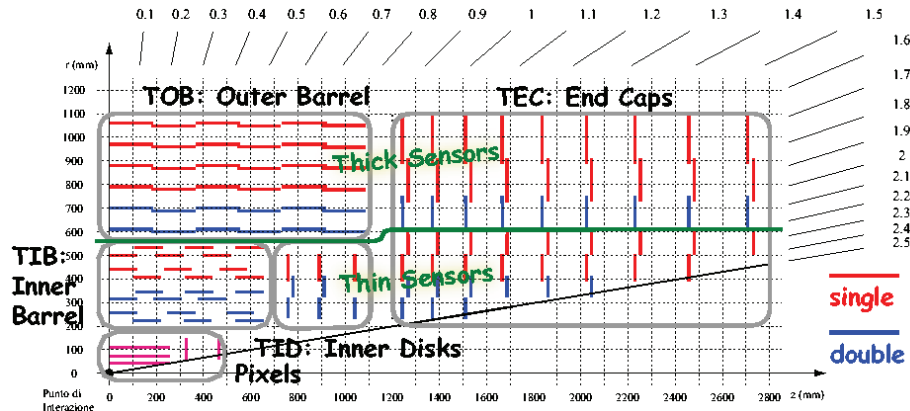
Figure 2.9: A schematic view $r - z$ of the CMS tracker subdetector.

(Tracker Inner Barrel), TID (Tracker Inner Disks), TOB (Tracker Outer Barrel) and TEC (Tracker EndCap) that cover a tracking volume up to $r = 1.1$ m with a length of 5.4 m.

The expected resolution for the CMS tracker, for three different parameters of the tracks as a function of pseudorapidity for muons with transverse momentum of 1, 10, 100 GeV respectively, is presented in figure **2.10**.

Figure 2.10: Resolution for three different track parameters for muons with p_T =1/10/100 GeV. Muons are reconstructed using only the tracker information: transverse momentum (left), transverse impact parameter (center), longitudinal impact parameter (right).

2.3.3 The Electromagnetic Calorimeter (ECAL)

The main goal of the ECAL(35) is the precise measurement of the energy of electrons and photons. The architecture and the design of the ECAL were performed in order to satisfy the requests imposed by th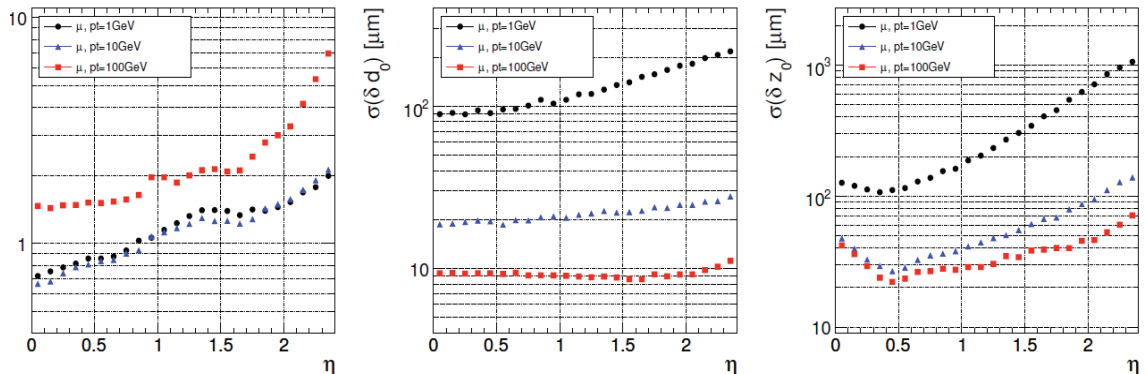e H$\rightarrow \gamma\gamma$ analysis, which is performed looking for an excess in the di-photon invariant mass distribution. Since the invariant mass width is dominated by the experimental resolution and is required to be order of 1%, high granularity is needed to improve the measurement of the angle between the two photons and to obtain a good $\pi^0 \rightarrow \gamma\gamma$ separation.

For this purpose an homogeneous calorimeter has been chosen. The ECAL is divided into a barrel and two endcaps containing almost 76000 Lead Tungstate (PbWO$_4$) scintillating crystals[1]. The barrel covers the $|\eta| < 1.479$ region. It consists of 36 supermodules built as a matrix of 20 crystals in $\phi \times 85$ crystals in η and covering an azimuthal angle of 20°. The supermodules are divided along η in 4 submodules, the basic unit of ECAL, made with 5×2 crystals each one. The barrel granularity is $\Delta\phi \times \Delta\eta = 0.0175 \times 0.0175$. The crystals are grouped into 5×5 arrays called trigger towers, providing information to the trigger system. To avoid that cracks might align with the particles trajectories, the crystal axis are tilted with respect to the direction from the interaction point, both in ϕ and in η. In figure **2.11** a schematic view of the ECAL is shown. Each endcap consists of two halves (Dees) and covers the $|\eta|$ region

---

[1]Different reasons brought to the choice of the PbWO$_4$ as active medium for ECAL. First of all the PbWO$_4$ crystals have a low light yield ($\sim$10 photo-electrons/MeV), which allows a good an internal amplification for the photodetectors. In addition the short radiation length (X_0 = 0.89 cm) and Moliere Radius (RM = 2.19 cm) of this material, allow one to build a compact and high granularity calorimeter. Another important aspect is its fast response ($\sim$80% of the light is collected within 25 ns), which is a crucial issue in the high LHC rate. Finally, since the PbWO$_4$ has a good intrinsic radiation hardness, it can work in the LHC environment.
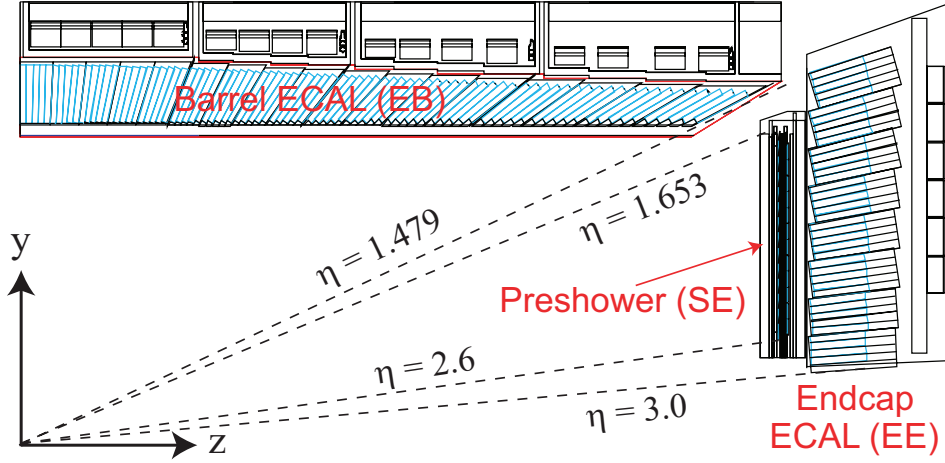
Figure 2.11: A schematic view of the CMS electromagnetic calorimeter (ECAL).

between 1.48 and 3. The size of each crystal is 220×24.7×24.7 mm³ and they are grouped in structures of 5×5 crystals called super-crystals. The granularity varies from $\Delta\phi \cdot \Delta\eta = 0.0175 \times 0.0175$ to 0.05×0.05.

To improve the π_0/γ separation and the vertex identification, a preshower is designed to cover the region between $|\eta|=1.6$ and 2.6. It works at a temperature of 5°C and it consist of two lead converters (2 X_0 and 1 X_0 thick) followed by silicon strips with a pitch of less than 2 mm. Important requirements are imposed on the front-end (FE) electronics since it has to be fast enough to sustain the 25 ns LHC crossing rate and it has to be radiation hard. The radiation, in fact, makes the crystals partially opaque to their emitted light. This effect is limited to ∼5% of the crystal light yield thanks to the radiation resistant properties of the PbWO$_4$, but in any case the status of the crystals has to be monitored. A monitoring system has been developed for this purpose, based on the injection of laser light into each crystal to measure its transparency and perform fast corrections to its calibration. During the data taking, the monitoring system provides regularly light pulses. Dedicated runs to follow the crystals behavior are also foreseen during the LHC refills.

Another crucial aspect that will be important for the offline data analysis is the energy resolution of the ECAL (36). Usually the energy resolution of an homogeneous calorimeter is written as

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{a}{\sqrt{E}}\right)^2 + \left(\frac{b}{E}\right)^2 + c^2 \qquad (2.7)$$

where a, b and c represent respectively the stochastic, noise and constant term of the energy resolution. The stochastic term represents the contribution of the

fluctuations in the number of electrons which are produced and collected. Since the fluctuations are poissonian, the stochastic term is $1/\sqrt{n_{pe}}$, where n_{pe} is the number of photo-electrons which are emitted per energy unit. The main contributions to this term come from:

- the efficiency in the light collection and from the quantum efficiency of the photo-detectors,

- the fluctuations in the multiplication process inside the photodetectors.

The typical value for stochastic term is 0.027 for the barrel and 0.057 for the endcaps. The noise term b represents contributions from the electronic noise, due to the photodetector, to the preamplifier, and from pileup[1] events. The mean value for this term is $b = 0.12$. The constant term c is the main term at high energies and it includes many different contributions like:

- the stability of the operating conditions, such as the temperature and the high voltage,

- the presence of dead material[2] in front of the crystals,

- the longitudinal non uniformity of the crystal light yield,

- the intercalibration errors,

- the radiation damage of the crystals.

The target value for the constant term c of the CMS ECAL is 0.5%.

2.3.4 The Hadronic Calorimeter (HCAL)

The CMS hadron calorimeter (HCAL) (37) works together with the ECAL in order to measure the energy and direction of charged and neutral hadrons, and the energy imbalance in the transverse plane, E_T^{miss}.

The central barrel and endcap HCAL detectors are placed in the high magnetic field of the solenoid and completely surround the ECAL subdetector. They provide a good segmentation, a decent energy resolution and cover the $|\eta| < 5$ region. The HCAL is made of four subdetectors (see figure **2.12**):

---

[1]With pileup term we refer to the several collision that take place in the same bunch crossing due to the high LHC luminosity.

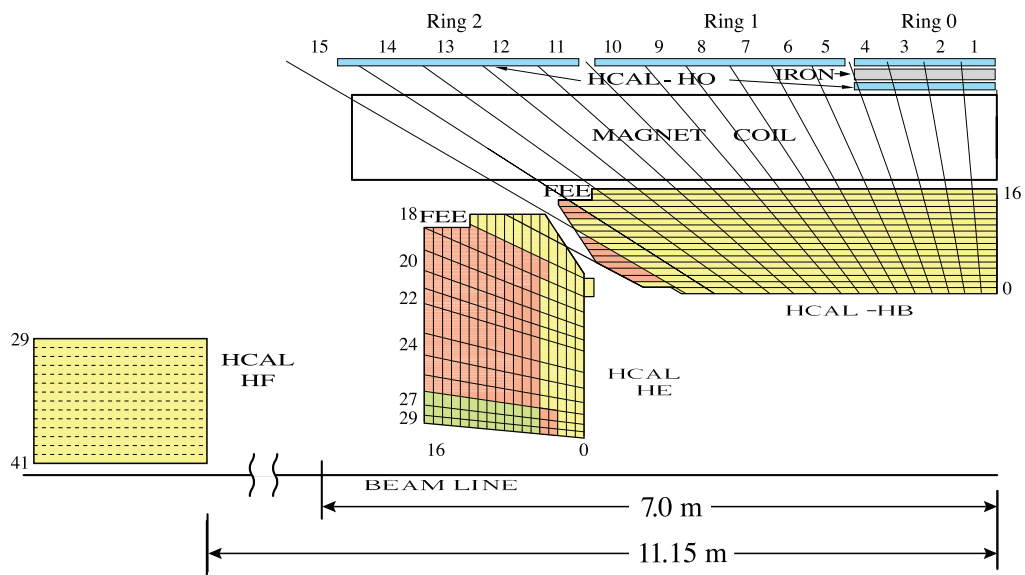[2]We usually call *dead material* the needed material in the detector that is not active for the detection.

Figure 2.12: A schematic view of the CMS hadron calorimeter (HCAL).

- Barrel Hadronic Calorimeter (HB): it is placed inside the magnetic coil and it covers the central pseudorapidity region, up to $|\eta| = 1.3$. The barrel hadron calorimeter consists of two half barrels, each composed of 18 identical 20° wedges in ϕ. Each wedge is composed of flat brass absorber plates parallel to the beam axis and interleaved with 3.7 mm thick plastic scintillators. The signal is readout through wavelength-shift fibres and hybrid photodiodes (HPD). The granularity is of the order of $\Delta\phi \times \Delta\eta = 0.087 \times 0.087$. The HB has an energy resolution for single pions of approximately $120\%/\sqrt{E}$. The minimum depth is about 5.8 λ_0.

- The Endcap Hadronic Calorimeter (HE) is located as well inside the magnetic coil and it is made of two endcaps extending the angular coverage up to $|\eta| = 3$. The HE has the same architecture as HB with the same granularity. The signal is read through wavelength-shift fibres and hybrid photodiodes.

- The Outer Hadronic Calorimeter (HO, or Tail Catcher) is placed in the barrel region, and was added outside the magnetic coil in order to extend the depth of the calorimeter in terms of nuclear interaction length. The total depth in the central region is thus extended to about 11.8 λ_0. It consist of two scintillator layers with the same granularity than the one of the HB.

- The Forward Hadronic Calorimeter (HF) consists of two units placed outside the magnetic coil, at ± 11.2 m from the interaction point along the beam direction. It extends the pseudorapidity coverage up to $|\eta| = 5$. The material of the two units, steel absorbers and embedded radiation hard quartz fibers, provide a fast collection of Cherenkov light. The granularity of HF is $\Delta\phi \times \Delta\eta = 0.17 \times 0.1745$.

In order to obtain a reference calibration and to measure the characteristics of the different parts of the HCAL, these were exposed to beams of electrons, pions, protons and muons. An ECAL module was also included in the test beam setup. The hadronic energy resolution of the barrel HCAL and ECAL combination is parameterized as $\sigma^2/E^2 = a^2/E + b^2$, where a corresponds to a stochastic term and b to a constant term. The energy resolution in the endcaps is similar to that of the barrel.

2.3.5 The Muon System

The CMS muon system(38) is dedicated to the identification of high p_T muons, in combination with the tracker. The system is placed outside the magnetic coil, embedded in the return yoke, in order to exploit the fully returning flux of about 1.8 T.

The system consists of three independent subsystems (figure **2.13**):
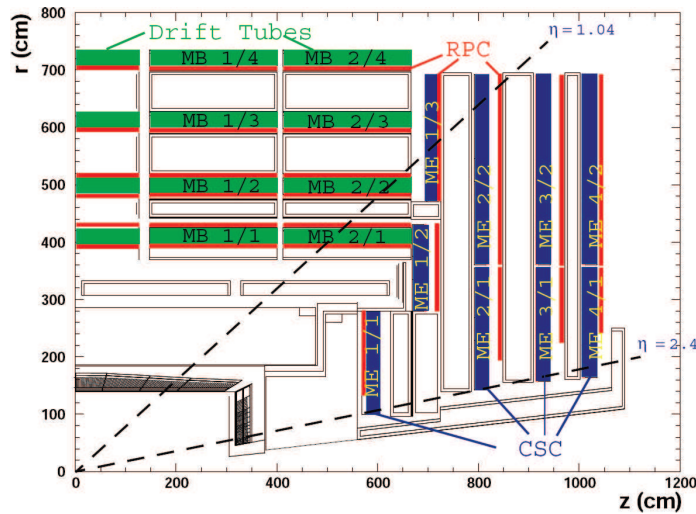
Figure 2.13: A quadrant of the CMS muon system.

- Drift Tubes (DT) are placed in the barrel, a region where the hits occupancy is relatively low (< 10 Hz/cm^2);

- Cathode Strip Chambers (CSC) are located in the endcaps, where the occupancy is much higher (> 100 Hz/cm^2);

- Resistive Plate Chambers (RPC) placed both in the barrel and in the endcaps regions.
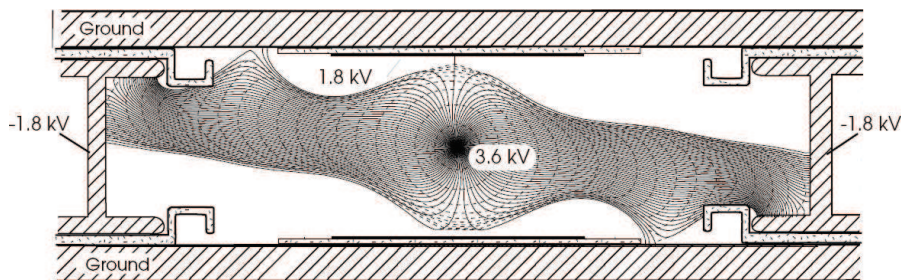
Figure 2.14: A schematic representation of a drift tube chamber. The drift lines in presence of the magnetic field are shown.

The Drift Tube system is made of chambers consisting of twelve layers of drift tubes. Each group of layers is packed in three independent substructures called

super-layers, for a total of four chambers with three super-layers per chamber. In each chamber two super-layers have anode wires parallel to the beam axis, and one has perpendicular wires. Thus, each chamber can provide two measurements of the $r\phi$ coordinate and one measurement of the z coordinate of the track hit positions. The position resolution is about 100 μm in both $r\phi$ and z. In figure **2.14** a representation of a drift tube chamber is shown. Each chamber is filled with a gas mixture of Ar(85%) and CO_2(15%).

The figure (2.15) shows the structure of the Cathode Strip Chambers. These are multi-wire proportional chambers with segmented cathodes. Each chamber can provide both hit position coordinates. Chambers are filled with a gas mixture of Ar(40%), CO_2(50%), CF_4(10%). The chamber spatial resolution is about 80-85 μm.

radial strips   wires
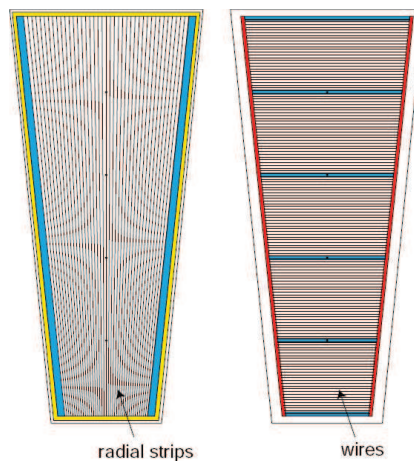
Figure 2.15: A schematic representation of the CSC cathode panel (on the left) and anod panel (on the right).

A schematic view of the Resistive Plate Chambers is shown in figure **2.16**. These are made of parallel bakelite planes, with a bulk resistivity of $10^{10} \div 10^{11}$ Ωcm. These are filled with a mixture of $C_2H_2F_4$ (96.2%), iso-C_4H_{10} (3.5%) and SF_6 plus water vapour (0.3%). They operate in avalanche mode. Those chambers have limited spatial resolution, but since they have excellent timing performances they are used for bunch crossing identification and even more for trigger purposes.

Another crucial aspect of the CMS detector is its trigger system. This argument will be explained in more detail in the next chapter since a specific trigger path has
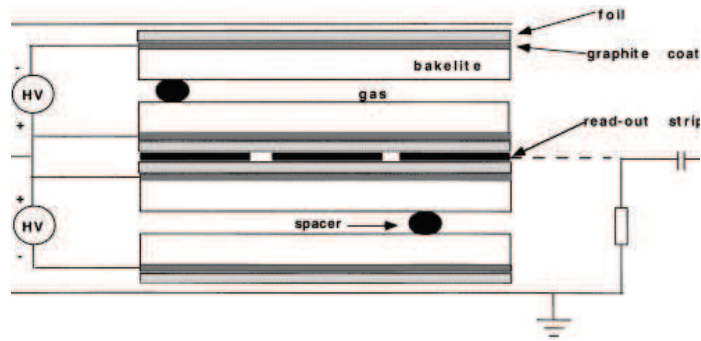
Figure 2.16: Schematic view of the RPC double-gap structure. The read-out strips in the Barrel chambers run along the beam direction.

been designed for the $t\bar{t} \rightarrow \tau_h +$ jets analysis in the context of this thesis.

Chapter 3

The CMS Trigger System

3.1 Introduction

The proton bunches at LHC cross at a rate of ~ 40 MHz. In real life it is impossible to register the full rate of events, due to the limited disk space and the limited speed of the readout electronics. Besides that, as we can see from figure **2.2**, the rate of interesting events is smaller by order of magnitudes than the total interaction rate. At a luminosity of $1 \cdot 10^{33}$ cm^{-2}s^{-1} for example, we expect a rate of 10^9 events with low momentum transfer, called minimum bias events, in which we are not interested. Hence a trigger system is built in order to reject by a factor 10^7 the collisions and to select in a short time the interesting physics events with high efficiency. To meet these goals the trigger system in CMS is divided in two main steps:

- Level-1 trigger (L1);

- High Level Trigger (HLT).

In figure **3.1** a schematic view of the CMS trigger system is presented. The L1 trigger runs on dedicated processors, and accesses coarse level granularity information from the calorimeter and muon systems. A L1 trigger decision has to be taken for each bunch crossing within 3.2 μs. The L1 trigger task is to reduce the flux of data from 40 MHz to 100 kHz. The High Level Trigger instead, should reduce again the L1 output rate down to the nominal rate of 100 Hz. In practice, the rate of events stored by the HLT is higher. During the 2011 data taking the output rate of the HLT was around 300 Hz in average.

The HLT code runs on a farm of commercial processors and can access the full granularity of all subdetectors. In the following more details of the main characteristics of the CMS trigger system are presented.
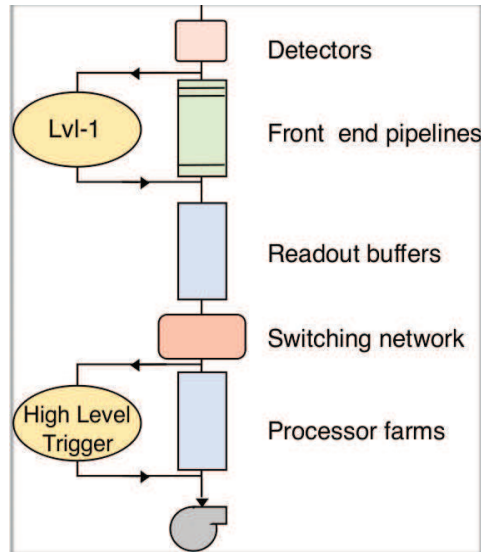
Figure 3.1: A schematic representation of the CMS trigger system.

3.2 L1 Trigger

The L1 trigger (39) is responsible for the identification of electrons, muons, photons, jets and missing transverse energy. It has to have an high and carefully understood efficiency. Its output rate and speed are limited by the readout electronics and by the performances of the Data Acquisition (DAQ) system. It consists of three main subsystems:

- L1 Calorimeter Trigger;

- L1 Muon Trigger;

- L1 Global Trigger.

The L1 Global Trigger is responsible for combining the output of the L1 Calorimeter Trigger and L1 Muon Trigger and for making the decision. The L1 Muon Trigger is actually a composed system itself: information from RPC, CSC and DT specific triggers are combined in the so called L1 Global Muon Trigger. The organization of the CMS L1 Trigger is summarized in figure **3.2**.

3.2.1 The Calorimeter trigger

The calorimeter Trigger uses $(0.35\eta\times0.35\phi)$ trigger tower energy sums formed by the ECAL, HCAL and HF upper level readout Trigger Primitive Generator (TPG)
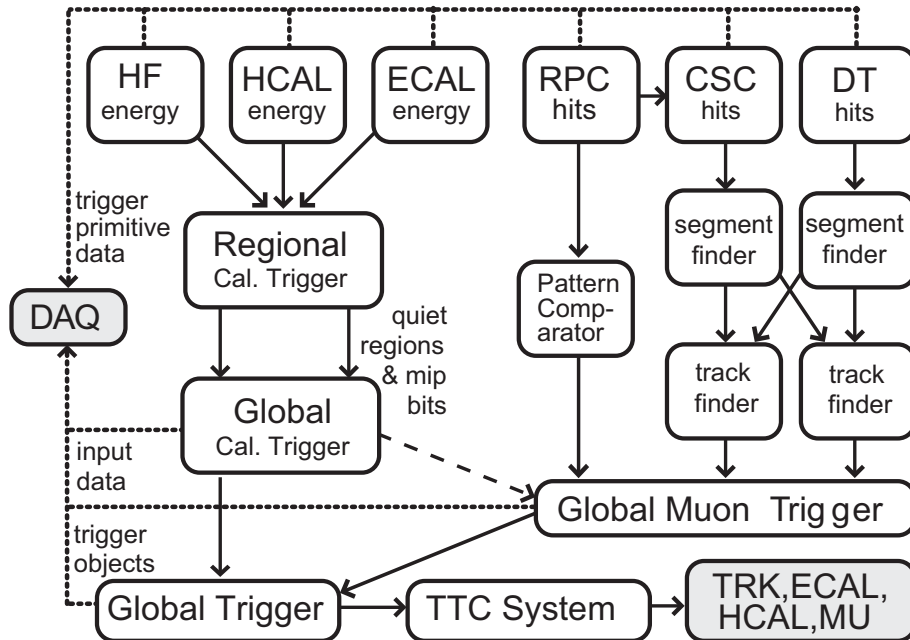
Figure 3.2: Overview of the L1 trigger system.

circuits from the individual calorimeter cell energies. The TPG information is then transmitted to the Regional Calorimeter Trigger (RCT). The RCT finds electron, photon, tau and jet (isolated or not) candidates, and it transmits the candidates along with the sums of transverse energy to the Global Calorimeter Trigger (GCT). The GCT sorts the electron, photon, tau, and jet candidates and forwards the top four candidates of each type to the Global Trigger (GT). Then the GCT calculates the total transverse energy (E_T defined as the vectorial sum of the energy in the transverse plane (x,y) of the electron, photon, tau, and jet candidates) and total missing energy vector ($\overrightarrow{E_T^{miss}}$). It transmits this information to the GT as well. The RCT also transmits an (η, ϕ) grid of quiet regions to the global muon trigger to define muon isolation cuts.

3.2.2 Electron and photon triggers

In each calorimeter region (4×4 trigger towers) the highest isolated and non-isolated E_T electron/photon candidates are separately found. Only the top four candidates (isolated or not), are retained to be processed by the CMS Global Trigger.

The isolation and shower shape trigger cuts are programmable and can be adjusted to the running conditions. For example, at high luminosity the isolation cuts can be relaxed to take into account higher pile-up energies. It is also possible to define different trigger conditions for different rapidity regions.

To connect the L1 threshold to an effective requirement on the electron transverse momentum, the electron p_T at which the L1 Trigger is 95% efficient is determined as a function of the L1 threshold. From this result, the rate for electron/photon triggers as a function of the effective cut on E_T, can be computed. Double, triple and quadruple electron/photon triggers can be defined. Different energy thresholds and different cuts on η, ϕ for different objects can be defined.

3.2.3   Jet and τ triggers

The jet trigger starts by computing the transverse energy sums (electromagnetic and hadronic) in calorimeter regions defined by 4×4 trigger towers, except in the HF region where single trigger towers are used.

The jet trigger uses a 3×3 calorimeter region sliding window technique capable to perfectly cover the entire CMS calorimeter (η, ϕ) range (figure **3.3**). The E_T of the central region (3×3 cells), is required to be higher than the E_T of each of the eight neighbors. In addition, the central region E_T is required to be greater than a fixed value, 5 GeV, to suppress spurious soft deposits. The jets are labelled by (η, ϕ) indexes of the central calorimeter region.

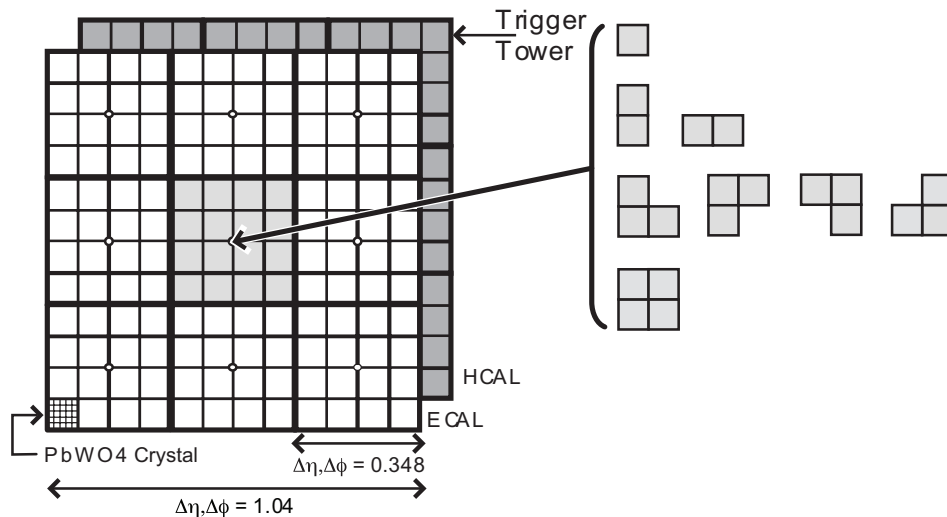**Figure 3.3:** Jet and τ trigger algorithms. The calorimeter is divided into 3×3 regions, made of 12×12 trigger towers. The entire calorimeter is scanned for calorimeter regions where the central region E_T is greater than the E_T of the eight neighbors. The τ jets are identified if the central region deposits shows one of the 8 distinctive patterns presented in the right part of the figure.

Single and three-prong decays of τ leptons form energy deposits in the calorimeter with a specific shape as shown on figure **3.3**. Since the τ decays involve charged pions which deposit their energy in the hadron calorimeter, the electron/photon trigger does not capture them. The transverse profile of the active tower is analyzed to tag the jets as taus candidate. A jet is defined as a τ if the active tower in the central region verifies at least one of the 8 patterns shown in figure **3.3**. Jets and taus occurring in a calorimeter region where an electron is identified are not considered.

In addition, counters of the number of jets above programmable thresholds are implemented in various η regions in order to trigger events with a large number of low energy jets. Jets in the forward and backward HF calorimeters are sorted and counted separately. This separation is a safety measure to mask the high η region, more background sensitive, from the central η region, although the central and forward jets are sorted and tracked separately through the trigger system.

3.2.4 H_T and E_T^{miss} triggers

The E_T triggers use the transverse energy sums (electromagnetic and hadronic) computed in calorimeter regions (4×4 trigger towers in barrel and endcap). E_x and E_y are computed from E_T using the coordinates of the calorimeter region center.

The H_T trigger is defined as the scalar sum of the E_T of jets above a given threshold with a typical value of jet $E_T > 10$ GeV. This trigger is not as sensitive as the total E_T, given by the sum of the calorimeter regions E_T deposits, to both noise and pileup effects. The H_T trigger can capture high jet multiplicity events such as those from fully hadronic top decays, hadronic decays of squarks and gluinos. Even if these events have several hundred GeV of energy, they can fail the jet triggers since the E_T of a single jet is lower than the defined threshold.

3.2.5 L1 muon trigger

The structure of the muon trigger is shown in figure **3.4**.

The RPC trigger electronics builds track segments and provides an estimation of the muon p_T. Then it transmits the information to the Global Muon Trigger. It also provides information to the CSC logic unit in order to solve hit position ambiguities, in case two or more muon tracks cross the same CSC chamber.

The CSC trigger builds Local Charged Tracks (LCT), i.e track segments made out of the cathode strips only. A p_T value and a quality flag are assigned to the LCTs. From each sector of the nine CSC chambers, the three top LCTs are kept and passed to the CSC Track Finder. It uses the full CSC information to build tracks, to assign them a p_T and a quality flag. At the end, the information from the CSC Track Finder are addressed to the Global Muon Trigger.
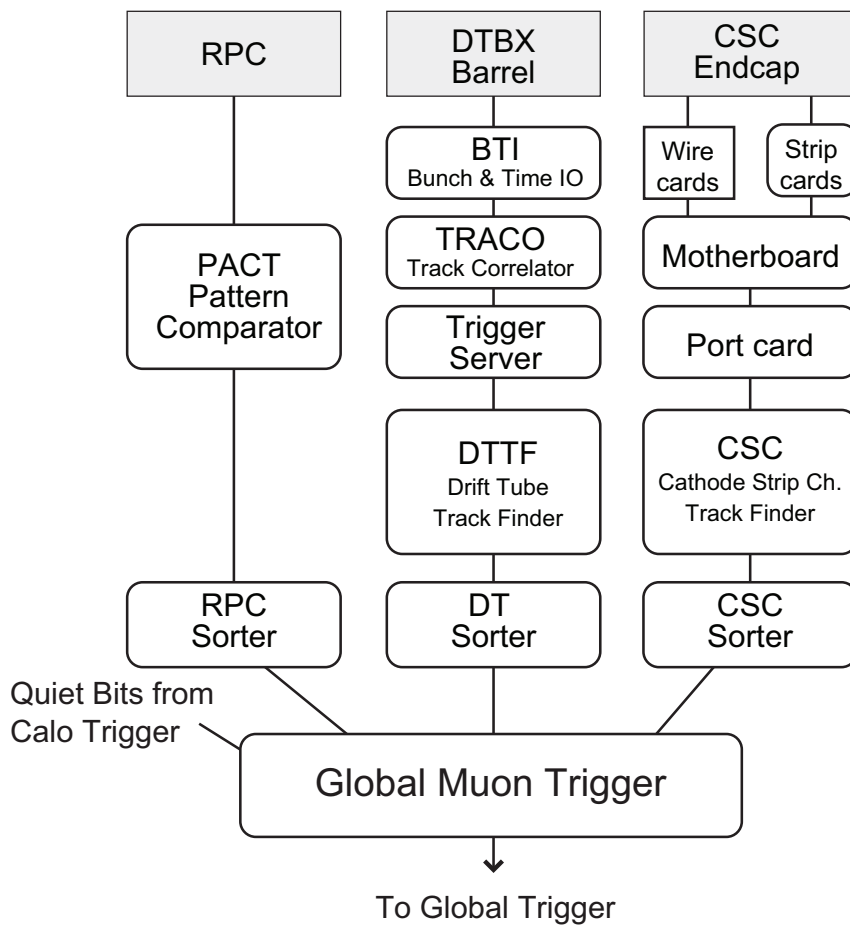
Figure 3.4: Block diagram of the L1 muon trigger.

DTs are equipped with Track Identifier electronics, which is able to find groups of aligned hits in the four chambers of a super-layer. Segments from two super-layers are then combined by the DT Track Correlator, measuring the ϕ coordinate. The best two segments are sent to the DT Track Finder that builds tracks and sends them to the Global Muon Trigger.

At the end the Global Muon Trigger sorts the RPC, CSC and DT muon tracks and tries to combine them. The final set of muons is ordered according to their quality. Only the best four tracks are passed to the L1 Global Trigger.

3.2.6 Global trigger

The Global Trigger accepts muon and calorimeter trigger information, synchronizes all information arriving from the different sub-systems and communicates the L1 decision to the timing in order to start the readout. The global trigger decision is made using logical combinations of the trigger data from the Calorimeter and Muon Global Triggers.

The L1 Trigger system sorts ranked trigger objects which are accompanied by their coordinates in the (η, ϕ) space. This allows the Global Trigger to vary thresholds based on the location of the trigger objects. It also allows the Global Trigger to require trigger objects to be close or opposite from each other. In addition, the presence of the trigger object coordinate permits a quick determination of the regions of interest where a more detailed HLT analysis will run. The Global L1 Trigger transmits a decision to either accept (L1A) or reject each bunch crossing. This decision is transmitted through the Timing Trigger and Control system (TTC) and to the Trigger Throttle System (TTS), that allows the reduction by prescaling or shutting off of L1A signals in case the detector readout is close to overflow.

3.2.7 Overview of the $1 \cdot 10^{33}$ cm^{-2}s^{-1} L1 menu

From the physics point of view we want the L1 trigger system to be able to select leptons and jets over the pseudorapidity range $|\eta| <2.5$, with high efficiencies, above selected thresholds in transverse momentum. The main feature of the L1 seeds is that they should be as general as possible in order to serve as many HLT paths as possible. They should contain low thresholds on the physics objects in order to not limit the efficiency of the final HLT path. Table **3.1** summarizes the L1 seeds used for the 2011 data taking for a luminosity of $1 \cdot 10^{33}$ cm^{-2}s^{-1}.

3.3 High Level Trigger

The main goal of the HLT is to reduce the L1 output rate until 100-(300) Hz. Only the HLT selected events will be definitively written to disk, while the other ones are lost forever. Hence it is clear that the correct functioning of the HLT is a crucial

Classification	L1 seed	prescale factor	rate (Hz)
JETS	L1_SingleJet52	1	11678
	L1_SingleJet68	1	3859
	L1_DoubleJet36	1	7279
	L1_DoubleJet52	1	2369
	L1_TripleJet28_Cen	1	1758
	L1_QuadJet20	1	644
	L1_SingleTauJet52	1	5724
	L1_SingleTauJet68	1	1590
	L1_SingleTauJet80	1	692
	L1_DoubleTauJet28	1	4983
	L1_DoubleTauJet40	1	1341
ELECTRONS	L1_SingleEG12	1	12778
	L1_SingleIsoEG12_Eta2p17	1	8626
	L1_DoubleEG8	1	1934
	L1_DoubleEG_12_5	1	1586
MUONS	L1_SingleMu7	1	5889
	L1_SingleMu7_Eta2p1	1	5456
	L1_DoubleMu0	1	2889
E_T^{miss}	L1_MET20	20	504
	L1_MET30	1	1312
	L1_HTT50	1	3873
cross seeds	L1_EG5_Jet36_deltaphi1	20	594
	L1_EG8_Jet20_Central_deltaphi1	20	503
	L1_EG12_TauJet20_deltaphi1	1	2793
	L1_Mu3_Jet16_Central	1	3693
	L1_Mu7_EG5	1	865
	L1_Mu12_EG5	1	357

Table 3.1: Main L1 seeds used for physics analyses in the $1 \cdot 10^{33}$ cm^{-2}s^{-1} menu deployed during the 2011 data taking. The corresponding rate and prescale factors are presented.

Table 3.2: Dedicated trigger space disk for the different physics topics

Physics topics	trigger space in %
Higgs physics	24%
Beyond SM	41%
B-quark physics	7%
QCD	5%
Electronweak	9%
Top Physics	12%
Forward physics	2%

aspect of the CMS physics program. In particular the main requests from a physics point of view are:

- The HLT selection has to cover all analyses that are part of the CMS physics program without passing the total rate of acceptable events of 100-(300) Hz. The triggers dedicated to the different physics topics are roughtly distributed as detailed in table **3.2**;

- As for the L1 trigger, the efficiency for the physics objects (electrons, muons, jets, taus...) must be as high as possible;

- The selection must be as inclusive as possible. Since the LHC represents a new energy frontier and since we look for new physics phenomena, the HLT selection has to keep potential exotic events;

- The final selection of events should include data allowing to study the trigger and offline reconstruction efficiencies;

- The entire HLT code should be continuously monitored;

- The HLT should include all major improvements coming from the offline reconstruction. For example, until the end of the 2011, the reconstruction of HLT jets only relied on the calorimeter information. Starting from the 2012 data taking, the reconstruction of the particle flow jets (PF jets) will be possible, thanks to the latest improvements in the timing of the particle flow algorithm. The aim of the particle flow technique is to reconstruct the particles using the information coming from all CMS subdetectors. This technique will be described in more details in chapter 4.

In order to process events efficiently, the HLT code has to reject not interesting events as soon as possible; computationally expensive algorithms must be ran only on good candidate events. In order to meet this requirement the HLT code is organized in three virtual levels:

- Level 2 (L2): uses only muon and calorimetry information;

- Level 2.5 (L2.5): uses also the pixel information;

- Level 3 (L3): uses the full information from all the tracking detectors.

Each step reduces the number of events to be processed in the next step. The most computationally expensive tasks are executed at L3; time consuming algorithms such as track reconstruction are only executed in the regions of the detector where good candidates from L1, L2, L2.5 are found. Besides that, since a high precision is not required at HLT, the track reconstruction is performed on a limited set of hits, and is stopped when the required resolution is achieved.

3.3.1 Electron and photon identification

Two complementary algorithms are defined for electron and photon reconstruction, using different seeding strategies:

- 'tracker driven' seeding, used for low p_T electrons and for electrons inside jets;

- 'ECAL driven' seeding that starts by the reconstruction of ECAL 'superclusters'[1] with $E_T > 4$ GeV. It is optimized for isolated electrons in the p_T range relevant for Z or W decays down to $p_T \simeq 5$ GeV.

At HLT, electron and photon selection proceeds by requiring a supercluster with E_T above a given threshold matching an electromagnetic L1 candidate. The HLT runs the standard ECAL superclustering algorithm with almost the same configuration than the offline reconstruction. The electron paths also require an hit in the pixel layers of the CMS detector compatible with an electron trajectory. The matching requirements are looser with respect to the offline requirements in order not to compromise the offline analysis.

Since there are no other HLT requirements except that an online supercluster with E_T >15 GeV should match a L1 candidate, the HLT component of the trigger efficiency is close to 100% for photon triggers. The online pixel matching requirement of the electron triggers is close to 100% efficient with respect to the offline pixel-matching requirement up to $|\eta| < 2.0$.

3.3.2 Jets and missing energy

Jet reconstruction at HLT level is based on calorimeter jets, the best compromise between resolution and timing for the reconstruction. While during the 2010 data taking, uncorrected E_T of the jets were used at trigger level, from the beginning of

---

[1]With the name of 'supercluster' we usually define a group of one or more associated clusters of energy deposits found in the ECAL.

2011 corrected jets were used. The correction factors were derived from simulated events.

The missing energy, E_T^{miss}, is generally calculated as the magnitude of the negative vector sum of the momentum transverse to the beam axis of all final state particles reconstructed in the detector.

While jets at HLT level are reconstructed using calorimeter information only, the improvements done on the particle flow timing have allowed to use the particle flow algorithm to reconstruct the HLT E_T^{miss} with a consequent increase of the performances of all paths including E_T^{miss}.

3.3.3 Muon reconstruction

The HLT muon trigger follows an outside-in logic. It starts from an L1 trigger object used as a seed to reconstruct a standalone-muon track in the muon system. Then a full track fit is performed to improve the estimation of the p_T. At this point, a filter on p_T threshold is applied to the standalone (also called L2) muon. Then seeds in the silicon tracker are generated in the region around the L2 muon, and tracker trajectories are reconstructed. If this track matches a L2 muon, a global fit combining tracker and muon hits is performed. The output of the fit is a L3 muon track on which the final p_T requirement is applied.

3.3.4 Tau reconstruction

The tau identification at trigger level is different for L1 and HTL. At L1 taus are reconstructed using the information from the calorimeter only as explained in section **3.2.3**. The tau reconstruction at HTL instead, is based on the particle flow algorithm. At HLT level the reconstruction proceeds as follows:

- L2: jets and tau jets coming from the L1 output are matched with an HLT jet within a cone of radius ΔR<0.2 for $|\eta| < 3$;

- L2.5: on the matched object the particle flow algorithm is ran. The leading track of the particle flow tau has to be higher than a given threshold, usually set at 5-10 GeV;

- L3: the L2.5 objects are required to be isolated. Usually no tracks and photons with, respectively, p_T >1.5 GeV, and E_T >1.5 GeV are accepted in an isolation cone of radius ΔR=0.5 surrounding the signal cone of radius 0.15.

More details about the particle flow tau reconstruction will be given in section **4.2** of the next chapter.

3.3.5 Overview of the $5 \cdot 10^{32}$-$1 \cdot 10^{33}$ cm^{-2}s^{-1} HLT menus

As we saw in the table **3.2** a large bandwith is given to the Beyond SM (BSM) and SM Higgs searches without leaving uncovered the other SM measurements. In the following of this section we intend to present the main features of the triggers used for the different physics analyses in the main menus ($5 \cdot 10^{32}$, $1 \cdot 10^{33}$ cm^{-2}s^{-1}) deployed during the 2011 data taking. For all analyses the main idea is to combine different objects in order to keep under control the rate and the pileup effect.

Regarding the BSM analyses and especially those looking for supersymmetry (SUSY) the main goal is to cover as much as possible of the SUSY phase space. The SUSY events are characterized by a wide variety of objects in the final state: leptons, jets, and generally by the presence of E_T^{miss}. For purely hadronic analyses, the hadronic activity is charactarized using the H$_T$ variable (the scalar sum of the transverse energy of the jets) combined with the presence of E_T^{miss}. Another important variable used is M$_{eff}$, defined as the scalar sum of H$_T$ and H$_T^{miss}$. It is used in order to enlarge the accessible phase-space at low value of E_T^{miss} and H$_T$. For photon analyses, instead, double photon triggers or combined triggers of single photon and H$_T$ are used. For SUSY leptonic final states, three classes of triggers mainly exist: single-lepton+H$_T$, double-lepton, and multi-lepton triggers. In table **3.3** the main triggers used during the 2011 data taking for SUSY analyses are summarized.

Regarding the SM Higgs boson search, it is important to stress that there are two production ways of the SM Higgs boson. The first one is by gluon-gluon fusion (40), the second one by vector boson fusion (VBF) (41). The cross section at LHC for the first kind of production is higher than the second one, but the VBF production has a distinctive signature due to the presence of two jets in the forward pseudorapidity region of the detector. A dedicated trigger strategy based on the identification of forward jets can lead to an important improvement of the trigger efficiency and compensate for the lower cross section. This feature is exploited for the VBF H$\rightarrow$ $WW \rightarrow e + \nu_e + jets$ channel. The trigger for this analysis contains the HLT identification of two well separated jets in pseudorapidity ($\Delta\eta > 2$). The triggers used with the corresponding L1 seeds and rates for the $5 \cdot 10^{32}$ cm^{-2}s^{-1} menu are presented in table **3.4**. Another important channel for the VBF Higgs boson search is the τ + lepton final state. In table **3.5** the trigger used for the decays of the VBF Higgs boson in lepton+τ final state are presented.

About the Higgs boson produced via gluon-gluon fusion, instead, the main analyses done are relative to the diphoton final state and the decay of the Higgs boson in two W(Z) bosons which decay in leptonic final states. The trigger used for the H$\rightarrow$ $\gamma\gamma$ relies on a L1_SingleEG12 seed and requires two identified photons at HTL level with thresholds of 26 and 18 GeV respectively. For the analyses of H$\rightarrow WW(ZZ) \rightarrow l\nu + l\nu(l_1l_2 + l_3l_4)$ the trigger used relies on the presence of single

Trigger name	rate (Hz)
HADRONIC SUSY ANALYSES	
HT260_MHT60	$\sim 10-20$
Meff520	$\sim 10-20$
Photon32_Photon26	~ 5
Photon70_HT200	~ 7
Photon70_MHT30	~ 5
LEPTONIC SUSY ANALYSES	
Ele10_HT200	8
Mu5_HT200	5
Ele5_HT200_pfMHT30	< 1
Mu5_HT200_pfMHT30	< 1
Tau10_HT200_pfMHT35	< 1
Ele17_Ele8	~ 10
DoubleMu6	~ 10
Mu17_Ele8	~ 2
Ele17_Mu8	~ 4
DoubleMu3_HT160	~ 2
DoubleEle8_HT160	~ 2
Mu3_Ele8_HT160	~ 3
TripleMu5	< 1
DoubleMu5_Ele8	< 1
Mu3_DoubleEle8	~ 3
TripleEle10	~ 1

Table 3.3: Trigger rates for SUSY searches at a luminosity of $5 \cdot 10^{32}$ cm^{-2}s^{-1}.

Trigger name	L1 seed	rate (Hz)
HLT_Ele15_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT_CleanDiJet_35_20_DeltaEta2	L1_SingleEG12	3.9

Table 3.4: Trigger used for the VBF H$\to WW \to e + \nu_e + jets$ analysis with the corresponding rate at a luminosity of $5 \cdot 10^{32}$ cm^{-2}s^{-1}.

Trigger name	L1 seed	rate (Hz)
HLT_Ele15_Id_Iso_LooseIsoPFTau15	L1_SingleEG12	11.9
HLT_Ele15_Id_Iso_LooseIsoPFTau20	L1_SingleEG12	7.71
HLT_Ele18_Id_Iso_LooseIsoPFTau15	L1_SingleEG12	9.23
HLT_Ele18_Id_Iso_LooseIsoPFTau20	L1_SingleEG12	6.20
HLT_IsoMu12_LooseIsoPFTau10	L1_Mu7	10.85
HLT_Mu15_LooseIsoPFTau20	L1_Mu12	13.68
HLT_IsoMu15_LooseIsoPFTau15	L1_Mu12	3.13

Table 3.5: HLT paths used for the SM Higgs search in the lepton $+ \tau$ final state, with the corresponding rates for a luminosity of $1 \cdot 10^{33}$ cm^{-2}s^{-1}.

Trigger name	L1 seed	rate (Hz)
HLT_Ele17_CalIdL_CalIsoVL_Ele8_alIdL_CalIsoVL	L1_SingleEG12	5.5
HLT_DoubleMu7	L1_DoubleMu3	3.7
HLT_Mu17_Ele8_CaloIdL	L1_Mu3_EG5	3.6
HLT_Mu8_Ele17_CaloIdL	L1_Mu3_EG5	1.2

Table 3.6: HLT paths used for H$\rightarrow WW \rightarrow l\nu + l\nu$ and H$\rightarrow ZZ \rightarrow l_1 l_2 + l_3 l_4$ analyses with the corresponding rates for a luminosity of $5 \cdot 10^{32}$ cm^{-2}s^{-1}.

or double leptons at L1 (L1_EG12, L1_EG12_EG5, L1_Mu3_EG5) and on the presence of two leptons at HLT. In table **3.6** an overview of the HLT paths used for these analyses with the corresponding L1 seeds and rates for a luminosity of $5 \cdot 10^{32}$ cm^{-2}s^{-1} are presented.

Finally the main paths used for top physics are presented. To record dilepton final state events, the same HLT paths as the ones used for the dilepton final state of the SM Higgs boson produced via gluon-gluon fusion, are used. For the semileptonic final states, combined triggers requiring the presence of jets and of a lepton are used. The b-jet identification is exploited at the HLT in order to keep the rate under control in particular in presence of pileup. The fully hadronic channel is triggered using a multijet path including b-tagging. The trigger used to record $t\bar{t} \rightarrow \tau_h + jets$ events, HLT_QuadJet40_IsoPFTau40, has been developed in the context of this thesis. In the following of this chapter we will see in more details how this path was built and tested before to be inserted in the official CMS trigger menu.

3.4 Development of the HLT_QuadJet40_IsoPFTau40 trigger

The design of the multijet with tau identification trigger was made in order to cope with the high number of expected pileup events at the LHC and in order to keep the rate below 10 Hz. The general configuration of the studied trigger requires the presence of four jets at HLT level, one of them being matched to an identified hadronic

HLT path	L1 seed	rate (Hz) at $5 \cdot 10^{32}$	involved analysis
HLT_Mu17-TripleCentralJet30	L1_Mu10	2.3	$t\bar{t} \to \mu$+jets
HLT_IsoMu17_BTagIP_CentJet40	L1_Mu10	0.22	$t\bar{t} \to \mu$+jets
HLT_Ele25-CaloIdVT_TrkIdT_TripleCentJet30	L1_SingleIsoEG12	4.5	$t\bar{t} \to e^-$+jets
HLT_QuadJet40_IsoPFTau40	L1_QuadCenJet20	4.5	$t\bar{t} \to \tau$+jets
HLT_QuadJet50_Jet40_BtagIP	L1_QuadCenJet20	2.1	$t\bar{t}$ full hadronic channel

Table 3.7: HLT paths used for top physics for a luminosity of $5 \cdot 10^{32}$ cm^{-2}s^{-1} with the corresponding rates.

tau (with $\Delta R < 0.4$). All jets are required to be inside the tracker acceptance, $|\eta| < 2.5$, since the matching to a tau requires the reconstruction of the tau leading track. The main parameters which were studied are the number and the transverse momentum threshold of the jets as well as the tau isolation.

For trigger studies the CMS collaboration developed a specific trigger emulator, called *openHLT*, which allows to simulate the trigger design and compute the expected rates for the different expected luminosities. This package was used for the studies presented in the following part of this chapter.

3.4.1 L1 seed

The chosen L1 seed relies on the presence of four central jets with a pseudorapitidy $|\eta| < 3$ and corrected jet transverse momenta above 20 GeV at a luminosity of $5 \cdot 10^{32}$ cm^{-2}s^{-1} : L1_QuadJetCen20. The choice of a multijet L1 seed was made in order to keep the threshold on the tau transverse momentum low, the only available tau L1 seed, L1_SingleTau52, being fully efficient only above 65 GeV. The rate for the L1_QuadJetCen20 at $5 \cdot 10^{32}$ cm^{-2}s^{-1} was measured to be 16 kHz.

3.4.2 HLT paths

The trigger used in the $5 \cdot 10^{32}$ cm^{-2}s^{-1} and $1 \cdot 10^{33}$ cm^{-2}s^{-1} HLT menu is HLT_QuadJet40_IsoPFTau40. For the $2 \cdot 10^{33}$ cm^{-2}s^{-1} menu the thresholds on the momenta of the jets and particle flow tau needed to be raised to 45 GeV. In the late part of the $2 \cdot 10^{33}$ cm^{-2}s^{-1} period the L1 thresholds were moved to 28 GeV. The HLT reconstruction sequence is based on the following parameters :

- jets : 4 central jets in $|\eta| < 2.5$ and $p_T > 40$ GeV.

- tau leg :

  - L2 jet made with a radius of 0.2 matching the L1 jet object, $E_T > 40$ GeV, $|\eta| < 2.5$;

  - L2.5: fixed cone particle flow tau matching the L2 jet object, tau $p_T > 40$ GeV, $|\eta| < 2.5$;

- L2.5: leading track finding, leading track $p_T > 5$ GeV;

- L3: tau isolation, with matching cone 0.2, signal cone 0.15 and isolation cone 0.5. The maximum track p_T is set to 1 GeV, and maximum gamma E_T to 1.5 GeV.

Table **3.8** summarizes the expected L2 and L2.5 rates for a luminosity of $5 \cdot 10^{32}$ $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ for the different studied trigger configurations (including HLT_QuadJet40_IsoPFTau40).

	expected rates at $5 \cdot 10^{32}$ (Hz)	
Trigger Name	L2	L2-5
QuadJet20_PFTau20_eta2.5	94.9±14.3	77.6±13.
QuadJet30_PFTau30_eta2.5	25.9±7.5	19.4±6.5
QuadJet40_PFTau40_eta2.5	**8.6±4.3**	**6.5 ± 3.7**

Table 3.8: Expected L2 and L2.5 rates at a luminosity of $5 \cdot 10^{32}$ $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ for the investigated trigger configurations.

Tables **3.9** and **3.10** summarize the expected final HLT rates at a luminosity of $5 \cdot 10^{32}$, $1 \cdot 10^{33}$ and $2 \cdot 10^{33}$ $\mathrm{cm}^{-2}\mathrm{s}^{-1}$.

Trigger Name	rate at $5 \cdot 10^{32}$ (Hz)
QuadJet30_IsoPFTau30_eta2.5	18.1±0.6
QuadJet35_IsoPFTau35_eta2.5	9.7±0.5
QuadJet40_IsoPFTau40_eta2.5	**4.5 ± 0.3**
QuadJet45_IsoPFTau45_eta2.5	2.3±0.2
QuadJet40_IsoPFTau_eta3	7.0±0.4
QuadJet40_IsoPFTau_eta2	2.8±0.2
QuadJet40_IsoPFTau_eta1.5	1.09±0.16

Table 3.9: Estimated trigger rates for the investigated triggers for the $5 \cdot 10^{32}$ $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ menu.

Trigger Name	rate at $1 \cdot 10^{33}$ (Hz)	rate at $2 \cdot 10^{33}$ (Hz)
QuadJet40_IsoPFTau40_eta2.5	2.85 ± 0.02	5.06± 0.03
QuadJet45_IsoPFTau45_eta2.5	**5.72 ± 0.05**	10.13 ± 0.06

Table 3.10: Estimated trigger rates for the investigated triggers for the $1 \cdot 10^{33}$ and $2 \cdot 10^{33}$ menus.

Figure **3.5** shows the final trigger rate dependency versus the transverse momentum threshold and η range of the HLT jets. Isolation1 and Isolation2 refer to two

possible tau isolations. Isolation1 is defined as follows: signal cone R=0.2, isolation cone R=0.3 and no track and photon with $p_\mathrm{T}(E_T) > 1(1)$ GeV in the isolation annulus are accepted. Isolation2, instead is defined by a signal cone of R=0.15 an isolation cone R=0.5 and no track and photon with $p_\mathrm{T}(E_T) > 1(1.5)$ GeV in the isolation annulus are accepted. Isolation2 is the one chosen for the $5 \cdot 10^{32}$ cm^{-2}s^{-1} menu.

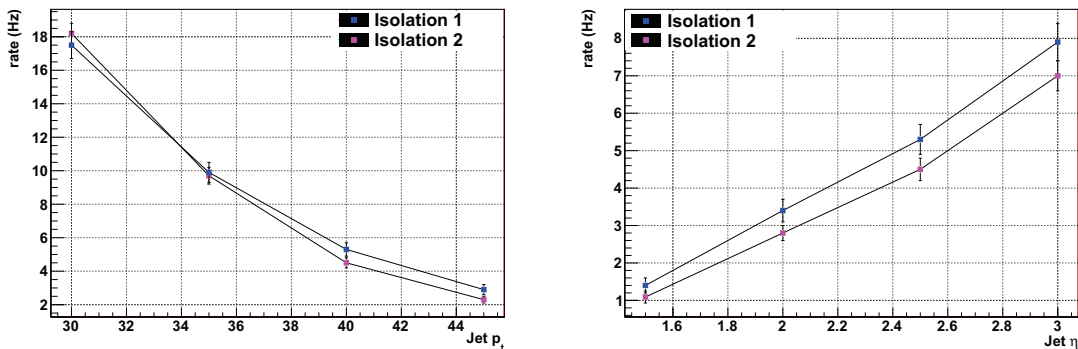**Figure 3.5:** Estimated trigger rates ($5 \cdot 10^{32}$ cm^{-2}s^{-1}) versus the transverse momentum threshold and η range (for $p_T = 40$ GeV) of the selected HLT jets.

3.4.3 Expected signal efficiency

A preliminary estimation of the trigger efficiency was done using a simulated sample of semileptonic hadronic tau top-antitop events. The following offline criteria were applied to mimic the offline analysis :

- 4 PF jets with $p_T > 30$ GeV and $|\eta| < 2.5$;

- 1 PF tau reconstructed with the shrinking-cone algorithm (see section **4.3.2.1** of the next chapter), $p_T > 30$ GeV, $|\eta| < 2.5$;

- specific criteria on the tau isolation as well as on the electrons and muons identification were applied in order to remove electrons and muons faking taus (section **4.3.2.1**).

The estimated efficiencies are summarized in table **3.9** for the different investigated trigger configurations. The trigger efficiency was also estimated on the charged Higgs signal since the trigger was implemented to be shared with the $H^\pm \to \tau \nu_\tau$ analysis (42).

Figures **3.6**-**3.8** show the efficiencies of finding an HLT PF tau in semileptonic hadronic top-antitop events. On figure **3.7** the presence of a leading track with $p_T > 5$ GeV is required. On figure **3.8** the HLT tau is in addition requested to be isolated. No preselection on the number of reconstructed jets is applied.
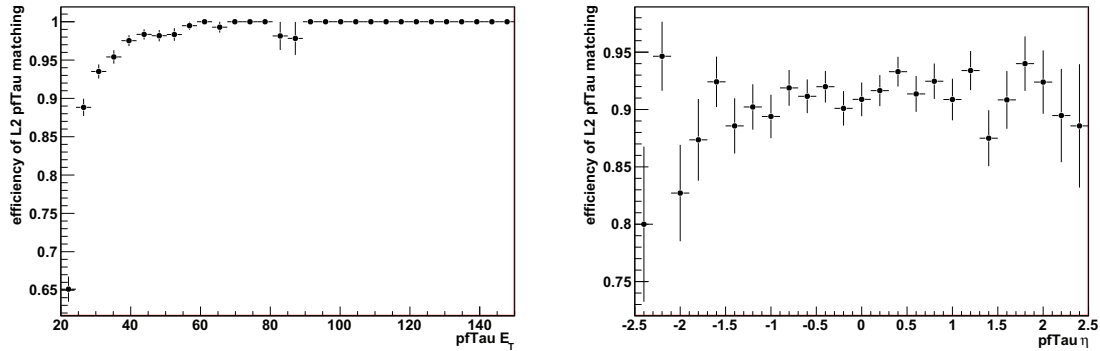
Figure 3.6: Efficiency of finding an online PF tau with respect to the tau E_T and η in semileptonic hadronic tau top-antitop events.
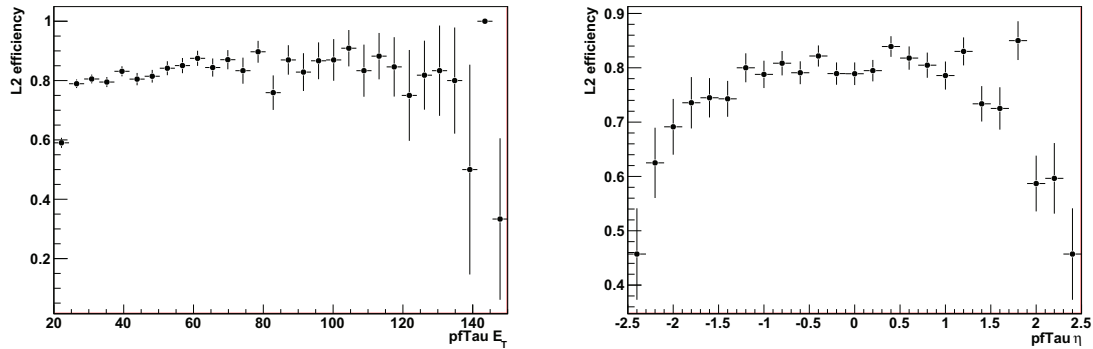

Figure 3.7: Efficiency of finding an online PF tau with leading track $p_T > 5$ GeV with respect to the tau E_T and η in semileptonic hadronic tau top-antitop events.
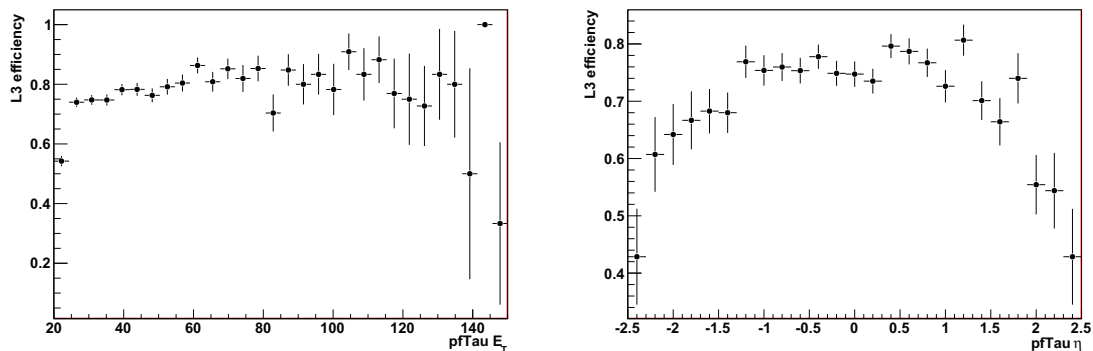

Figure 3.8: Efficiency of finding an isolated online PF tau with leading track $p_T > 5$ GeV with respect to the tau E_T and η in semileptonic hadronic tau top-antitop events.

3.4.4 Other investigated trigger configurations

As we can see on table **3.11**, the efficiency of the HLT_QuadJet40_IsoPFTau40 is not optimal for the $H^{\pm} \to \tau\nu_{\tau}$ signal. In order to increase the efficiency we tried to implement other configurations, relaxing the thresholds on the jet and tau momenta,

Trigger Name	$\epsilon(t\bar{t})$	$\epsilon(H^{\pm}_{m=90\text{GeV/c}^2})$	$\epsilon(H^{\pm}_{m=160\text{GeV/c}^2})$
QuadJet30_IsoPFTau30_eta2.5	0.77±0.02	0.40	0.32
QuadJet35_IsoPFTau35_eta2.5	0.68±0.02	0.39	0.32
QuadJet40_IsoPFTau40_eta2.5	**0.58±0.02**	0.39	0.31
QuadJet45_IsoPFTau45_eta2.5	0.49±0.02	0.38	0.30

Table 3.11: Estimated trigger efficiencies for the $5\cdot10^{32}$ cm^{-2}s^{-1} menu for the pre-elected $t\bar{t}$ and $H^{\pm}$ events.

and including the presence of PF H$_T^{miss}$ (missing scalar sum of E_T of pfJets (see section **4.2**)). More specifically, we tried to build an HLT path relying on the same L1 seed but with the following requests at HLT:

- 4 jets with $|\eta| < 2.5$ and $p_T > 30$ GeV;

- 1 PF tau with $|\eta| < 2.5$ and $p_T > 30$ GeV;

- PF H$_T^{miss}$ >30 GeV.

The estimated rate for a luminosity of $5\cdot10^{32}$ cm^{-2}s^{-1} for this path is of the order of 19.4±6.5 Hz, exceeding the allocated bandwidth of 10 Hz.

Regarding the signal efficiency, the studies done on simulated events showed that it increases for the $H^{\pm}$ signal but not for our analysis, (table **3.12**).

Trigger Name	$\epsilon(t\bar{t})$	$\epsilon(H^{\pm}_{m=160\text{GeV/c}^2})$
QuadJet400_IsoPFTau40_pfMHT20_eta2.5	0.51 ± 0.03	0.66 ± 0.04
QuadJet30_IsoPFTau30_pfMHT30_eta2.5	0.63 ± 0.04	0.60 ± 0.02

Table 3.12: Estimated trigger efficiencies for the $5\cdot10^{32}$ cm^{-2}s^{-1} menu for the pre-elected $t\bar{t}$ and $H^{\pm}$ events.

The HLT_QuadJet40_IsoPFTau40 trigger was finally chosen instead of the QuadJet30_IsoPFTau30_pfMHT30, not only for the lower rate but also to reduce the complexity of the trigger. The second path presented also problems of stability in the high pileup condition that indeed did not affect the HLT_QuadJet40_IsoPFTau40 trigger as we shall see in section **6.2.2**. The commissioning of the HLT_QuadJet40_IsoPFTau40 trigger on data will be described in section **6.2.2** while in the next chapter we shall describe the offline particle reconstruction in CMS.

Chapter 4

Offline reconstruction of particles at CMS

4.1 Introduction

In this chapter we will describe the offline particle reconstruction in CMS. In particular we will introduce the particle flow (PF) reconstruction technique. We will focus our attention on the reconstruction of the objects present in the $t\bar{t} \to \tau_h+$jets final state. Hence, the reconstruction of the jets with their energy calibration, the reconstruction of the τ leptons and of the missing transverse energy (E_T^{miss}) will be presented. The next chapter will be dedicated to the b quark jet identification.

4.2 The Particle Flow reconstruction

The PF algorithm (43) is an event reconstruction technique that benefits from the information coming from all CMS subdetectors with the aim to identify and reconstruct all particles produced in the collision: charged and neutral hadrons, photons, muons and electrons (figure **4.1**). The combination of all informations coming from the CMS detector allows the best determination as possible of the direction, energy and type of all the stable particles in the event. The PF benefits from the silicon tracker immersed in an uniform axial magnetic field of 3.8 T that allows an efficient reconstruction of the charged particles with small fake rate, for particles down to a transverse momentum p_T of 150 MeV, and pseudo-rapidities $|\eta| <$2.4. Starting from the list of single particles, the particle flow technique provides the reconstruction of higher level objects and observables like jets, hadronically decaying taus, missing transverse energy (E_T^{miss}), as well as lepton and photon isolations and b-jet tagging.
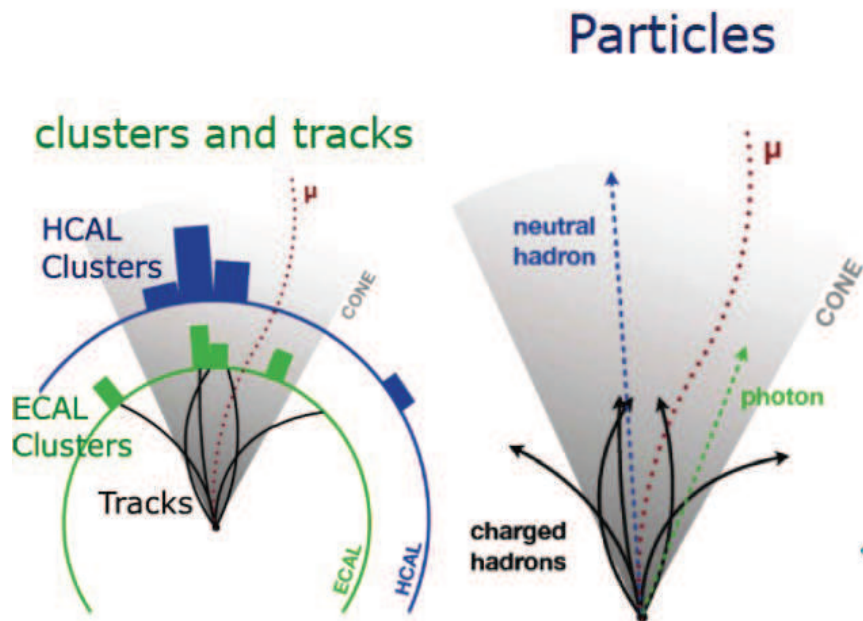
Figure 4.1: Illustrative plot of the functioning of the Particle Flow algorithm.

4.2.1 Iterative tracking and calorimeter clustering

Each element of the PF algorithm (charged particle tracks, calorimeter clusters as well as muons tracks), has to be reconstructed with high efficiency and with a low fake rate in a high-density tracks environment. These constraints impose to develop advanced tracking and clustering algorithms which will be described in the following of this section.

The iterative tracking

The main goal of the tracking algorithm is to provide a precise measurement of the direction of the charged particles starting from the tracker detector information. This step is crucial for the PF event reconstruction for two different reasons: the first one is that about two thirds of the energy of the jets is carried by the charged particles, the second one is due to the fact that each charged hadron missed by the tracking algorithm would only be detected by the calorimeters with a lower efficiency, a degraded energy resolution and a biased direction. Besides that, charged hadrons could be classifed as neutral ones. The tracking efficiency has to be as close as possible to 100% while the fake rate has to be kept small. To satisfy these requirements the CMS collaboration has chosen an iterative-tracking strategy (44). At the first step, all tracks are seeded and reconstructed applying some identification criteria on the p_T of the reconstructed tracks, as well as on the number of hits or the χ^2 of the fitted tracks (45). These criteria are chosen in order to keep low the fake rate

64

(under 1%). In the second step the algorithm proceeds by removing from the entire list of hits, the ones already assigned to the tracks in the previous iteration. Then the tracking algorithm runs again on the remaining hits, adapting the identification criteria in order to obtain the best compromise as possible between efficiency and fake rate. The optimization of the seeding criteria allows one to increase the tracking efficiency as well as to keep low the fake rate thanks to the reduced number of combinatorics. In the fourth and fifth iterations the constraints on the origin of the vertex are relaxed. The reconstruction of secondary charged particles, like particles originating from photon conversions can be done.

The calorimeter clustering

The goals of the calorimeter clustering are mainly the following ones:

- detection and measurement of the energy and direction of stable neutral particles;

- separation of the neutral particles from charged hadrons;

- reconstruction and identification of the electrons and all accompanying bremsstrahlung photons;

- improvement of the energy resolution of high p_T charged hadrons, for which the track parameters are measured with lower precision and for which the calorimeter energy resolution is more accurate.

In order to achieve these requirements, a specific clustering algorithm has been developed for the particle-flow event reconstruction. The clustering algorithm runs separately in each subdetector: ECAL barrel, ECAL endcap, HCAL barrel, HCAL endcap, PS first layer and PS second layer, and it consists of three steps:

- the "cluster seeds" are identified using *calorimeter-cells* if the energy exceeds a given threshold;

- "topological clusters" are formed starting from the seeds, by aggregating cells which have already at least one side in common with a cell present in the cluster, and with an energy above a given threshold. These thresholds represent two standard deviations of the electronics noise in the ECAL (i.e. 80 MeV in the barrel and up to 300 MeV in the end-caps) and amount to 800 MeV in the HCAL. A topological cluster gives rise to a so called "PF seed";

- One topological cluster can contain more than one seed due to the presence of two close particles. In this case the energy of the topological cluster is computed adding the energy of the neighbor cells by giving to each cell a weight according to the distance to the seeds. In this way we are able to take into account possible energy overlaps between two close particles.

4.2.2 The linking algorithm

A given particle is, in general, expected to give rise to several PF elements in the various CMS subdetectors: one or more charged-particle tracks, several calorimeter clusters, even muon tracks. These elements have to be connected to each other by a linking algorithm in order to fully reconstruct each single particle and avoid any possible double counting between subdetectors. Pair of elements are linked and the distance between any two linked elements is used to define the quality of the link. The algorithm then produces "blocks" of elements linked directly or indirectly. Thanks to the granularity of the CMS detectors, blocks typically contain few elements, which constitute the inputs for the particle reconstruction and identification algorithm. In any case they can contain several tracks linked to one ECAL or HCAL cluster.

More specifically, we can describe the link between a charged particle track and a calorimeter cluster with the following steps:

- the track is extrapolated from the last measured hit in the tracker to the two layers of the PS and to the ECAL, at a depth corresponding to the expected maximum of a typical longitudinal electron shower profile;

- The track is extrapolated to the HCAL, at a depth corresponding to one interaction length, typical of a hadron shower;

- The track is linked to any cluster found in the boundaries of the extrapolated position in the calorimeter. In order to account for the presence of gaps between calorimeter cells, and calorimeter modules, for the uncertainty on the position of the shower maximum as well as for the effect of multiple scattering for low-momentum charged particles, the cluster envelope can be enlarged by up to the size of a cell in all directions. The link distance between the extrapolated track position and the cluster position in the (η, ϕ) plane is finally defined.

In a similar way one can define a link between two calorimeter clusters, for example between HCAL and ECAL, or between ECAL and PS.

At the end, a link between a charged particle track in the tracker and a muon track in the muon system is defined when a global fit between the two tracks returns a χ^2 less than a given threshold. This link gives rise to a global muon and the χ^2 is used as variable to define the quality of the fit. Only the smallest χ^2 global muon is retained.

4.2.3 Particle reconstruction and identification

A list of reconstructed particles, stemming from blocks, is finally derived by the particle flow algorithm. This list provides a global description of each event, available

for subsequent physics analyses. In the following we will see how the blocks are used to provide the list of single particles.

- Each global muon gives rise to a "PF muon" if its combined momentum is compatible, within three standard deviations, with the momentum determined by the tracker information only. Then the corresponding track is removed from the block;

- The algorithm continues with electron identification. Each track of the block is submitted to a pre-identification stage which exploits the tracker and the pre-shower information. It is done since electrons tend to leave their energy by bremsstrahlung along their trajectory to the ECAL. Tracks pre-identified using the tracker and pre-shower information are then refitted with the Gaussian-Sum Filter algorithm (46) in order to reconstruct their trajectories up to the ECAL. This algorithm is used since it is able to take into account eventually missing hits in the tracker due to the bremsstrahlung photons produced in the interaction between electrons and the tracker material. The ECAL deposit due to the bremsstrahlung photon has to be taken into account. For this purpose the tangents to the tracks from the intersection points between the track and each of the tracker layers, are extrapolated to the ECAL. If a calorimeter cluster compatible with the tangent of a track is found in the ECAL, its energy is assigned to the electron. The final identification is performed by combining the tracker and calorimeter information with a proper weight in order to match the expected energy resolution. The track and ECAL cluster (also the one representing the bremsstrahlung photon) corresponding to the identified electron (PF electron) are then removed from the blocks;

- The algorithm proceeds to the identification of charged and neutral hadrons as well as of photons. If the calibrated energy of the closest ECAL and HCAL clusters linked to the remaining track(s) is significantly larger than the total momentum associated with the charged particles, the particle flow algorithm looks for photons and neutral hadrons. In particular if the excess is larger than the total ECAL energy, a photon is created with the ECAL excess and a neutral hadron is created with the remaining part of the excess. On the other hand it is possible that the total calorimeter energy is smaller than the total remaining track momentum. In this case the particle flow algorithm proceeds in two steps. First it looks for additional muons and fake tracks, and removes the corresponding tracks from the blocks. Then it builds charged hadrons with the remaining tracks.

4.3 Reconstruction of the $t\bar{t} \to \tau_h$+jets channel

As we already mentioned in Chapter 1, the $t\bar{t} \to \tau_h$+jets channel corresponds to the case in which one of the two W bosons involved in the process $pp \to t\bar{t} \to W^+bW^-b$ decays in an hadronic τ, and the other one hadronizes. Hence in the final state we have two jets from the hadronization of b-quarks, two light jets stemming from the decay of one of the two W bosons and one hadronically decaying τ with its corresponding neutrino. It means that the final state will also be characterized by the presence of missing transverse energy due to the presence of the neutrinos.

The ingredients needed to extract the signal from the background are: a good jet identification especially for b-jets, and τ identification. In the following of this chapter the jet, τ and E_T^{miss} reconstruction will be presented. The b-jet identification, instead, will be detailed in the next chapter.

4.3.1 Jet reconstruction

Four main types of jet reconstruction have been developed in CMS (47): calorimeter jets, Jet-Plus-Track jets (JPT), PF jets , and track jets. They differ from each other in the way the individual contributions from subdetectors are combined to form the inputs to the jet clustering algorithm.

The calorimeter jets are reconstructed combining the information coming from the ECAL and HCAL only, properly corrected in order to reduce the electronic noise as well as the pileup effects. Since calorimeter jets are the ones that provide the best compromise between resolution and execution time they are used for the reconstruction at trigger level.

The PF jets instead, are formed starting from the list of single particles reconstructed by the particle flow algorithm. Hence, since they are reconstructed with information coming from all CMS subdetectors, the PF jets are the ones with the best resolution and this is why they are used for the "offline" analysis.

For all jet types the used clustering algorithm is the "anti-kt" (48) with the parameter size $R = 0.5$. It relies on the definition of two distances: the distance d_{ij} between entities (particles, pseudojets) i and j, and the distance d_{iB} between the particles i and the beam (B). Their definition is given by the following expressions:

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p})\frac{\Delta_{ij}^2}{R^2}, \tag{4.1}$$

$$d_{iB} = k_{ti}^{2p}, \tag{4.2}$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and k_{ti}, y_i and ϕ_i are respectively the transverse momentum, rapidity and azimuth of a particle i.

The clustering proceeds in order to identify the smallest distances. If the smallest distance is d_{ij}, the entities i and j are recombined, while if the smallest distance is d_{iB} the entity (particle) i is labeled as a jet and is removed from the list of the entities.

As we can see in the Eq.(4.1, 4.2) in addition to the usual radius parameter R, a parameter p is used. It serves to switch between the "anti-kt" algorithm ($p = -1$), the inclusive k_t algorithm ($p = 1$), and the inclusive Cambridge/Aachen algorithm ($p = 0$). The "anti-kt" algorithm is generally used in CMS since it presents mainly one advantage. Thanks to the definition of the distance for $p = -1$, the low k_T particles tend to cluster with the high k_T ones instead to cluster among themselves. Hence the low k_T does not modify the shapes of the jet that will have a more regular shape with respect to the other algorithms (figure **4.2**). It improves the momentum resolution and the calorimeter performance.
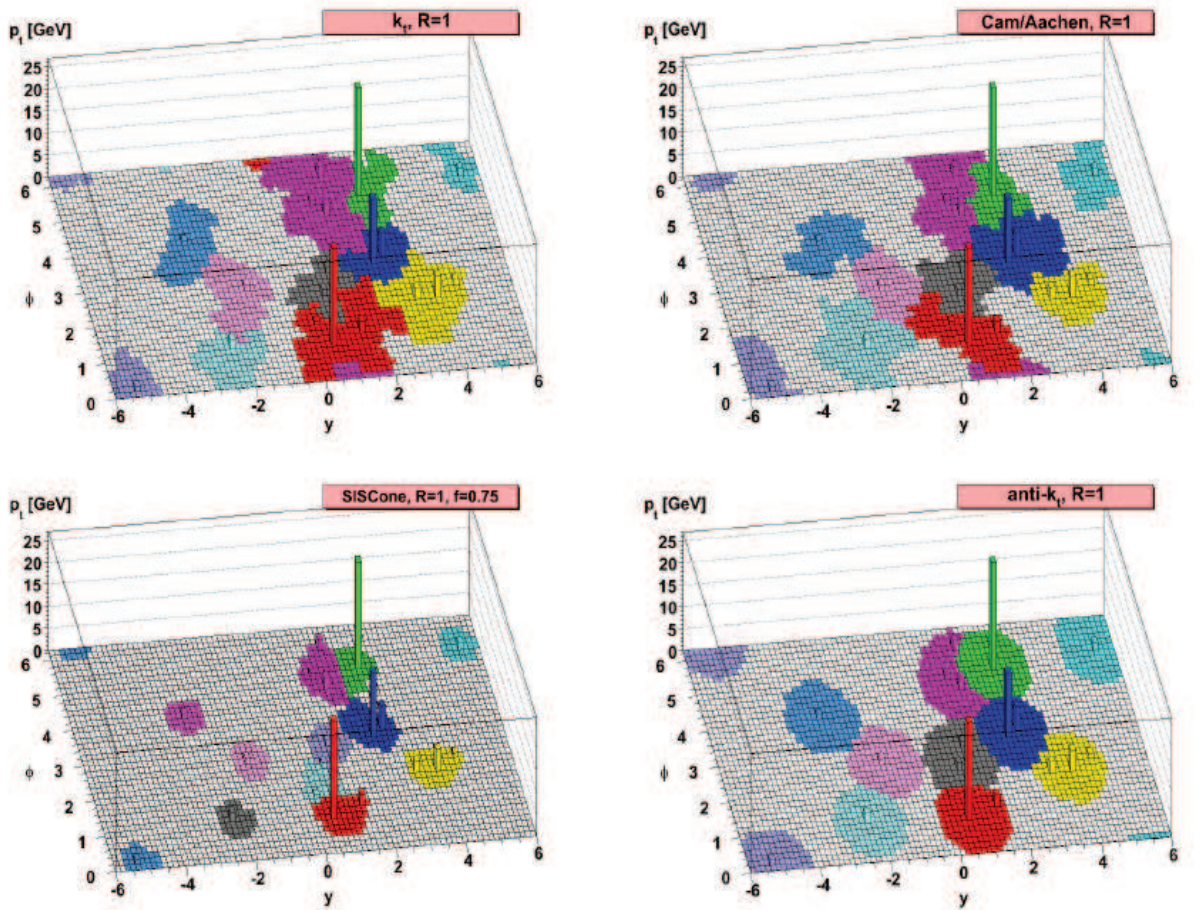
Figure 4.2: Jets are clustered with four different algorithms. The simulated events contain a few high energetic jets and many low k_T particles. Jets clustered with the "anti-kt" algorithm (bottom right) present a more regular shape with respect to the other ones: k_T corresponding to R=1 and p =1 (upper left), Cam/Aachen corresponding to R=1 and p =0 (upper right), SisCone corresponding to R=1 and p =0.75 (bottom left).

4.3.1.1 Jet energy corrections

The reconstructed energy of a jet is usually different from the corresponding real energy of the jet. The main source of this energy mismatch is the non-uniform and non-linear response of the CMS calorimeters as well as to the electronics noise and the presence of pileup events. The goal of the jet energy correction is to recalibrate the measured energy in the detector in order to match the real energy of the jet.
The jet energy correction in CMS is a factorized multi-step procedure (49). We have the following steps of correction: offset, relative and absolute corrections. The goal of the offset step is to correct the jet energy spoiled by the effect of electronics noise and pileup. The relative correction, instead, is performed to remove variations in the jet response versus pseudorapidity, relative to a unifor central region. At the end the absolute correction removes variations in jet response versus jet p_T. We can express the default sequence for the jet energy corrections with the following expression:

$$E_{Corrected} = (E_{Uncorrected} - E_{Offset}) \times C_{Rel}(\eta, p_T^{''}) \times C_{Abs}(p_T^{'}) \tag{4.3}$$

where $p_T^{''}$ is the transverse momentum of the jet corrected for the offset and $p_T^{'} = p_T^{''} \times C_{Rel}(\eta, p_T^{''})$ is the transverse momentum of the jet corrected for the offset and for the pseudorapidity dependence.

In CMS two complementary approaches to determine the jet energy correction factors are developed. The first one relies on the MC truth information (MC truth JEC), the second one instead, uses physics processes from pp collisions. The second approach is used for the offline analysis.

Jet energy correction using Monte Carlo information   The approach based on the MC information, derives the correction using a sample of simulated QCD events at $\sqrt{s}$ =7 TeV. The MC truth jet energy corrections do not factorize out the offset correction. Rather, the offset is lumped together with the relative and absolute corrections(50). For the relative correction the generated jets (GenJets) are matched to the calorimeter/JPT/PF jets in (η, ϕ) space by requiring $\Delta R < 0.5$. For the matched jets the quantity p_T^{Jet}/p_T^{GenJet} is studied to extract jet calibration factors as a function of uncalibrated jet p_T and η. The extraction procedure is divided in two steps: first the relative correction $C_{Rel}(\eta, p_T)$ is extracted by comparing the response at a given η with one relative to jets in the central region $|\eta| < 1.3$. In the second step, the absolute correction $C_{Abs}(p_T^{'})$ is computed. This correction removes the p_T dependence of the jet response and brings it to unity. In figure **4.3** the combined factor $C(P_T, \eta)$ for different jet types as a function of jet η is shown. In the left plot the correction factor required to get a corrected jet $p_T = 50$ GeV is presented, while in the right we have correction factor required to get a corrected jet $p_T = 200$ GeV. The combined factor $C(P_T, \eta)$ is given by the following expression:

$$C(p_T, \eta) = C_{Rel}(\eta, p_T) \times C_{Abs}(p_T \times C_{Rel}(\eta, p_T)). \tag{4.4}$$
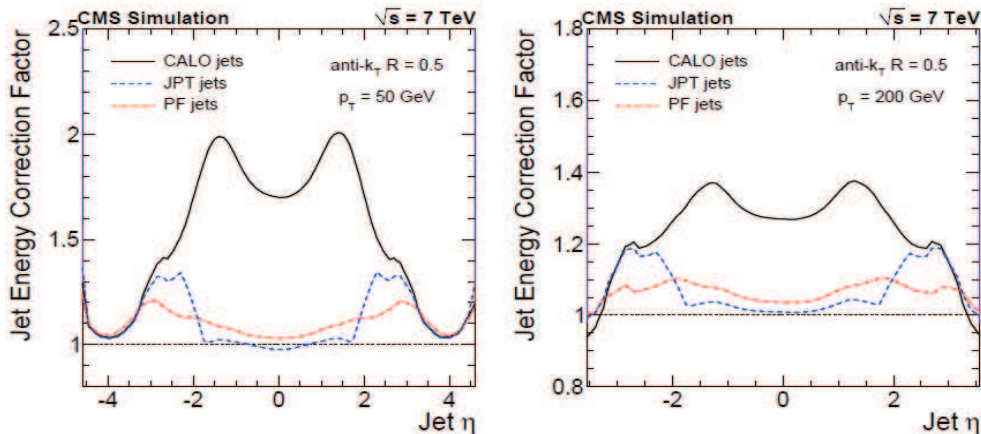

Figure 4.3: Monte Carlo jet-energy-correction factors for the different jet types, as a function of jet η. Left: correction factor required to get a corrected jet $p_T = 50$ GeV. Right: correction factor required to get a corrected jet $p_T = 200$GeV (49).

As expected the shape of the correction factor distribution presents a clear dependence in pseudorapidity that is more evident for calorimeter jets, which are reconstructed dividing the calorimeter in three different $|\eta|$ regions: barrel ($|\eta| < 1.3$), endcaps ($1.3 < |\eta| < 3$) an forward ($3 < |\eta| < 5$). Besides that, overall correction factors are smaller for PF jets, since they also profit from the tracking information.

Jet energy correction using real data To estimate the correction factor from data, the procedure starts by estimating the offset (49). This contribution, due to the electonics noise and pileup, is estimated using zero bias events. They are collected using a random trigger in presence of a beam crossing. Minimum bias events are excluded from the sample of zero bias events. Then the energy inside a cone of radius R = 0.5 in the $\eta - \phi$ space is summed. The noise contribution has been measured to be less than 250 MeV in p_T over the entire η range. The total average offset is than classified according to the total number of reconstructed vertices. In figure **4.4** the average offset p_T as function of η for different pileup condition, for calorimeter jets (right) and PF jets (left) is presented.

To retrieve the relative correction factor a special sample of dijet events is used. This sample is built by selecting events with 2 jets where one has to be in the central $|\eta| < 1.3$ region (barrel jet), no third jet with energy above a given threshold is accepted. The used technique is the p_T balance of the back to back dijet events: the p_T of the barrel jet is balanced with the p_T of the other jet (probe jet). The
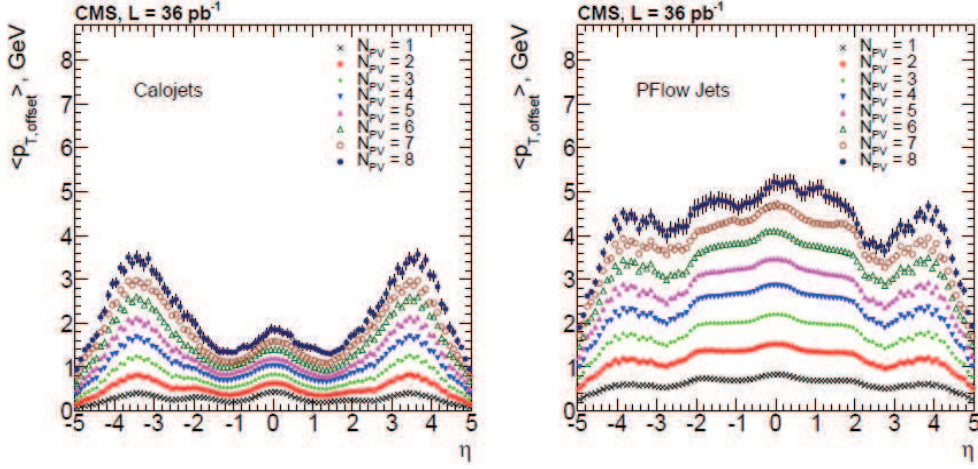
Figure 4.4: Average offset in p_T, as function of η, measured in zero bias events for different pileup conditions, for calorimeter jets (left) and for PF jets (right).

average of the balanced quantity is given by the following expression:

$$B = \frac{p_T^{probe} - p_T^{barrel}}{p_T^{dijet}} \qquad (4.5)$$

with $p_T^{dijet} = \frac{(p_T^{probe} + p_T^{barrel})}{2}$. The balance variable B is used to retrieve the relative correction factor:

$$R(\eta^{probe}, p_T^{dijet}) = \frac{2 + <B>}{2 - <B>}. \qquad (4.6)$$

In figure **4.5** the relative jet energy correction for calorimeter jets in different p_T^{dijet} bins is presented. A detailed description of the dijet method can be found in (49).

Jets in data are additionally corrected for a residual correction factor computed as the difference between the relative response in data with respect to the MC in order to take into account the observed shift in data (figure **4.6**).

To determine the absolute correction factor a MPF (Missing E_T Projection Fraction) method (50) is applied to a sample of γ+jets events. The MPF method is based on two assumptions:

- the γ+jet events have no intrinsic missing E_T;

- the photon is perfectly balanced by the hadronic recoil in the transverse plane.

Starting from these hypotheses the absolute correction factor is computed and the corresponding result is presented in figure **4.7**.
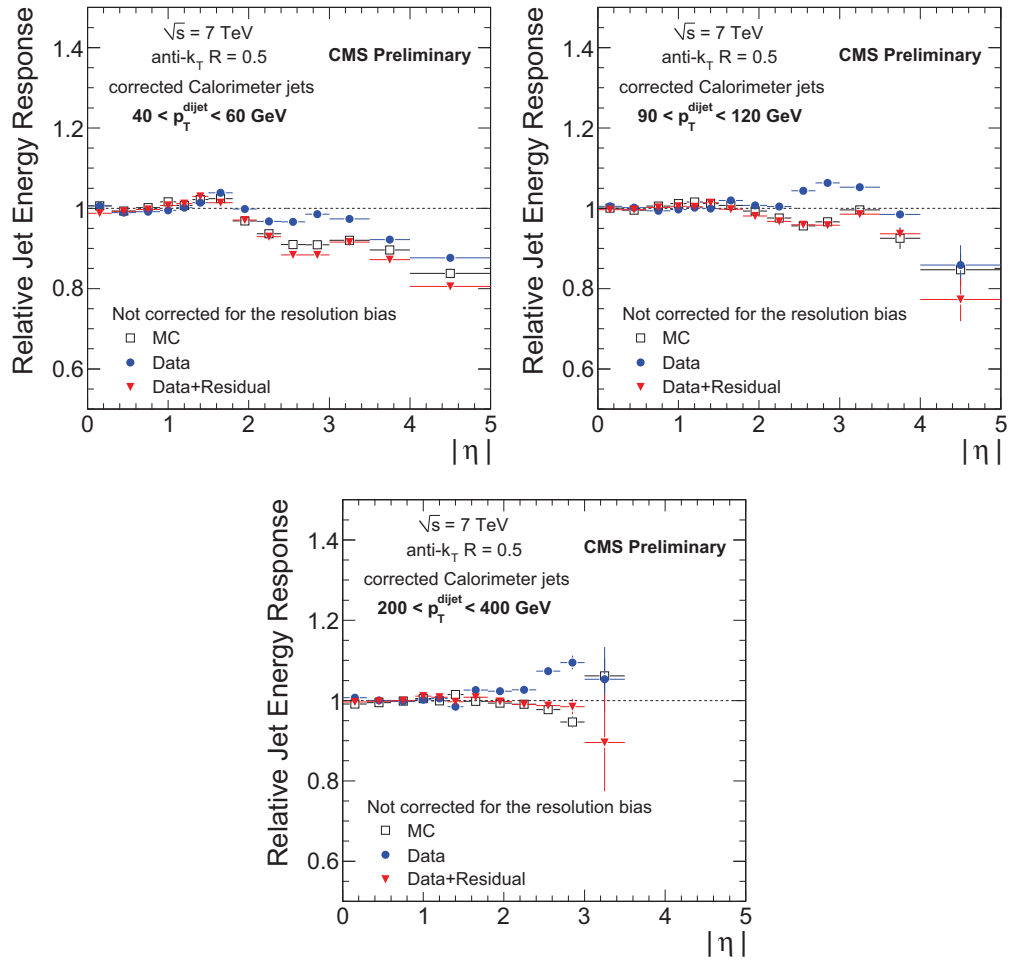
Figure 4.5: Relative jet energy response for calorimeter jets in various p_T^{dijet} bins.
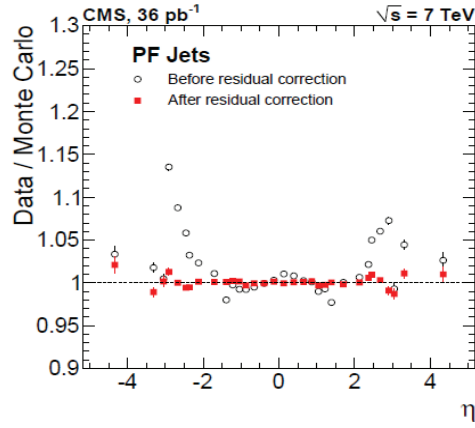
Figure 4.6: Data/MC ratio for the relative response obtained from the dijet p_T balance method for PF jets. Data are shown before (open circles) and after (solid squares) applying the residual correction and are compared to the uncertainty of the measurement (49).
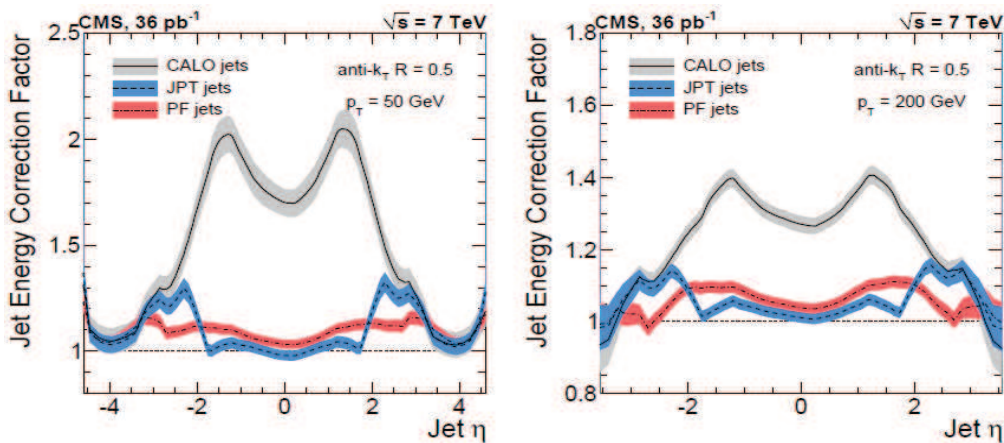

Figure 4.7: Total jet energy correction factor and its uncertainty (band) as a function of the jet η for two jet p_T values (49).

Table 4.1: Properties and decay modes of the τ lepton (51).

Mass		1.777 GeV
Charge		± 1
Spin		1/2
Lifetime		$(290.6 \pm 1.0)\cdot10^{-15}$s
Decay mode	Resonance	Branching ratio
Leptonic		
$\tau^- \rightarrow \mu^- \nu_\tau \nu_\mu$		17.36%
$\tau^- \rightarrow e^- \nu_\tau \nu_e$		17.85%
Hadronic		
$\tau^- \rightarrow h^- \nu_\tau$		11.6%
$\tau^- \rightarrow h^- \pi^0 \nu_\tau$	$\rho(770$ Mev$)$	26.0%
$\tau^- \rightarrow h^- \pi^0 \pi^0 \nu_\tau$	$a_1(1230$ Mev$)$	10.8%
$\tau^- \rightarrow h^- h^+ h^- \nu_\tau$	$a_1(1230$ Mev$)$	9.8%
$\tau^- \rightarrow h^- h^+ h^- \pi^0 \nu_\tau$		4.8%
Other hadronic decay mode		1.7%
Total hadronic decay mode		64.8%

4.3.2   Tau particle reconstruction

The tau lepton is the heaviest lepton of the Standard Model with a mass of 1.777 GeV. It is also the only one that can hadronize. In table **4.1** its characteristics are summarized.

The short lifetime of the τ leptons doesn't allow to identify them easily with a direct measure of a displaced secondary vertex. In this sense we refer to a **tau** as a jet from light mesons coming from the hadronic tau decay mode, while the leptonic tau decay modes are reconstructed by the standard muon and electron PF identification. Two main characteristics distinguish the tau jets from the normal QCD light jets:

- the tau jet is not color linked to the gluon sea of the underlying event. For this reason the tau jets are usually more confined and isolated with respect to the light jets from QCD events;

- the tau jets are the result of a small number of decay modes, so we can try to reconstruct each of them.

The main issue of the tau identification is the large background from which the tau jets have to be separated. In the following a brief overview of the main background processes is given:

- QCD events represent the largest background. Tau jets, as we already mentioned, are more isolated, and generally have a lower track multiplicity, as well

as a displaced secondary vertex, but QCD jets can fake all these properties especially in a high pileup environment;

- Electrons can fake taus decaying into $\pi^{\pm}\pi^0\nu_\tau$. This contribution can be reduced by looking for different signatures of electron and hadrons in the ECAL/HCAL subdetectors, as well as for bremsstrahlung photons;

- Muons can also fake taus if some energy is deposited in the HCAL. This contribution can be removed by comparing the HCAL deposit with the tracking momentum and by looking for deposits in the muon chambers.

As we already mentioned the identification of tau leptons is mainly done in the hadronic decay modes. Three approaches to identify the tau jets are mainly available in CMS using PF jets as input:

- the cut based method;

- the hadron plus strip (HPS) method;

- the Tau Neural Classifier (TaNC) method.

4.3.2.1 Cut based tau identification

The simplest form of tau jets are PF jets considering particles only in a given *signal cone* centered around the highest p_T charged hadron (figure **4.8**). All other particles identified outside the signal cone and contained in a larger isolation cone are used to calculate the isolation.
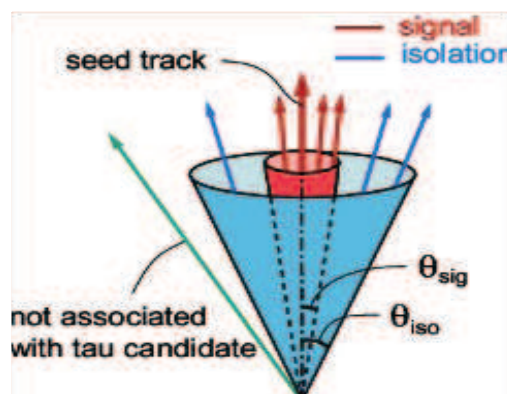

Figure 4.8: Reconstruction of the a tau jet using a signal and an isolation cone around the tau leading track.

Two different ways to build the signal cone exist:

- **Fixed cone**: the signal cone has a fixed size around the leading hadron, usually ΔR$<$ 0.1;

- **Shrinking-cone**: the radius of the signal cone is a function of the jet E_T, usually ΔR$=\frac{5GeV}{E_T}$ with a cone size constrained between 0.07 and 0.15.

To the PF tau candidates found with one of the two methods explained above, one can apply a list of discriminators in order to distinguish true taus from fake ones.

- **Leading track** p_T: we apply a cut on the leading track $p_T >$10 GeV;

- **ECAL isolation**: no PF neutral hadrons or photons with E_T above 1 GeV are accepted in the annulus cone;

- **Tracker isolation**: no PF charged hadrons with p_T above 1.5 GeV are accepted in the annulus cone;

- **Charged particle multiplicity**: tau jets have to contain one or three charged hadrons;

- **Against muons**: the tracks identified in the signal cone must not match a global muon track;

- **Against electrons**: the leading track is tested using the PF pre-identification algorithm (see section **4.2.3**).

This algorithm is the simplest one and for this reason is used in the high level of the trigger using the shrinking-cone definition, without applying the discriminators against electrons and muons.

4.3.2.2 Hadron plus Strip (HPS) algorithm

The goal of the HPS algorithm is to reconstruct the tau jets identifying the neutral and charged hadrons stemming from the tau decay. It this why the main hadronic decay modes can be reconstructed (52) except the $3\pi^{\pm}\pi^0\nu_{\tau}$ decay mode. Since the decay of the tau gives rise, in most of the cases, to a neutral pion (π^0), an efficient π^0 reconstruction is needed. The typical signature of the neutral pions are photons with broad deposits in the ECAL due to photon conversions.

Hence a **strip** reconstruction is used for the electromagnetic fraction of the taus as described on figure **4.9**. More specifically the strip algorithm starts from the electromagnetic object with largest E_T in the jet cone, then searches for additional electromagnetic objects in a strip of size $\Delta\eta <$0.5 and $\Delta\phi <$0.2. Each found object is added to the previous one and the strip is re-centered on the aggregated center, until no more objects are found. Additional strips are created for each object in the jet not associated with the first strip. Finally only strips with aggregated $p_T >$1
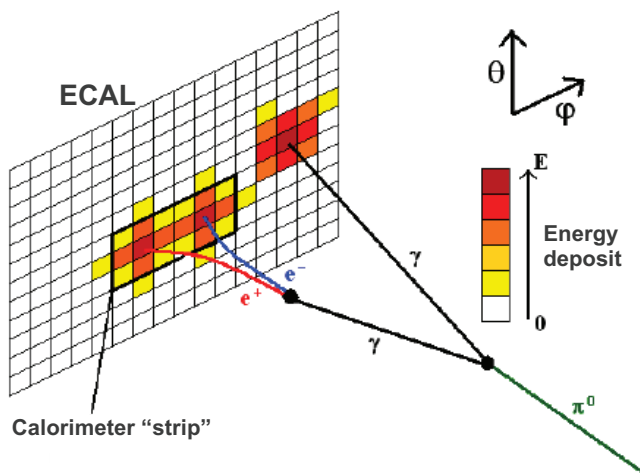
Figure 4.9: Sketch of the strips algorithm.

GeV are kept. Then the center of the charged hadrons and strip is computed and is retained if located inside a shrinking cone with $0.05 < \Delta R < \frac{2.8}{p_T} < 0.1$ (using the p_T of the summed hadrons and strips). Then the algorithm attemps to classify the selected hadrons and strips in possible decay modes:

- single hadron, equivalent to $h^- \nu_\tau$ or $h^- \pi^0 \nu_\tau$ (if low p_T π^0);

- hadron + strip, equivalent to $h^- \pi^0 \nu_\tau$ (the two photons stemming from the π^0 are close on the calorimeter surface);

- hadron + two strips, equivalent to $h^- \pi^0 \nu_\tau$ (the two photons stemming from the π^0 are well separated), or to $h^- \pi^0 \pi^0 \nu_\tau$;

- 3 hadrons, equivalent to $h^- h^+ h^-$.

For each decay mode a different cut on the invariant mass of the final state is imposed:

- single hadron: no mass cut, the p_T of the hadron has to be higher than the p_T cut imposed on the *tau*;

- hadron + strip: $0.3 < m_{\pi^\pm \pi^0} < 1.3$ GeV (reconstruction of the ρ resonance);

- hadron + two strips: $0.4 < m_{\pi^\pm \pi^0 \pi^0} < 1.3$ GeV (reconstruction of the a_1 resonances);

- 3 hadrons: $0.8 < m_{\pi^\pm \pi^\pm \pi^\pm} < 1.5$ GeV (reconstruction of the a_1 resonance).

	p_T cut on PF h^+ [GeV]	E_T cut on PF γ [GeV]	Eff MC [%]	Eff Data [%]	Data/MC
Loose	1.0	1.5	50.2	50.1±2.9	0.997 ±0.057
Medium	0.8	0.8	30.1	30.4±2.3	0.945 ±0.070
Tight	0.5	0.5	16.4	14.9±1.0	0.908 ±0.062

Table 4.2: Definition of the working point for the HPS algorithm. The corresponding efficiencies computed on MC and data (Run2011A+B) are presented.

working point	isolation definition	Eff MC [%]	Eff Data[%]	Data/MC
Loose	Sum[(PF h^+ + PF γ) p_T] < 2 GeV	58.2	58.9±3.3	1.012 ± 0.057
Medium	Sum[(PF h^+ + PF γ) p_T] < 1 GeV	46.2	45.4±2.6	0.982 ± 0.057
Tight	Sum[(PF h^+ + PF γ) p_T] < 0.8 GeV	42.8	42.1±2.7	0.983 ± 0.063

Table 4.3: Definition of the working points for the combined discriminators including $\Delta\beta$ corrections. The corresponding efficiency computed on MC and data (Run2011A+B) are presented.

If the tau candidate is reconstructed in several modes, the one keeping the maximum fraction of the entire jet energy is kept.

The final step of the algorithm is the computation of the isolation. The isolation is computed in a solid cone of ΔR=0.5 around the reconstructed decay mode axis. It can be estimated using all candidates in the cone or adding only the energy of the candidates above a given threshold. In both cases the constituents corresponding to the reconstructed tau are subtracted. The main advantage of this procedure with respect to the standard isolation annulus method is that one is able to reject also narrow jets with high electromagnetic or hadronic fractions. Three isolation working points are defined. Their definitions with the corresponding efficiencies are presented in table **4.2**. The efficiency is estimated using $Z \rightarrow \tau\tau$ events as explained in reference (53). The efficiency values given in tables **4.2** are estimated on the entire 2011 statistics (Run2011A+B corresponding to the first and second part of the 2011 data taking). In figure **4.10** the fake rate corresponding to the loose working point as function of jet p_T is shown. In the bottom pannel the comparison between data and MC expectation is presented.

As already mentioned, the pileup conditions during the 2011 data taking complicated the offline reconstruction of the particles. This is why the tau group provided a new set of discriminators for the HPS algorithm, in order to remove eventually charged particles or photons stemming from pileup from the isolation computation (54), (55). More specifically the PF charged hadrons stemming from pileup are rejected by requiring Δz <2 mm with respect to the highest p_T track of the τ candidate. PF photons instead, are corrected using the so called $\Delta\beta$ corrections. An estimation of the pileup is given using the sum of the transverse momenta of the particle flow charged hadrons with Δz >2 mm: $\Delta\beta = k \cdot \sum p_T^{\pm}$ (k is a proportional factor depending on the parametrization). The pileup-corrected PF photon isolation is computed as: $\sum p_T^0$ - $\Delta\beta$. Using the $\Delta\beta$ correction procedure a new combined

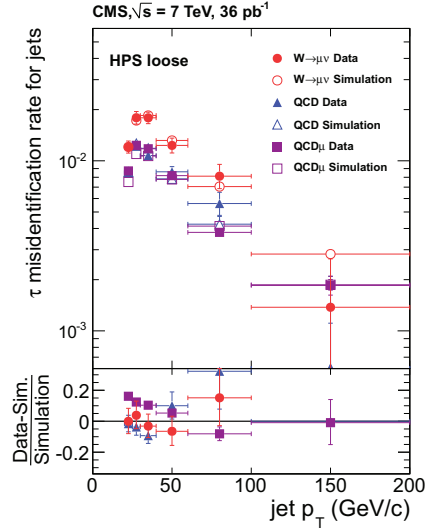**Figure 4.10:** Fake rates for the HPS algorithms (loose working point), derived from W→ $\mu\nu_\mu$ and QCD events, as function of jet p_T. The comparison between data and MC expectation is presented in the bottom pannel.
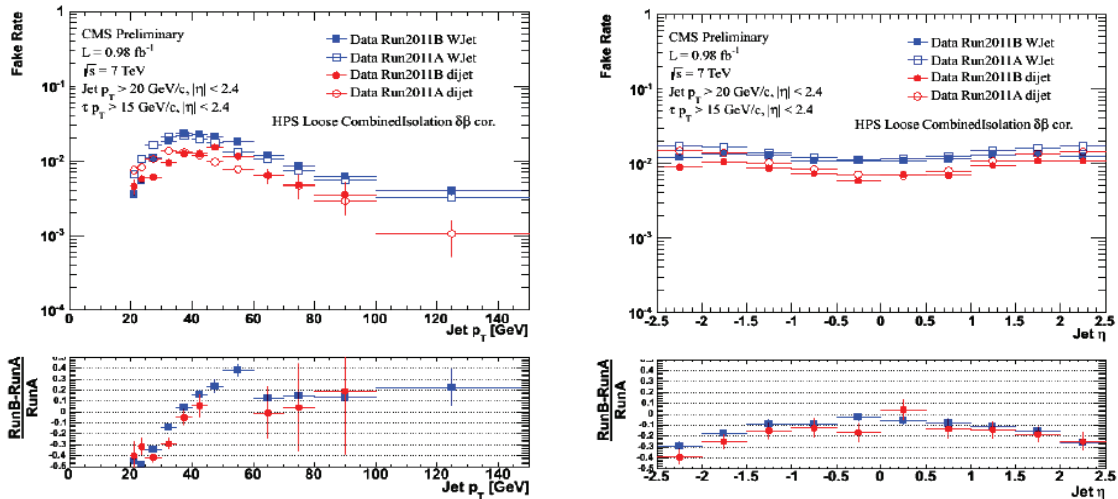

Figure 4.11: Fake rates for the HPS combined algorithm (loose working point), derived from W+jets and QCD dijets events, as function of jet p_T and η. The comparison between Run2011A and Run2011B is presented in the bottom pannel.

	$|\eta|$ <1.5 MC [%]	$|\eta|$ <1.5 Data [%]	$|\eta|$ >1.5 MC [%]	$|\eta|$ >1.5 Data[%]
Run2011A	2.79 ± 0.01	2.58 ± 0.05	2.37 ± 0.02	2.42 ± 0.01
Run2011B	2.75 ± 0.01	2.48 ± 0.1	2.84 ± 0.1	2.43 ± 0.2

Table 4.4: Fake rate for the tight discriminator against electrons in Run2011A and Run2011B. The fake-rate seems lower in data with respect to the MC expectation.

discriminator is defined with three different working points as summarized in table **4.3**. The corresponding values for the efficiency are also presented. In figure **4.11** the jet→ τ fake rate for the loose combined discriminator $\Delta\beta$ corrected as function of jet η (left) and jet p_T (right), is presented. In our $t\bar{t} \to \tau_h$+jets analysis we will use the combined discriminator including $\Delta\beta$ corrections with its medium working point as recommended by the tau group since it represents the best compromise between efficiency and fake rate.

In order to reduce the background due to the electrons and muons, special disciminators are developed. The discrimination against electrons relies on a boosted decision tree, that combines variables which characterize the presence of the neutral particles reconstructed in the tau decay (e.g. number of constituents, spread, energy fraction) as well as the presence of a charged hadron and electromagnetic particles (e.g. energy fractions, electron-pion discriminator). To suppress the muon contamination the leading track of the tau candidate should not be reconstructed as a muon that passes segment and track muon criteria. A single charged hadron tau candidate should in addition not be reconstructed as a minimum ionizing particle (H/P> 0.2 where H is the linked HCAL energy to the reconstructed charged hadron and P its track momentum). The fake rate corresponding to the discriminator against electrons is given in table **4.4**. It seems stable with respect to the different pileup scenarios (Run2011A and Run2011B).

Regarding the muon fake rate, in Run2011A the fake rate has been measured on data and is (1.2±0.36)·10^{-3} in perfect agreement with the one estimated from simulation. For Run2011B the fake rate on simulation is the same, while the computation on data is still ongoing. The estimation of the electron and muon fake rate on data is done using a tag and probe technique on an electron/muon + tracks data sample (56).

4.3.2.3   Tau Neural Classifier (TaNC) algorithm

The Tau Neural Classifier (57) is an identification algorithm that takes as input the tau candidates classified according to their different hadronic tau decay modes as described in (58).

The full collection of PF photons within the tau cone is examined in order to reconstruct all possible invariant mass pairs. Each pair with M<200 MeV is tagged

as a π^0 candidate, the closest one to the π^0 mass is retained (the less well recon-structed are removed). The unpaired photons (carrying at least 10% of the jet E_T) are analyzed in order to consider the case in which the two photons from the π^0 decay impact the ECAL too closely to be resolved, or the case where the decay was sufficiently asymmetric that one photon missed the detector acceptance.

The neural network input variables are a set of kinematic and topological prop-erties of the tau like the invariant mass of the signal tracks, the p_T of the leading track and the p_T of the Nth charged hadron or π^0 constituent.

The set of neutral and charged hadrons are then trained with a neural network. One training for each of the five leading decay modes is performed. Working points are defined for the tree level of fake rate of 1%, 0.5% and 0.25%. The training takes as input Monte-Carlo $Z\rightarrow \tau\tau$ events as signal and QCD as background. The fake rates versus the efficiency for the different working points for HPS and TaNC algorithms are presented in figure **4.12**

.
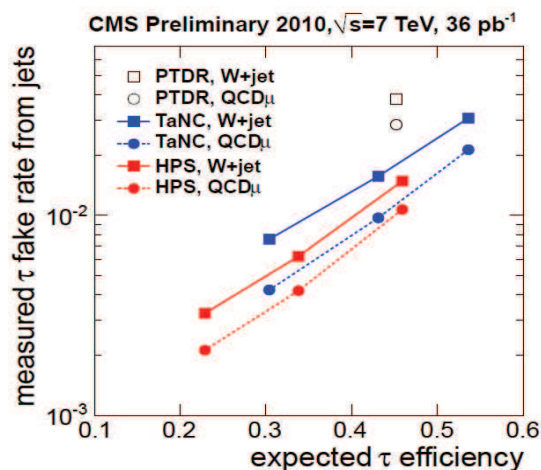
Figure 4.12: Fake rates (from W + jet and QCD events) for the HPS and TaNC algorithms, derived from W and Z events (59). The points correspond (left-to-right) to tight, medium and loose working points. The open points, instead, represent the expected tau fake rates/efficiencies from the TDR (60), for the shrinking cone algo-rithm.

4.3.3 Missing Transverse Energy reconstruction

The presence of neutrinos, since they don't interact on a short scale distance, can only be inferred by the apparent non-conservation of energy in the transverse plane. It is impossible in fact, to know the actual momentum of the interacting partons in the beam direction, this is why the inference can only be done perpendicular to the beamline. Furthermore it is not possible to know if one ore more neutrinos are involved in the process. In CMS three different E_T^{miss} reconstructions are available:

- **Calo $\mathbf{E}_T^{miss}$**: it corresponds to the negative sum of the vector for ECAL and HCAL towers, with energy scale for jets (type I) and low-p_T unclustered particles (type II);

- **TC $\mathbf{E}_T^{miss}$**: it is computed using the Calo $\mathbf{E}_T^{miss}$ by substituting the calorimeter deposits with the tracker information of the identified charged hadrons. In this way we profit of the better resolution of the tracker with respect to the HCAL one.

- **PF $\mathbf{E}_T^{miss}$**: it represent the negative vector sum of all PF particles reconstructed as described in section **4.2.3**.

The PF $\mathbf{E}_T^{miss}$ is the one providing the best resolution, this is why we will use it in the offline analysis.

In figure **4.13** the resolution for the three different methods is shown.

.
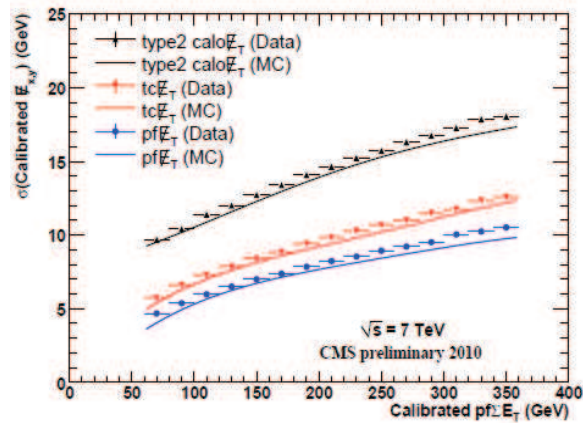
Figure 4.13: Resolution for the three main E_T^{miss} reconstruction methods as function of the scalar sum of the transverse energy in the event. As expected the PF $\mathbf{E}_T^{miss}$ gives the best performance.

84

Chapter 5

The b jets identification

5.1 Introduction

The b-jet algorithms in CMS relies on the long lifetime, high mass and large momentum fraction of b hadrons produced in b-quark jets, as well as on the presence of soft leptons from semi-leptonic b decays. The b-jet identification is a fondamental tool to select top-antitop events which are characterized by the presence of two b jets in the final state. In this chapter the b-tagging algorithm and their performance will be detailed. We shall describe the observables used to build the b-tag algorithms, then the selection of events, tracks and jets used for the commissioning studies, finally the main b-tag algorithms will be presented. In section **5.5** the JetProbability algorithm is presented in more details since it is the one used for the offline analysis and the one for which I provided the calibration since 2009. Finally in section **5.7** the performance of the different taggers are summarized.

5.2 The b-tagging observables

The b-tagging algorithms rely on the measure of three main variables:

- The impact parameter (IP). It is defined as the distance in space between tracks (linearized around the point of closest approach to the PV) and the jet axis at the point of closest approach to the primary interaction vertex (figure **5.1**). To account for resolution effects, we use the IP divided by its error: $IP/\sigma(IP)$. The impact parameter can be signed as positive (negative) if the associated tracks are produced downstream (upstream) with respect to the primary interaction vertex;

- the measure of the position of the secondary vertex (SV). It is the point where the b-hadron decays (figure **5.1**). It is possible to measure the position of this point thanks to the high resolution of the CMS tracker system. The secondary vertex reconstruction is performed using adaptive vertex fitter (61) algorithm.
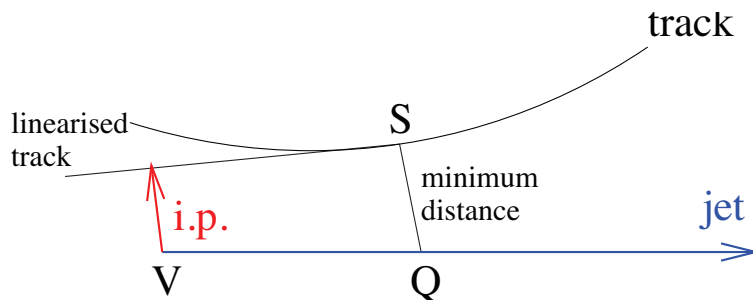
Figure 5.1: Geometric meaning of the impact parameter (IP). The sign is positive (negative) when the angle between the impact parameter direction and the jet axis is smaller (larger) than 90^0. The primary vertex (PV) is shown in green while the secondary vertex, the point where the b hadron decays, in red.

> The resulting list of vertices is then subject to a cleaning procedure, rejecting SV candidates that share 65% or more of their tracks with the PV;

- the p_T of the muons with respect to jet axis. This variable is used to tag the b-jets since semileptonic decays of b hadrons give rise to b jets that contain a muon with a branching ratio of about 11%, or 20% when b→c→l cascade decays are included.

Starting from these variables the different algorithms provide the *discriminators*, i.e. output variables on which the users can cut on to select b-jets in different regions of the efficiency versus purity phase space. Varying the cuts on the discriminator we obtain different efficiencies of the taggers. We establish standard operating points as, loose (L), medium (M), and tight (T), being the value at which the tagging of light jets is estimated from MC to be 10%, 1%, or 0.1%, respectively, for jets with p_T of about 80 GeV.

5.3 Selection for b-tagging commisioning and performance studies

For the commisioning of the taggers as well as for the study of their performance a specific selection of the events, tracks, and jets is applied. In the following the selection is explained in more details.

5.3.1 Event selection

We select the events that pass the following requirements: colliding events selected by removing *beam−scraping* events (the fraction of reconstructed high-purity tracks

in the event has to be larger than 25%) and by requiring a good primary vertex, PV (number associated tracks > 10, PV coordinates within $|z| < 24$ cm along the beam axis and within $|d0| < 2$ cm in the transverse plane). The primary vertex is reconstructed from all tracks in the event which are compatible with the beam spot. The Deterministic Annealing Filter algorithm (62), (61) is used for the PV reconstruction.

A pre-scaled single jet trigger, HLT_Jet60, was used to collect the data, and a p_T threshold of 80 GeV was applied offline in data and simulation to produce the calibration for the JetProbability algorithm as we shall see in the section **5.5**.

5.3.2 Jet selection

The anti-kT Particle-Flow algorithm (48) is used to reconstruct jets within a cone $\Delta R < 0.5$ around the jet axis (where $\Delta R = \sqrt{\Delta \phi^2 + \Delta \eta^2}$). Jets with corrected p_T larger than 20 GeV and within the tracker acceptance, $|\eta| < 2.4$, are considered. Other loose identification criteria are required: a neutral hadron energy fraction < 1.0, a neutral electromagnetic (photon) energy fraction < 1.0, a charged hadron energy fraction > 0, a charged electromagnetic (electron) energy fraction < 1.0, and a number of charged+neutral particles ≥ 2.

In the simulation, a parton flavor is assigned to a jet if the direction between the generated parton and the reconstructed jet is within a cone of radius $\Delta R < 0.3$. At first a b jet is looked for, otherwise a c jet is considered and, if no b or c flavor is found, a light flavor is assigned to the jet.

5.3.3 Track selection

As we already mentioned, during the 2011 data taking the number of pile-up events per bunch crossing increased significantly. These new pile-up conditions imposed to apply a special selection on tracks in order to remove tracks coming from the pile-up events without compromising the performance of the different taggers. Tracks are required to satisfy the following quality criteria:

- number of pixel hits (NpHits) ≥ 2;

- number of tracker hits (including pixel) ≥ 8;

- transverse impact parameter with respect to the primary vertex (PV) $|IP_{2D}| > 0.2$ cm;

- transverse momentum $p_T > 1$ GeV;

- normalized $\chi^2 < 5$;

- longitudinal impact parameter with respect to the PV, $|IP_z| < 17$ cm;

- distance to the jet axis < 0.07 cm. It is defined as the spatial distance between the track trajectory and the jet axis at their point of closest approach;

- decay length < 5 cm, defined as the spatial distance between the primary; vertex and the point of closest approach between the track trajectory and the jet axis.

The effect of the selections is described in figure **5.2** where we see the number of tracks per event before (left) and after (right) applying the selection. The track selection is clearly rejecting those additional tracks from nearby primary vertices.
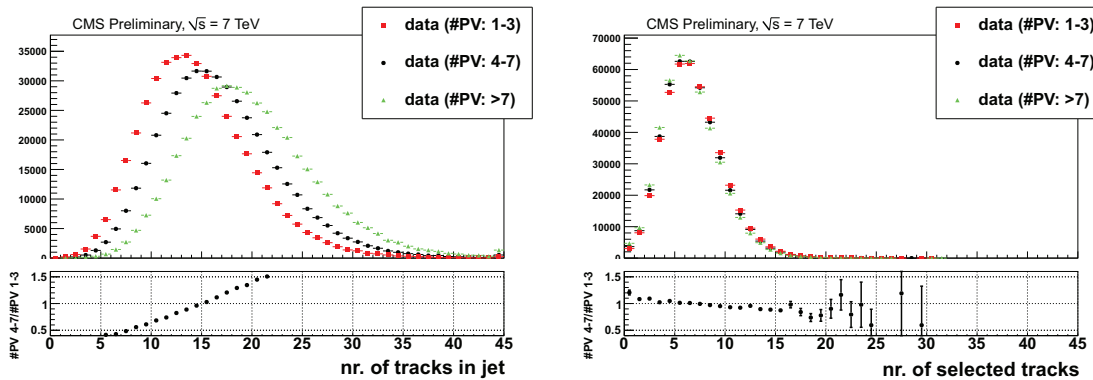
Figure 5.2: Number of tracks before (left) and after (right) the quality selection.

5.4 Track Counting algorithm

The track counting algorithm identifies a b-jet if the jet contains at least N tracks with a significance of the impact parameter above a given threshold (63). The tracks
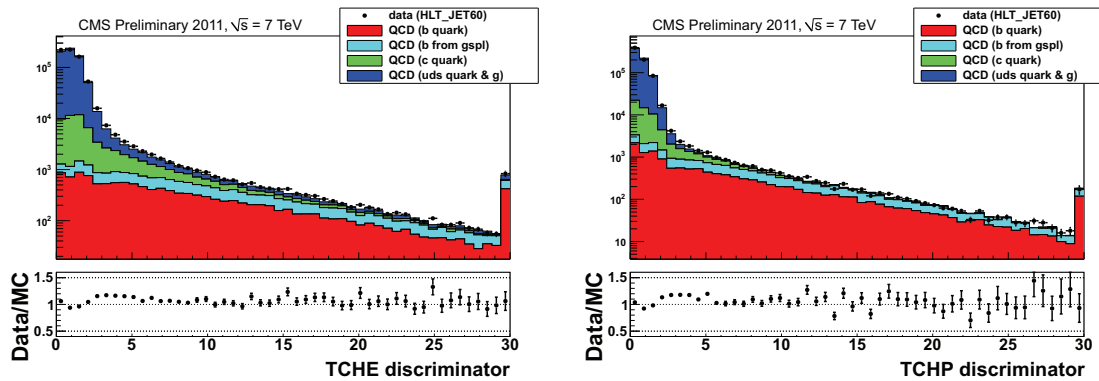
Figure 5.3: Discriminator (impact parameter) for the Track Counting High Efficiency (left) and High Purity (Right) algorithms.

are ordered in decreasing $IP/\sigma(IP)$ and the discriminator is the impact parameter

significance of the Nth track. To get an high b-jet efficiency we can use the $IP/\sigma(IP)$ of the second track (Track Counting High Efficiency, TCHE)(figure **5.3** left), to select b-jets with high purity the third track is the better choice (Track Counting High Purity, TCHP)(figure **5.3** right). The plots presented in figure **5.3** are obtained for 2011 data.
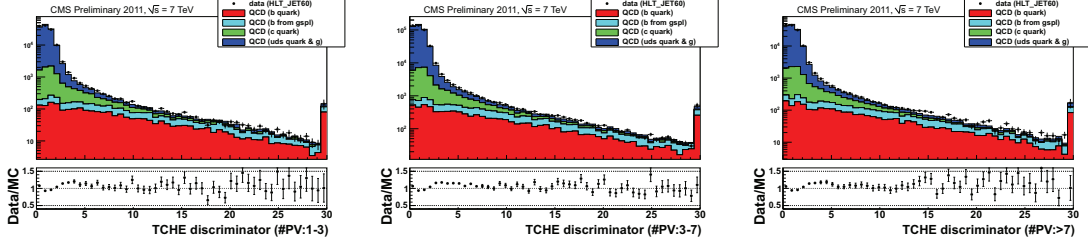

Figure 5.4: Data/MC comparison of the track counting high efficiency discriminator for three samples characterized by different primary vertex multiplicities: (left) NPV: 1-3, (middle) NPV: 4-7, (right) NPV: > 7.
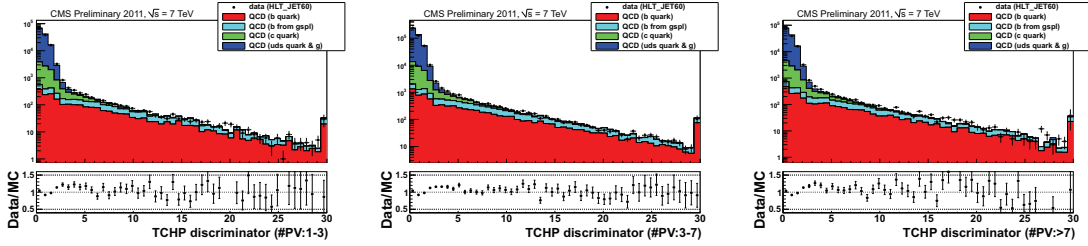
Figure 5.5: Data/MC comparison of the track counting high purity discriminator for three samples characterized by different primary vertex multiplicities: (left) NPV: 1-3, (middle) NPV: 4-7, (right) NPV: > 7.

In figures **5.4**, **5.5** the distribution of the discriminators for the TCHE and TCHP algorithms, for three different primary vertex multiplicities, are presented. The agreement between data and simulation is stable for the three bins of primary vertex multiplicity.

5.5   Jet Probability algorithm

The Jet Probaility algorithm (63),(64) relies on a specific feature of the $IP/\sigma(IP)$ distribution. Since the b-hadrons decay with a measurable decay length, the particles that come from b vertex have a larger $IP/\sigma(IP)$ value than the particles coming from the primary vertex itself. This is why the contamination from b-jets in the negative part of the $IP/\sigma(IP)$ distribution is small as shown in figure **5.6**. For the light jets instead, the $IP/\sigma(IP)$ distribution is about symmetric due to the IP resolution. For this reason we assume that all tracks with negative $IP/\sigma(IP)$ come from the PV and
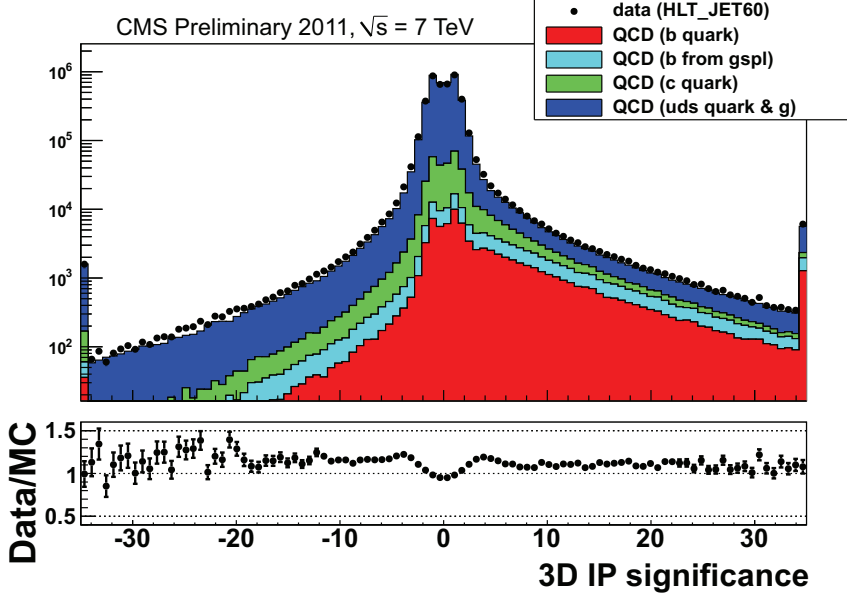
Figure 5.6: IP significance for all selected tracks in the jets. The tracks selection described in Section 5.3.3 is applied.

we use the negative tail of the IP/σ(IP) distribution to compute the probability of tracks to come from the PV. To do so, starting from the negative tail of the IP/σ(IP) distribution obtained from the data themselves, we define a *Resolution Function* $R(X)$ and we compute a probability from the following expression:

$$P_{tr}(S_{IP}) = \int_{|S_{IP}|}^{\infty} R(x)dx. \tag{5.1}$$

Combining the track probabilities, we define the probability of a jet, P_{jet}, to come from the primary vertex:

$$P_{jet} = \Pi \sum_{j=0}^{N-1} \frac{(-ln\Pi)^j}{j!} \tag{5.2}$$

$$\Pi = \prod_{i=1}^{N} P_{tr}(i). \tag{5.3}$$

with N being the number of tracks in a jet.

One can also defined a *Negative JetProbability* by only using the negative IP/$\sigma(IP)$ tracks. In the same way we can define a *Positive JetProbability* using only the positive IP/$\sigma(IP)$ tracks, which corresponds to the *JetProbability* tagger.

The shape of the negative side of IP/σ(IP) distribution is strongly dependent of the choice of the parameters of the tracks, as the NpHits, χ^2, p and η, as we can expect knowing the geometry of the CMS tracker.

Table 5.1: Definition of the 9 track categories. In the latest row is reported the number of tracks per categories corresponding to an integrated luminosity of 191 pb^{-1} collected at the beginning of the 2011 data taking with the HLT_Jet60 trigger requeste.

	Cat1	Cat2	Cat3	Cat4	Cat5
NpHits$_{Min}$	2	3	3	3	2
NpHits$_{Max}$	8	8	8	8	2
χ^2_{Min}	2.5	0	0	0	0
χ^2_{Max}	5	2.5	2.5	2.5	2.5
p_{Min}	1	1	1	1	1
p_{Max}	5000	8	8	8	8
$|\eta|_{Min}$	0	0	0.8	1.6	0
$|\eta|_{Max}$	2.5	0.8	1.6	2.5	2.5
nb of tracks	$1.19{\cdot}10^6$	$1.19{\cdot}10^7$	$6.2{\cdot}10^7$	$1.75{\cdot}10^7$	$3.45{\cdot}10^7$

	Cat6	Cat7	Cat8	Cat9
NpHits$_{Min}$	3	3	3	2
NpHits$_{Max}$	8	8	8	2
χ^2_{Min}	0	0	0	0
χ^2_{Max}	2.5	2.5	2.5	2.5
p_{Min}	8	8	8	8
p_{Max}	5000	5000	5000	5000
$|\eta|_{Min}$	0	0.8	1.6	0
$|\eta|_{Max}$	0.8	1.6	2.5	2.5
nb of tracks	$5.43{\cdot}10^7$	$5.56{\cdot}10^7$	$4.75{\cdot}10^7$	$3.22{\cdot}10^7$

To improve the performance of the tagger, we can define several categories of tracks corresponding to different shapes of the IP/σ(IP) distribution. The table **5.1** shows the definition of the categories used to calibrate the JetProbability algorithms. In figure **5.7** the distribution of the IP/σ(IP) for each category is presented.

Hence the calibration is computed for the 9 different categories separately and merged at the end in only one output file.

5.5.1   Calibration on 2011 Data

In figure **5.8** one can see the negative track probability (IP/σ <0) for 2011 data. By construction this distribution has to be uniform between 0 and 1. In figure **5.9** the distribution for the negative jet probability is presented. In both plots the black distribution represents the distribution of 2011 data calibrated with 2010 data (old calibration), and the red distribution is relative to the 2011 data calibrated with an indipendent sample of 2011 data itself (new calibration).
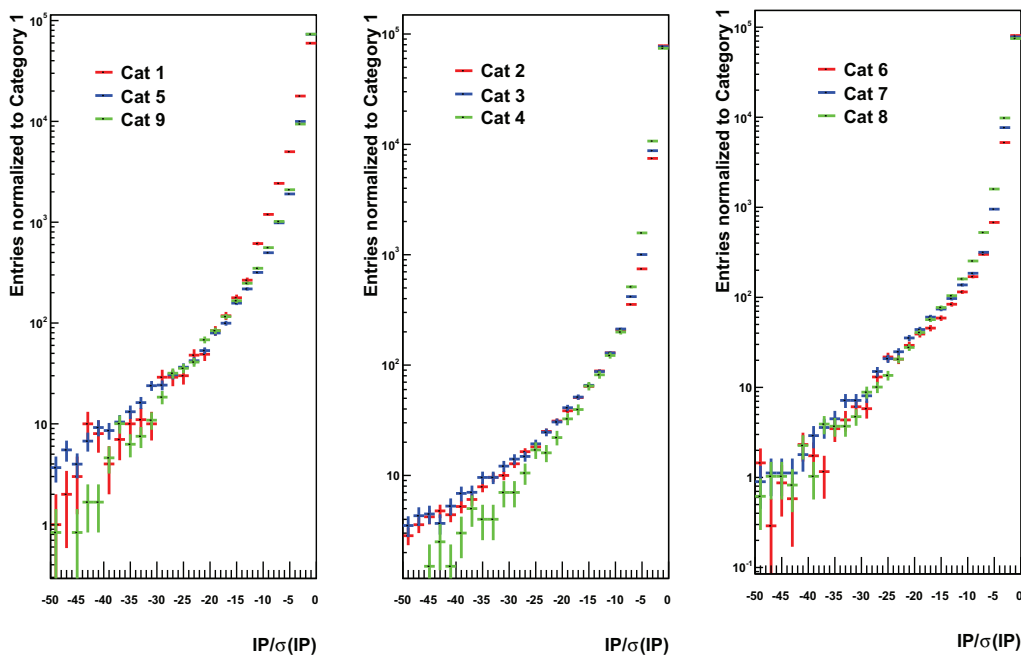
Figure 5.7: Distributions of the negative IP significance, for each category for the 2011 data. The number of entries is normalised to the entries in the first category.
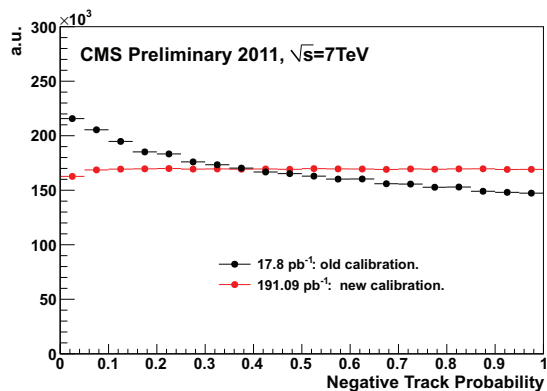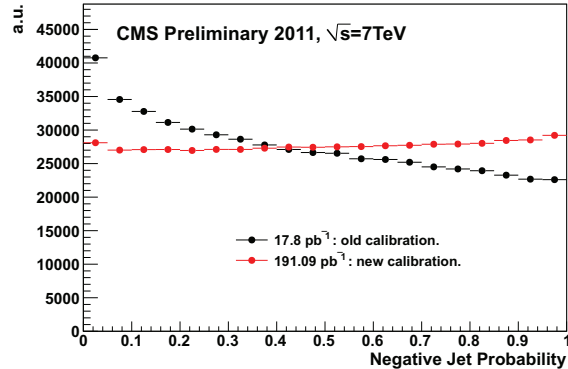

Figure 5.8: Distributions of the negative track probability for 2011 data: in black the distribution obtained using the 2010 calibration and in red with the 2011 calibration.

5.5.2 Calibration on MC

The number of pile-up vertices is not exactly reproduced in the MC simulation, so the MC distribution needs to be reweighted according to the data. This has been accomplished by dividing each $\hat{p}_t$ bin of the MC samples into five different PV multiplicity bins. The resulting calibrations are merged with their respective weights. More details about the Jet Probability calibration can be found in (65). Figure **5.10** shows the distribution of the track probability (left) and Jet Probability (right) for tracks with negative IP significance, for MC samples reweighted to 2011

Figure 5.9: Distributions of the negative jet probability for 2011 data. In black the distribution obtained using the 2010 calibration and in red is shown the distribution obtained with the 2011 calibration.
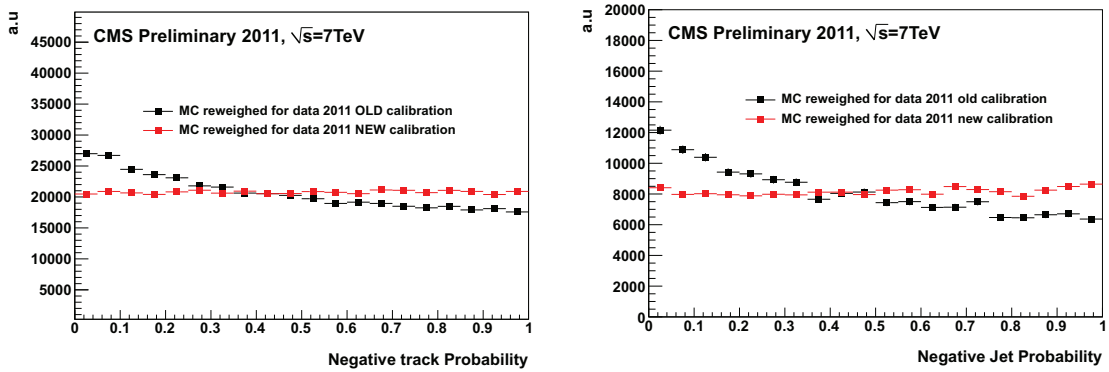
Figure 5.10: Track probability (left) and jet probability (right) for negative IP significance tracks for MC samples reweighted to 2011 data pileup.

.

data. The comparison between old and the new calibrations is displayed.

5.5.3 Jet probability discriminators

In addition to the Jet Probability tagger calculated by the Eq.(5.2), we define the JetBProbability. As the average charged particle multiplicity from b-hadron decay is about 5, and from the average track reconstruction efficiency, around 90% for tracks in jets, the JetBProbability estimates how likely the four most displaced tracks in a jet are compatible with the primary vertex.

To compute this tagger we apply some additional cut:

- ΔR (Tracks, jet axis) < 0.3 (this cut is applied only for the Jet Probability tagger);

- to avoid that a single track drives the whole jet probability, any track probability less than 0.005 is set to 0.005.

93

The Jet Probability discriminator is defined as: $-log(P_{jet}/4)$. In the same way we define the discriminator of the JetBProbability algorithm. In figure **5.11** the distribution of the discriminators are presented.
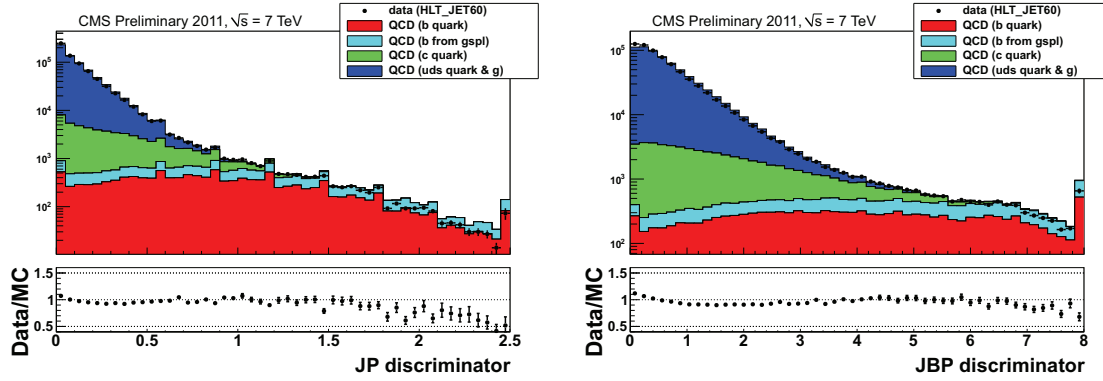

Figure 5.11: Discriminators for the Jet Probability (left) and JetBProbability (right) algorithms.

.

In figures **5.12, 5.13** the discriminators for Jet Probability and JetBProbability algorithms for three samples with different primary vertex multiplicities are shown. As we can see a proper calibration provides stable results in all bins of primary vertex multiplicities.
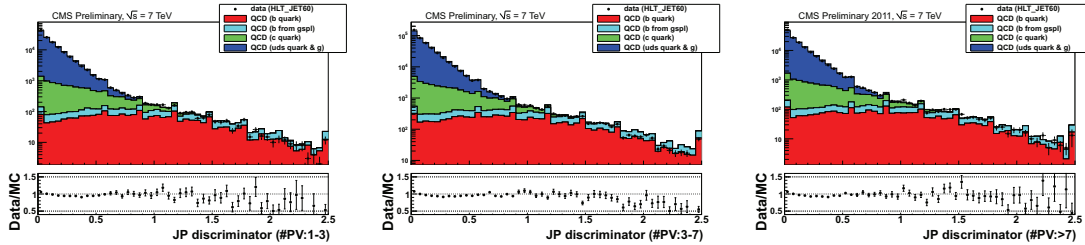
Figure 5.12: Data/MC comparison of the Jet Probability discriminator for three different primary vertex multiplicities: (left) NPV: 1-3, (middle) NPV: 4-7, (right) NPV: > 7.
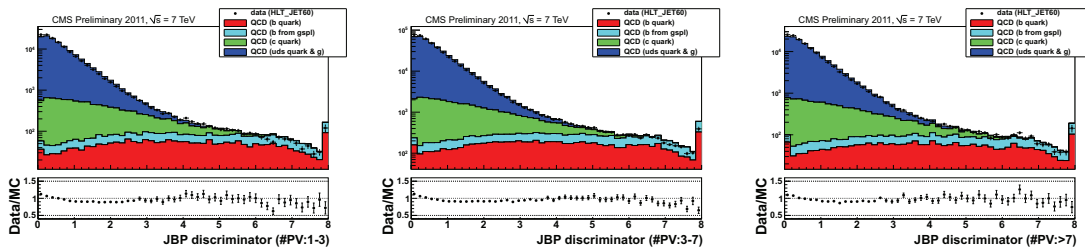
Figure 5.13: Data/MC comparison of the JetBProbability discriminator for three different primary vertex multiplicities.

5.6 Secondary Vertex algorithms

There are mainly two algorithms based on the direct reconstruction of the secondary vertex. The simplest one, called simple secondary vertex, returns a discriminator if a secondary vertex is found. Analogous to the IP based tagger, we consider two variants based on the number of tracks, N_{trk}, assigned to the vertex: $N_{trk} >= 2$ yields a high efficiency (SSVHE), requiring $N_{trk} >= 3$ corresponds to the high purity (SSVHP). In figure **5.14** the distributions of the discriminators (3D flight distance significance) for SSVHE and SSVHP are shown.
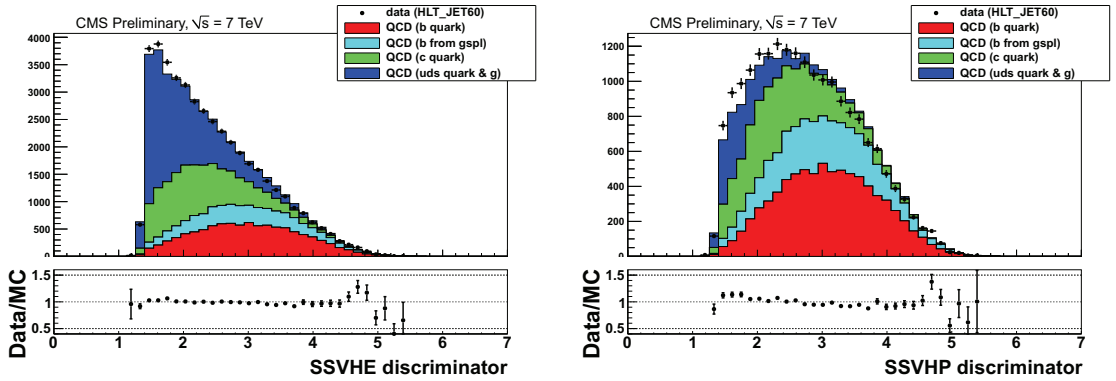
Figure 5.14: Discriminators for the Simple Secondary Vertex High Efficiency (left) and High Purity (Right) algorithms.
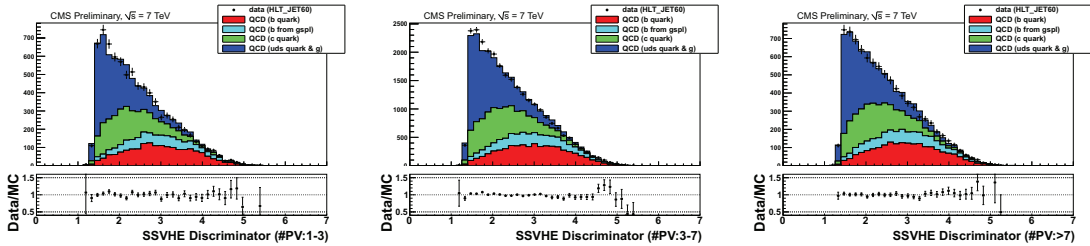

Figure 5.15: Data/MC comparison of the simple secondary vertex high efficiency discriminator for three samples characterized by different primary vertex multiplicities.
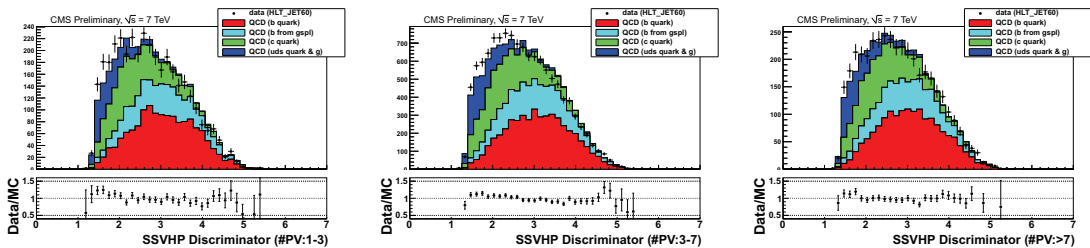

Figure 5.16: Data/MC comparison of the simple secondary vertex high purity discriminator for three samples characterized by different primary vertex multiplicities.

In figures **5.15, 5.16** the distributions of the discrimintators for the SSVHE and SSVHP algorithms, for three different primary vertex multiplicities, are presented. The agreement between data and simulation is stable for the three bins of primary vertex multiplicity.

This approach is limited by the efficiency of the reconstruction of the secondary vertex which is about 60-70%. A more complex approach combines the secondary vertex reconstruction with the information from the IP of the tracks. By using these additional variables, the Combined Secondary Vertex algorithm (CVS) provides discrimination even when no secondary vertices is found, so the maximum possible b-tagging efficiency is not limited by the secondary vertex reconstruction efficiency. In many cases, tracks with an IP significance > 2 can be combined in a so-called pseudo vertex, allowing for the computation of a subset of secondary vertex based quantities even without an actual vertex fit. When even this is not possible, a no vertex category reverts simply to track based variables similarly to the jet probability algorithm.

5.7 Performance of the tagger

5.7.1 Efficiency measurements

From the analysis point of view, it is crucial to know the efficiency of each b-tagging algorithm in order to select real b-jets. Several techniques are developed in CMS, that can be applied on data in order to reduce the dependence on simulations (66). One of these relies on the kinematic properties of the muon in jets. As we already mentioned, due to the large b-quark mass, the p_T of the muon relative to the jet axis (p_T^{rel}) is larger for b-jets than for the other flavor jets. Hence the technique consists in fitting the distribution of the p_T^{rel} of the muon in jets, using a simulated template for c, light and b jets. The template for light jets was validated on data as explained in (66). The efficiency is computed as the fraction of muon jets which satisfy the requirement of a specific tagger. Some other methods rely on the template fits to the muon-track impact parameter or on solving a system of eight equations where the b-tagging efficiency is one of the unknowns.

In a similar way we can use the properties of the Jet Probability tagger that has different distributions for different jet flavors at any value of jet p_T (67). The JP tagger also has the virtue of being calibrated directly in the data, where tracks with negative impact parameter can be used to compute the probability for those tracks to come from the primary vertex. This technique will be described in more details in the following.

Fits to the JP discriminant for the TC, SSV and CSV taggers. Thanks to its properties the JP algorithm can be used as reference tagger for estimating the fraction of b jets in a data sample, and also for estimating the fraction of b jets in a subsample that has been selected by an independent tagging algorithm. This allows a measurement of the efficiency of the independent algorithm. The efficiency measurement is performed in inclusive jet events in which at least one jet must be above a given p_T threshold, and separately in events in which at least one jet is a muon jet. To increase the fraction of b jets in the inclusive sample, an additional jet tagged by the JPM algorithm is also required. The sample with muon jets is already sufficiently enriched in b jets by the muon requirement. The same set of samples can be established with simulated events, so that the tagging efficiency can be measured there and a data-to-simulation scale factor computed. Because a value of the JP discriminant can be defined for jets that have as few as one track with a positive impact parameter significance, the discriminant can be calculated for most b jets, regardless of their p_T. The fraction of jets with JP information (before tagging) in Data and Monte Carlo is shown in figure **5.17** (top) for muon jets and inclusive jets as a function of the jet p_T.
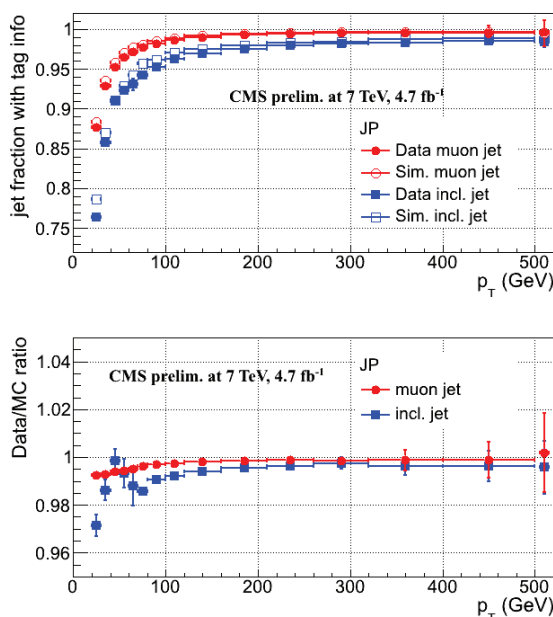

Figure 5.17: For muon jets and inclusive jets before tagging: (top) fraction of jets with JP information in data and MC; (bottom) data-to-MC ratio. The last bin includes all jets with $p_T > 500$ GeV.

In this figure, the last bin includes jets with $p_T > 500$ GeV and corresponds to an average jet p_T of 586 GeV. The data-to-MC ratio is given in the bottom plot: it remains close to one with at most a 3% difference observed at low p_T in inclusive jets. The fraction of b jets with JP information (before tagging) in the Monte Carlo,

C_b, is presented in figure **5.18** (top) for muon jets and inclusive jets as a function of the jet p_T. Whereas C_b is about 0.91 at $p_T = 20$ GeV, it rapidly raises up to more than 0.98 for $p_T > 50$ GeV. An illustration of fits (68) to the JP discriminant
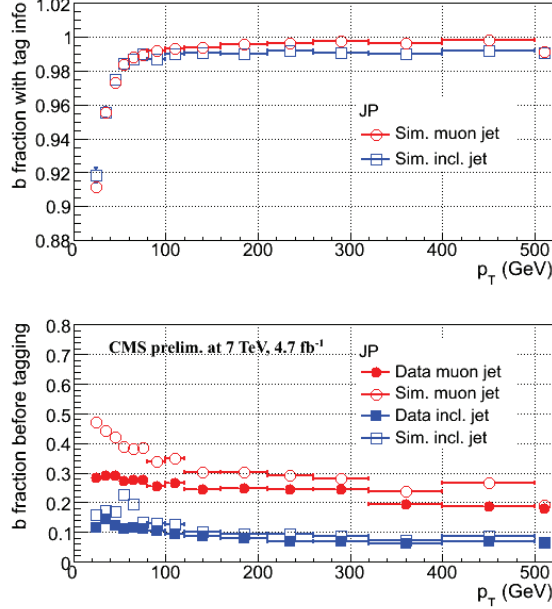

Figure 5.18: For muon jets and inclusive jets before tagging: (top) fraction C_b of MC b jets with JP information; (bottom) fitted fraction of b jets in the JP data distribution, compared to MC truth.

in the jet p_T bin 260-320 GeV is presented in figure **5.19** in the muon-jet sample (top panels) and in the inclusive jet sample (bottom panels), before (left panels) and after (right panels) applying a b-tagging criterion, here CSVM. The 2011 data are fitted to the sum of three MC shape distributions for b-, c- and light flavour jets. The normalisation of the relative flavour fractions (f_b, f_c, f_{light} respectively) is left free, with the constraint $f_b + f_c + f_{light} = 1$. The fit results, expressed as a fraction of b, c and light fractions are quoted in the inset legend of the figure, together with the expectation from MC truth (within parentheses). The fitted fraction of b jets in data, before tagging, is compared to the MC truth in figure **5.18** (bottom). The b-tagging efficiency can be computed as the ratio of the number of fitted b jets after and before tagging, correcting for the C_b fraction (Equ.6.5). The statistical uncertainty is calculated by taking into account the binomial terms (Equ.6.6):

$$\epsilon_b^{tag} = \frac{C_b \cdot f_b^{tag} \cdot N_{data}^{tag}}{f_b^{before\ tag} \cdot N_{data}^{before\ tag}} \tag{5.4}$$

$$= \frac{C_b \cdot f_b^{tag} \cdot N_{data}^{tag}}{f_b^{tag} \cdot N_{data}^{tag} + f_b^{untag} \cdot N_{data}^{untag}} \ . \tag{5.5}$$
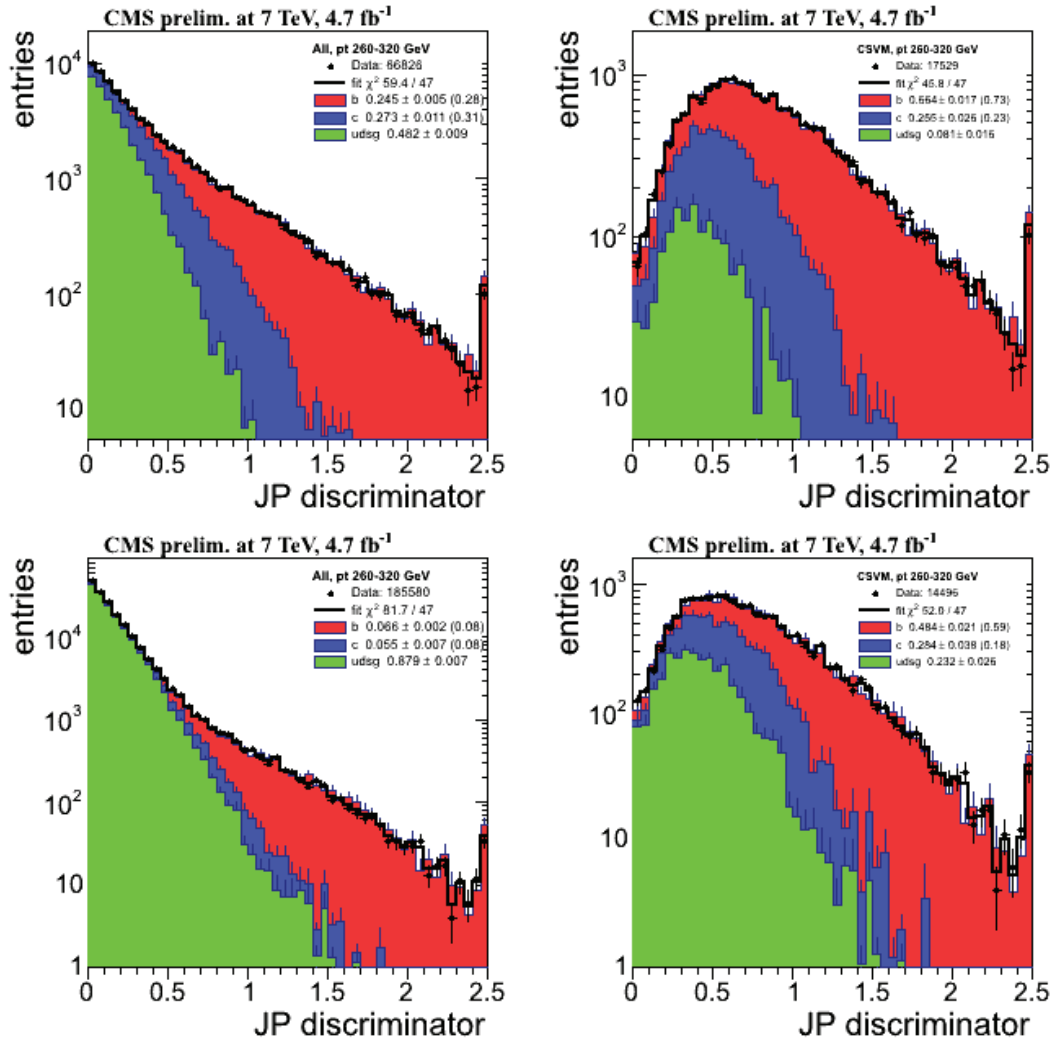
Figure 5.19: Fit to the JP discriminant in 2011 data with (top) a muon jet and (bottom) an inclusive jet with 260< p_T <320 GeV, (left) before and (right) after b-tagging with CSVM. Overflows are displayed in the upper right bin.

The measured b-tagging efficiencies are displayed in the upper panels of figure **5.20-5.23** for the TC, SSV and CSV taggers at their medium working point, for the muon jet (red dots) and inclusive jet (blue squares) samples. The fitted b-tagging efficiencies in data (close symbols) are compared to the MC expectation (open symbols) as a function of the b-jet p_T. The larger b-tagging efficiency in muon jets, especially at high p_T, is attributed to the higher probability for the presence of a $b\bar{b}$ pair due to gluon splitting inside the muon jet. Indeed an away tagged jet is requested for inclusive jets, which reduces the presence of close b-hadron decays, whereas no away tag is requested for muon-jet events.

The data-to-MC ratio is fitted with a straight line in the range $80 < p_T < 500$ GeV (lower panels) and the statistical error of the fit (dashed area) is superimposed. In general, the b-tagging efficiency from data agrees with the simuation within a few percent uncertainty in the overall p_T range.
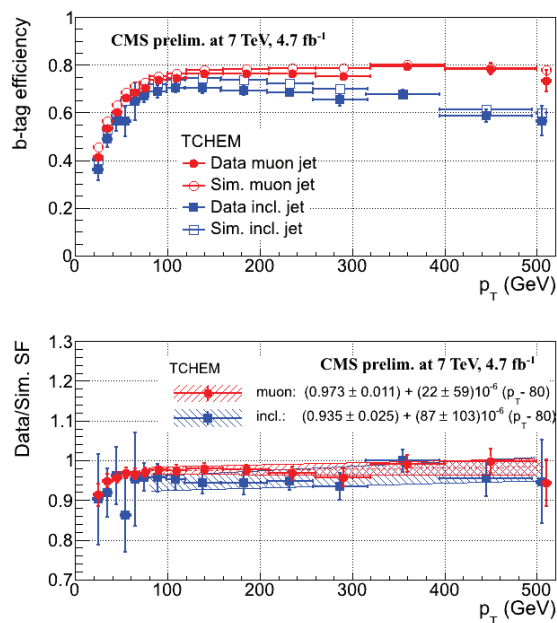

Figure 5.20: For muon jets (red dots) and inclusive jets (blue squares) with the TCHEM tagger: (top) b-tagging efficiency in data and MC; (bottom) data-to-MC scale factor.

Fits to the CSV discriminant for the TC, JP and JPB taggers. As the proposed method is based on fits to the Jet Probability discriminant, it cannot be applied to measure the b-tagging efficiency of the JP and JBP taggers. The Combined Secondary Vertex discriminant, which is mainly constructed on secondary vertex informations, could be used instead. As shown in figures **5.24-5.26**, the CSV discriminant has similar properties than for JP. However, contrary to JP, the CSV discriminant is built from MC and is not calibrated using data. The correlation due
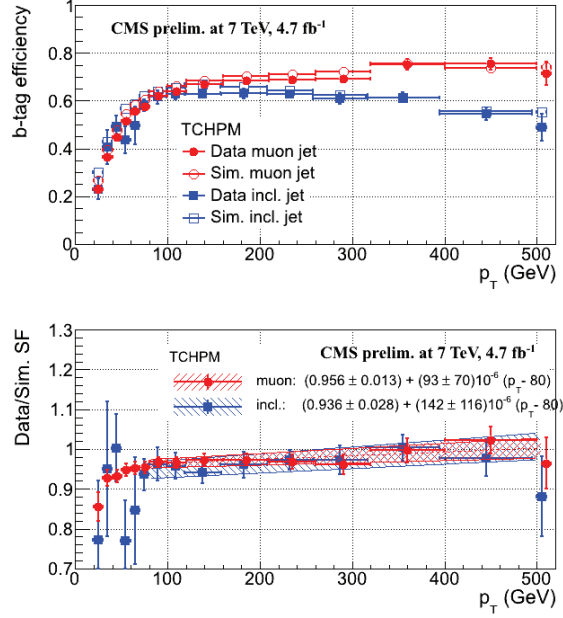
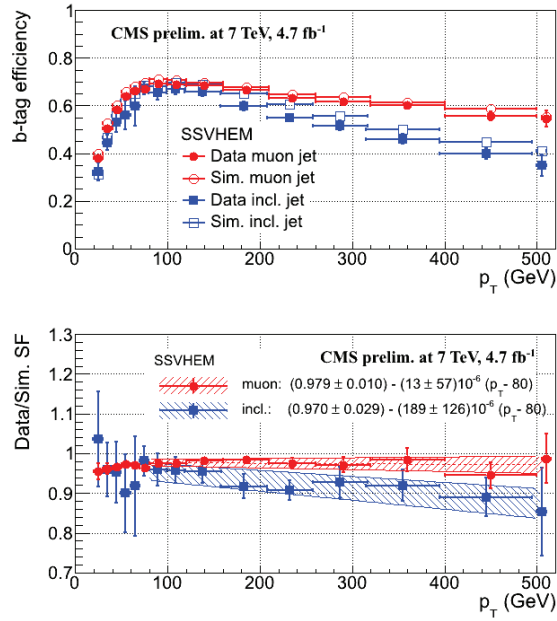**Figure 5.21:** Same as figure **5.20** for the TCHPM tagger.

Figure 5.22: Same as figure **5.20** for the SSVHEM tagger.
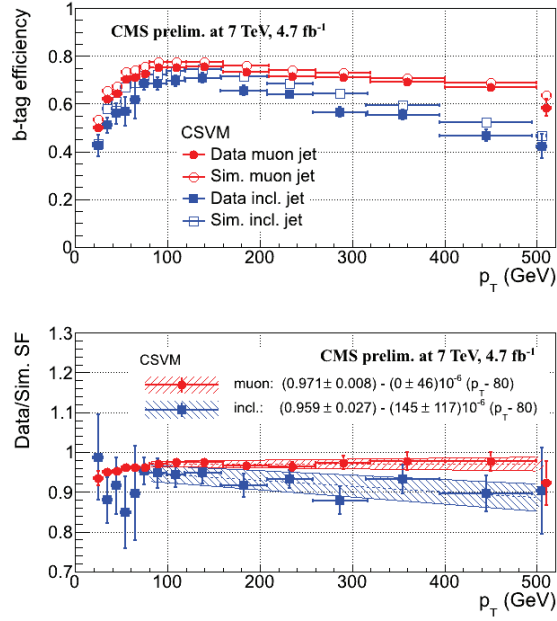
Figure 5.23: Same as figure **5.20** for the CSVM tagger.
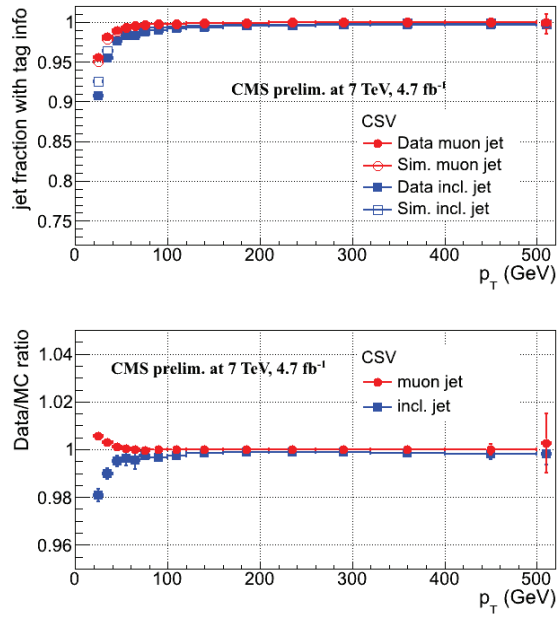

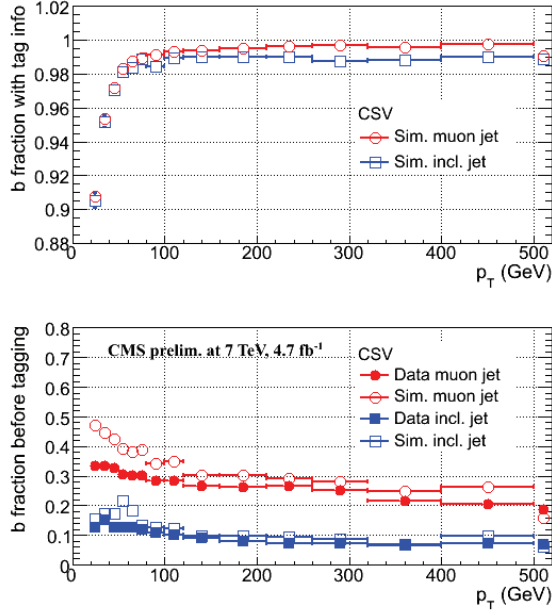**Figure 5.24:** Same as figure **5.17** for the CSV discriminant.

Figure 5.25: Same as figure **5.18** for the CSV discriminant.

to some common lifetime informations used in CSV and in JP/JBP may be larger than in the previous studies. We propose to first apply CSV fits to evaluate the TC b-tagging efficiencies and compare the results with the previous JP fits with TC, in order to quantify the bias. Then CSV fits can be applied to evaluate the JP and JBP b-tagging efficiencies and the final results on JP/JBP can be obtained by correcting from the bias (including the bias in an additional systematic uncertainty for JP/JBP).

About a 0-2%, 4-6%, 6-9% positive bias is observed for the loose, medium, tight operating points, respectively (see figures **5.27-5.29**)

Then the b-tagging efficiencies and scale factors are measured for the JP and JBP taggers (medium working point) in figures **5.30, 5.31**: the scale factors have to be corrected for the bias, but the overall behavior is similar to the other taggers.

Systematic uncertainties on the b-tagging efficiency Several sources of systematic uncertainties affect the measurement of the data-to-MC scale factor, SF_b, of the b-tagging efficiency at hight jet p_T.

- **Fraction of b jets with JP information:** as already explained, the number of b jets before tagging is measured by a fit to the JP distribution and corrected by the fraction C_b of b jets with JP information. A systematic uncertainty of half the residual correction, $(1 - C_b)/(2\ C_b)$, is estimated from the MC simulation as a function of the b-jet p_T. To estimate the b-tagging efficiency of the JP and JBP taggers, a fit to the CSV discriminant is performed instead.
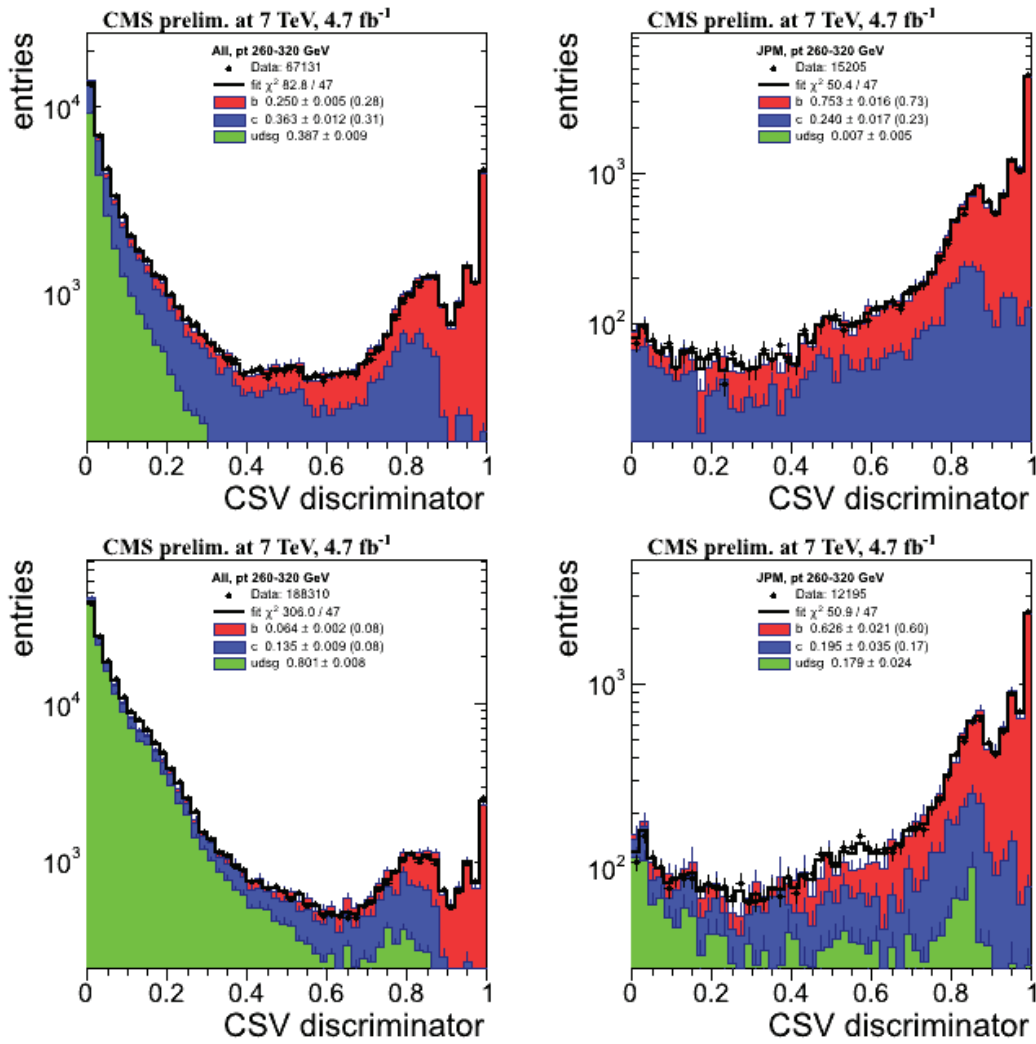
Figure 5.26: Same as figure **5.19** for the CSV discriminant and b-tagging with JPM.
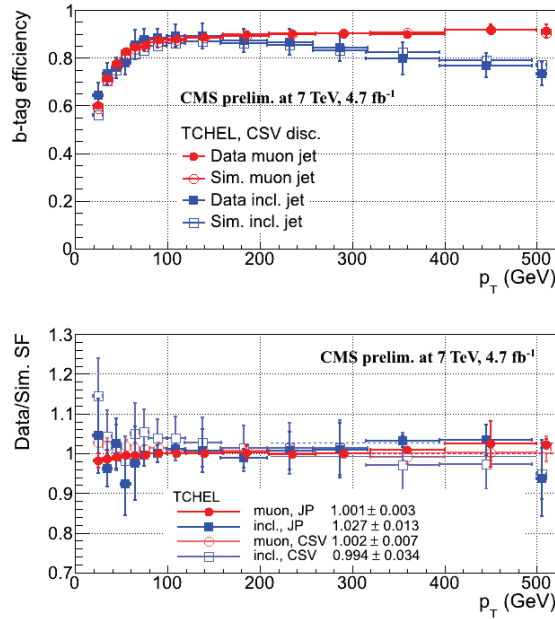
Figure 5.27: With the TCHEL tagger: (top) b-tagging efficiency in data and MC, using a fit to the CSV discriminant; (bottom) data-to-MC scale factor, using a fit to the CSV discriminant (open symbols) or to the JP discriminant (full symbols). The values quoted in the inset legend correspond to fits with a constant for $p_T > 210$ GeV (dotted lines).
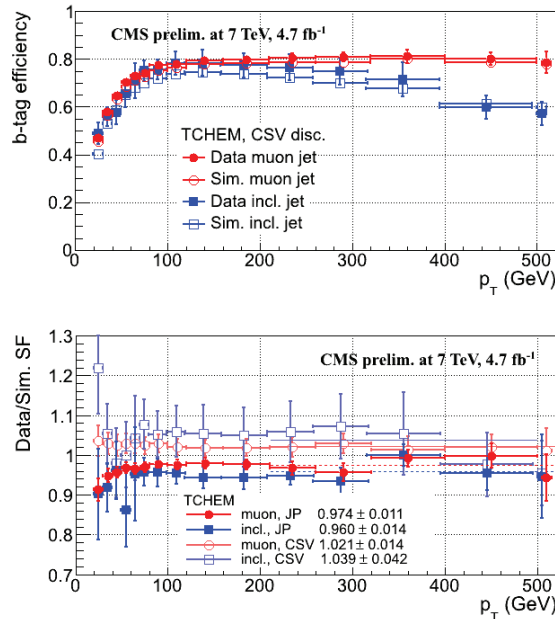

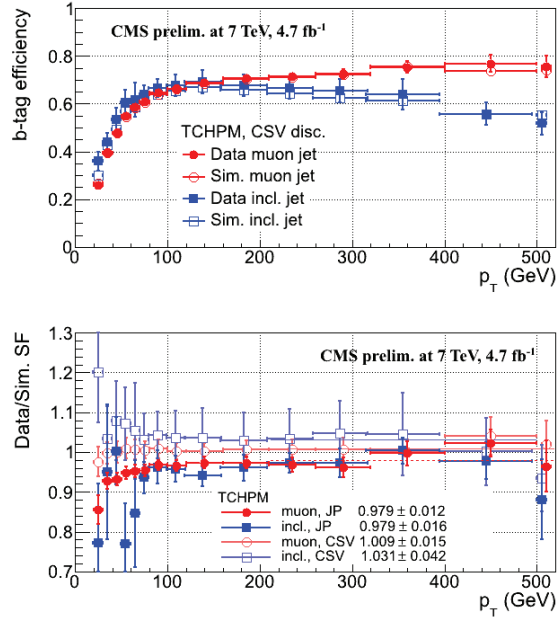**Figure 5.28:** Same as figure **5.27** for TCHEM tagger.

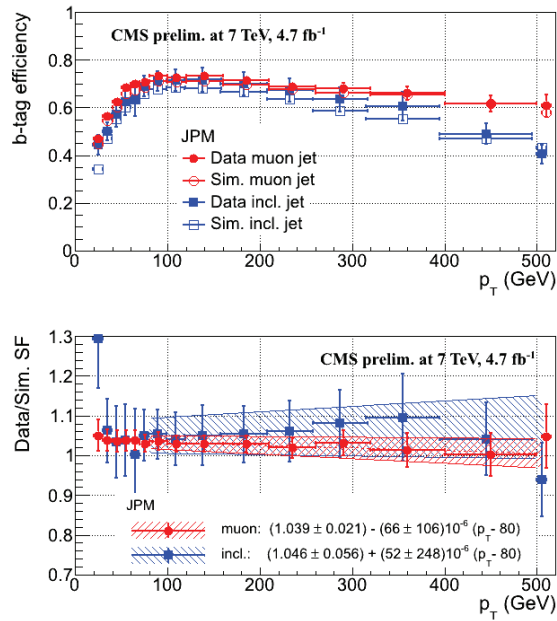**Figure 5.29:** Same as figure **5.27** for the TCHPM tagger.

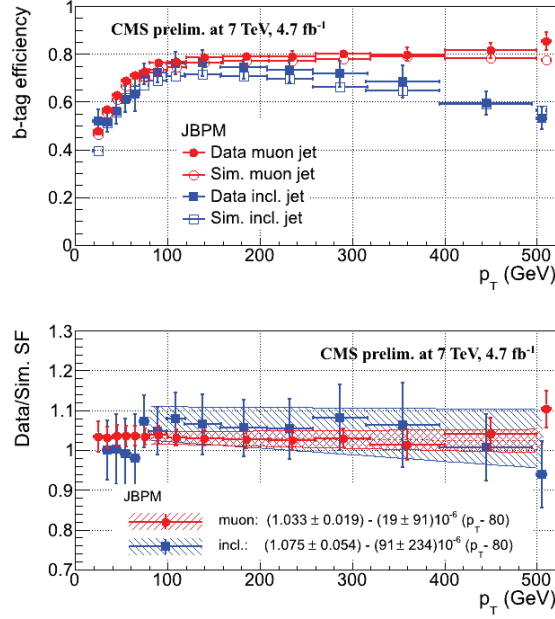**Figure 5.30:** Same as figure **5.20** for the JPM tagger.

Figure 5.31: Same as figure **5.20** for the JBPM tagger.

The corresponding C_b correction factor is computed with a similar systematic uncertainty. The order of magnitude of this error for the TC algorithm is of 4.4% for jet p_T between 20 and 30 GeV, and it decreases under the percent for jet with $p_T > 50$ GeV.

- **Difference between muon jets and inclusive jets:** the difference between the measured SF_b in muon jets and in inclusive b jets is taken as a systematic uncertainty. This is the largest contribution to the systematic uncertainty on SF_b with the JP-fit method. The order of magnitude of this error is between 5 and 25% for various algorithms (it changes for the different working points) for jet with $p_T < 50$ GeV and it decreases under 8% for high jet with $p_T > 80$ GeV.

- **Selection cut on the muon p_T:** the default selection on the muon transverse momentum is muon $p_T > 5$ GeV. This selection cut affects the shape of the light, c and b distributions used in the fit to the JP and CSV discriminants. In order to estimate a systematic uncertainty, the fits are repeated with muon $p_T > 9$ GeV and the observed difference in SF_b are considered. In order to reduce the statistical fluctuations, this systematic uncertainty is computed as a constant value for b-jet p_T in the ranges 20-80 GeV, 80-210 GeV and >210 GeV.

- **Gluon splitting:** gluon splitting into a charm-quark pair or into a bottom-quark pair affects the shape of the c and b distributions used in the fit to the

JP and CSV discriminants. The gluon-splitting rate is varied by $\pm 50\%$ in the MC simulation and the fits are repeated. The largest observed difference is taken as a systematic uncertainty in each b-jet p_T bin.

- **Generated pileup:** the pileup conditions have varied during the 2011 data taking as the instantaneous luminosity of the LHC was increasing. Each MC event is weighted in order to match the average in-time pileup distribution estimated in data. The data pileup is estimated within a $\pm 10\%$ uncertainty. First, new pileup distributions were obtained for data with an average pileup shifted by $\pm 10\%$. Then new pileup weights were computed for the MC and new shapes were inferred for the distributions of JP and CSV discriminants in simulated light, c and b jets. Fitting the JP and CSV distributions in data with these new shapes gives new values for the SF_b scale factor. The largest observed difference in SF_b is taken as a systematic uncertainty in each b-jet p_T bin.

- **Bias for the JP and JBP taggers:** A correction factor, computed as ratio between the scale factor obtained with JP and CVS algorithm ($C_{JP/CSV}$), is inferred for the loose, medium and tight operating points separately. The scale factors obtained for the JP and JBP taggers can then be multiplied by the $C_{JP/CSV}$ correction factor. The systematic uncertainty on $C_{JP/CSV}$ is taken as its difference with one. The order of magnitude of this systematic uncertainty is 1-10%.

- **Linearity of the template fits:** The template fits have been tested in the MC simulation by reducing the b-tagging efficiency of the simulated data by 20%. The observed SF_b scale factor between these pseudo-data and the MC expectation is found in excellent agreement with 0.8.

An overview of the systematic uncertainties entering the b-tagging efficiency scale factor for the different taggers in their medium working point, for the muon-jet p_T bin 80-120 GeV, is give in table **5.2**.

All methods to estimate the efficiency are then combined to get the overall Data/MC scale factor for the different algorithms. In figure **5.32** the individual (upper panel) and combined (lower panel) measurements of the Data/MC scale factor for the JPM algorithm is presented.

Efficiency measured on $t\bar{t}$ candidate events A probabilistic approach, has been used to estimate the efficiency of the b-tagging algorithms using the $t\bar{t}$ candidate events as it is well documented in (16), (69) . The main idea is to assign to each jet a probability to be a b-jet using the MC information. Then the probability
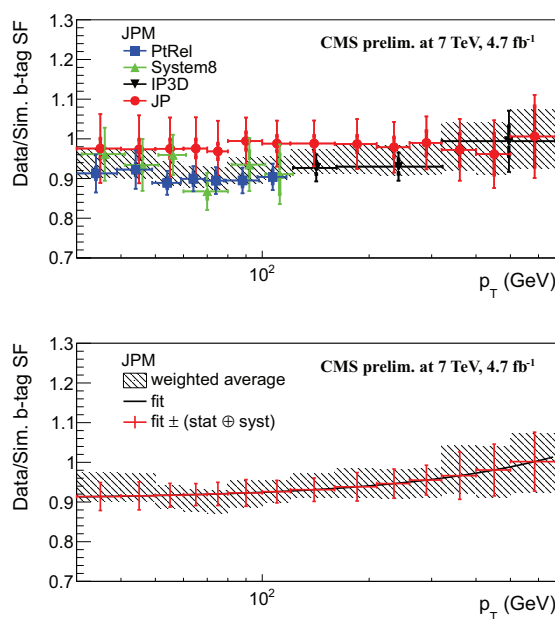
Figure 5.32: With the JPM algorithm: (upper panel) data-to-MC scale factor of the b-tagging efficiency as measured with the four methods used with muon-jet events, with (thick error bar) statistical error and (narrow error bar) overall statistical+systematic uncertainty; The combined SF_b value with its overall uncertainty is displayed as a hatched area; (lower panel) same combined SF_b value with the result of a fit function superimposed (solid curve). The overall statistical+systematic uncertainty is renormalized around the fit result (points with error bars).

Table 5.2: Relative systematic uncertainty on SF_b with the medium operating points of all taggers in the muon-jet p_T range 80-120 GeV.

b-tagger	C_b	inc. jets	p_T^μ	g split.	pileup	bias	total
JPM	0.1%	3.7%	0.8%	0.5%	0.1%	4.0%	5.5%
JBPM	0.1%	3.6%	0.9%	0.6%	0.1%	4.0%	5.5%
TCHEM	0.4%	2.8%	0.3%	0.8%	0.0%	—	2.9%
TCHPM	0.4%	2.0%	0.0%	0.2%	0.1%	—	2.1%
SSVHEM	0.4%	2.0%	0.3%	0.4%	0.0%	—	2.1%
CSVM	0.4%	2.6%	0.4%	0.8%	0.1%	—	2.8%

of all jets in the events are combined in order to assign to the events a probability to contains at least n b-tagged jets.

The individual results coming from the different methods to estimate the b-tagging efficiency and the Data-MC scale factor in muon-jet events have been combined to provide an optimal measurement as a function of the jet p_T in the range $30< p_T <670$ GeV. The combination, finally, was compared with the results coming from the studies on the b-tag efficiency done on $t\bar{t}$ candidate events. In order to allow for a comparison, the p_T dependent scale factors measured in and multi- jet events have been reweighted to match the jet-p_T spectrum observed in $t\bar{t}$ events. Table **5.3** gives the values and total uncertainties for the efficiency scale factors SF_b obtained in multi-jet and $t\bar{t}$ events for b jets in the expected p_T range of $t\bar{t}$ events. All results in the table are for the tagger with their medium working point. An excellent agreement between the two estimations is observed.

Table 5.3: Values and total uncertainties for the efficiency scale factors SF_b obtained in multi-jet and $t\bar{t}$ events for b-jets in the expected p_T range of $t\bar{t}$ events. For the $t\bar{t}$ results with the JP and JBP algorithms the profile likelihood ratio values (70) are quoted as they correspond to the same calibration as for the multijet results.

b tagger	SF_b in multijet events	SF_b in $t\bar{t}$ events
JPM	0.92 ± 0.03	0.95 ± 0.03
JBPM	0.92 ± 0.03	0.93 ± 0.04
TCHEM	0.95 ± 0.03	0.96 ± 0.04
TCHPM	0.94 ± 0.03	0.93 ± 0.04
SSVHEM	0.95 ± 0.03	0.96 ± 0.04
CSVM	0.95 ± 0.03	0.97 ± 0.04

5.7.2 Mistagging estimation

The mistagging[1] rate has also to be evaluated for each b-tagger (66).The measurements of the mistagging rate from light-flavor jets relies on the definition of negative discriminator values for each b-tagging algorithm. These negative taggers can then be used in the same way as the regular b-tagging algorithms both in data and in the simulation. As the negative-tagged jets are enriched in light flavors, the mistagging rate can be measured from the data, with the simulation used to extract a correction factor. The mistagging rate is evaluated from tracks with a negative impact parameter or from secondary vertices with a negative decay length. When a negative tagger is applied to jets of any flavor, the corresponding tagging efficiency is denoted *negative tag rate*.

The mistagging rate is evaluated as:

$$\epsilon_{data}^{mistag} = \epsilon_{data}^{-} * R_{light}. \tag{5.6}$$

whereϵ_{data}^{-}data is the negative tag rate as measured in jet data, defined as the fraction of jets that are negatively tagged. $R_{light} = \frac{\epsilon_{MC}^{mistag}}{\epsilon_{MC}^{-}}$ is a correction factor taken from simulation. It represents the ratio of the mistagging rate for light-flavor jets to the negative tag rate for jets of all flavors in the simulation.

To compare the measured mistagging rate to that predicted by the simulation, a scale factor SF_{light} is defined as:

$$SF_{light} = \frac{\epsilon_{data}^{mistag}}{\epsilon_{MC}^{mistag}}. \tag{5.7}$$

The measured mistagging rate and data/simulation scale factor with the corresponding relative systematic uncertainty are presented in figure **5.33** as a function of the jet p_T for the JPM tagger. The observed scale factors are close to one over a broad range of p_T and $|\eta|$.

As an illustration, the data mistagging rates and data/simulation scale factors are given in table **5.4** for jets with p_T between 80 and 120 GeV.

Systematic uncertainty on the mistagging estimation The following sources of systematic uncertainties on the mistagging rate are considered:

- **b and c fractions**: the fraction of b-flavour jets has been measured in CMS to agree with the simulation within a $\pm 20\%$ uncertainty. A $\pm 20\%$ uncertainty is conservatively estimated for the overall fraction of b and c jets. The b+c flavour fraction is varied in the QCD Monte Carlo, from which a systematic uncertainty on R_{light} is inferred.

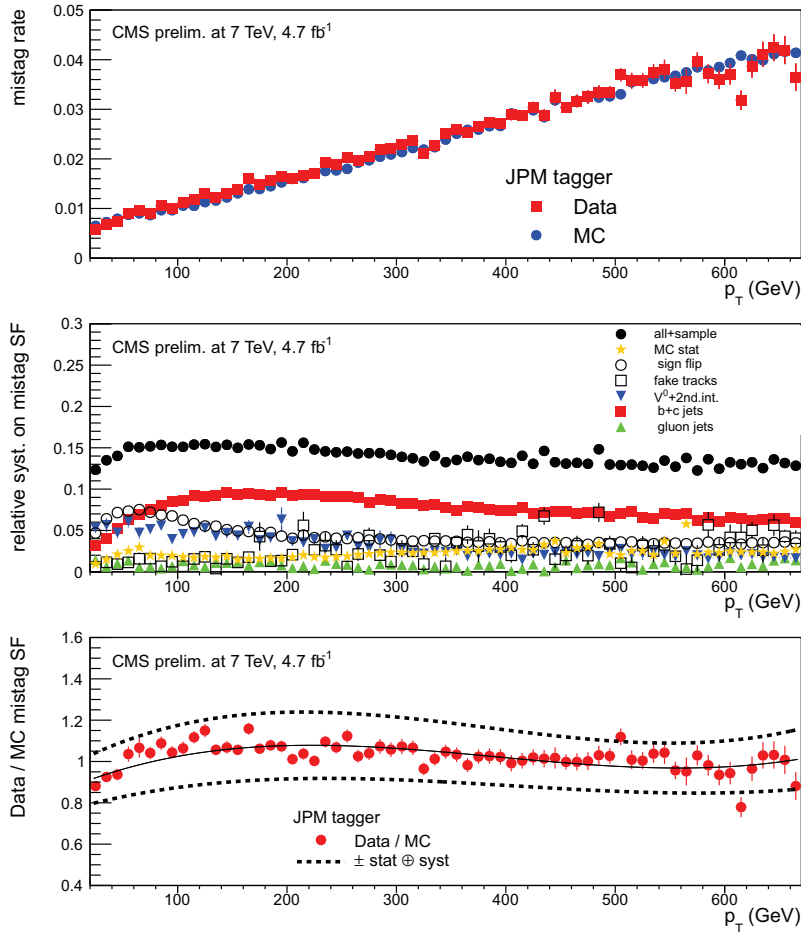[1]With *mistag* we refer to the percent of light jets misidentified as a b jet.

Figure 5.33: For the JPM tagger as a function of the jet p_T (the last p_T bin includes all jets with $p_T > 670$ GeV): (top) mistagging rate in data and MC. (middle) relative systematic uncertainties on the data-to-MC scale factor; (bottom) data-to-MC scale factor of the mistagging rate. The solid curve is the result of a fit to the data with a polynomial function, the dashed curves represent the overall statistical and systematic uncertainties on the measurements.

Table 5.4: Mistagging rate and data-to-simulation scale factor for the medium operating points for jet p_T in the range 80 - 120 GeV. The statistical errors are quoted for the mistagging rates and the statistical+systematic uncertainties for the scale factors.

tagger	mistag rate ($\pm stat$)	scale factor ($\pm stat \pm syst$)
JPM	0.0109 ± 0.0002	$1.02 \pm 0.02 \pm 0.16$
JBPM	0.0112 ± 0.0001	$0.94 \pm 0.01 \pm 0.11$
TCHEM	0.0286 ± 0.0003	$1.20 \pm 0.01 \pm 0.14$
TCHPM	0.0306 ± 0.0003	$1.24 \pm 0.01 \pm 0.12$
SSVHEM	0.0209 ± 0.0002	$0.93 \pm 0.01 \pm 0.08$
CSVM	0.0152 ± 0.0002	$1.10 \pm 0.01 \pm 0.11$

- **Gluon fraction**: this affects both the mistagging rate in simulation and the overall negative tag rates. The average fraction of gluon jets depends on the details of the parton density and hadronisation functions used in the simulation. An uncertainty of $\pm 20\%$ is extracted from the comparison of simulation with data.

- **Long lived K_S^0 and Λ decays**: the amount of reconstructed K_S^0 and Λ are found to be a factor 1.40 ± 0.15 and 1.50 ± 0.50 larger, respectively, in the data than in the simulation, with the quoted uncertainty accounting for the p_T dependence. To estimate the uncertainty on R_{light} due to the K_S^0 and Λ contribution, the simulated QCD events are reweighted in order to match the observed yield of K_S^0 and Λ in the data. Then this yield is varied within the quoted uncertainty and the inferred variation on R_{light} taken as a systematic uncertainty.

- **Photon conversion and nuclear interactions**: the rate of secondary interactions in the pixel detector layers has been measured with $\pm 5\%$ precision. The corresponding variation implies a systematic uncertainty on R_{light}.

- **Mismeasured tracks**: according to the simulation, jets with a reconstructed track not associated with a genuine charged particle also present an excess of positive over negative tags. To correct for residual mismeasurement effects, a $\pm 50\%$ variation on this contribution is taken into account in the systematic uncertainty on R_{light}.

- **Sign flip**: The ratio of the number of negative over positive tagged jets is computed in a muon-jet sample with a larger than 80% b purity. Data and simulation are found to be in good agreement. From the statistical uncertainty on the comparison, the absolute uncertainty on this ratio is estimated as 2% (1%, 0.5%) for loose (medium, tight) operating points, respectively. This sign flip uncertainty can be translated into a systematic uncertainty on R_{light}.

- **Pile-up:** the mistag rate depends on the pileup used in the MC simulation. However the simulated events are already reweighted according to the observed number of pileup in data. Differences between R_{light} values obtained for different running periods are used to estimate the systematic uncertainty, which is about 1% for all taggers.

- **Event sample:** physics analyses use jets from different event topologies. For a given jet p_T, the mistagging rate is different if the jet is the leading one or if there are other jets with higher p_T values in the same event. Measured mistagging scale factors for leading and sub-leading jets have a dispersion of about 7%. In addition, mistagging scale factors vary by 2-7%, depending on the tagger, for different running periods. These two uncertainties are added in quadrature to account for an uncertainty due to sample dependence. This is the dominant contribution to the overall systematic uncertainty on the mistagging rate.

In table **5.5** an overview of the systematic uncertainty on the mistag scale factor, SF_{light}, for all taggers in their medium working point, for the muon jet p_T bin 80-120 GeV, is shown.

Table 5.5: Relative systematic uncertainties on SF_{light} for jet in the range 80 - 120 .

b-tagger	b + c jets	gluon	$V^0 +$ 2nd int.	mis-meas.	sign flip	MC stat	sample + pile-up	all
JPM	8.6%	0.8%	4.7%	1.0%	6.4%	0.9%	9.4%	15.2%
JBPM	6.2%	1.2%	3.9%	0.5%	1.6%	0.9%	9.0%	11.9%
TCHEM	4.5%	0.8%	3.6%	1.2%	5.1%	0.7%	8.0%	11.3%
TCHPM	1.6%	1.0%	1.8%	0.6%	2.5%	0.6%	9.2%	10.0%
SSVHEM	1.0%	0.9%	2.4%	1.9%	2.9%	0.7%	7.3%	8.7%
CSVM	3.2%	1.8%	3.0%	0.7%	4.6%	0.7%	7.4%	10.1%

5.8   Conclusions

During the entire PhD period the different b-tag algorithms have been studied, with particular attention to the calibration of the Jet Probability algorithm for which I'm in charge since 2009. This is one of the most efficient algorithm and this is why it was chosen for the measure of the $t\bar{t}$ cross section in the τ_h + jets final state that will be presented in the next chapter. This is the first analysis using the Jet Probability algorithm to tag the b-jet in the final state and thanks to the high performance of this algorithm the CMS b-tagging group intend to use it as one of the main algorithms for the analyses of the 2012 data.

Chapter 6

Measurement of the top-antitop production cross section in the hadronic tau+jets final state

6.1 Introduction

In this chapter the first measurement of the top-antitop ($t\bar{t}$) production cross section in the hadronic tau plus jets final state in proton-proton collisions at $\sqrt{s} = 7$ TeV in the CMS experiment, is presented. Previous experiments at the Tevatron have measured the $t\bar{t}$ production cross section in the tau plus jets final state in proton-antiproton collisions at $\sqrt{s} = 1.96$ TeV as described in references (71) and (72).

The $t\bar{t}$ quark pairs are copiously produced at the LHC mainly via strong interaction through gluon-gluon fusion. The measurement of the $t\bar{t}$ production cross section and branching ratios (BR) is an important test of the Standard Model (SM), since the top quark is expected to play a special role in various extensions of the Standard Model due to its high mass (see chapter 1). Inversely, if not involved, the $t\bar{t}$ pairs are expected to be a large background for new physics searches.

The branching ratio of a top quark to a W boson and a b quark is close to 100% in the SM. The final states of the top quark decays are hence driven by the decay mode of the W boson. About 9.8% of the produced top-antitop quark pairs lead to a hadronic tau plus jets final state.

A charged Higgs boson could give rise to an enhanced cross section. The top quark would decay via $t \to H^{\pm} + b$ and the charged Higgs boson in turn via $H^{\pm} \to \tau\nu_{\tau}$. Such a deviation from the SM expectation would not necessarily be observable in the most studied $t\bar{t}$ decay channels.

The largest background for this analysis is due to the high multiplicity jet events where one of the jets fakes an hadronic tau. Given the low expected signal over background ratio, a neural network (NN) technique is used to discriminate the signal from the multijet background. A data driven approach is used to model the multijet

background while for the other electroweak processes (W+jets, Z+jets and single top) and remaining $t\bar{t}$ final states we rely on simulated events. The cross section is extracted using a fit to the NN output distribution.

This chapter is structured as follows: in section **6.2** the commissioning of the developed trigger and data samples are described. In section **6.3** the event preselection is discussed. Section **6.4** describes the design of the neural network (NN) used to enhance the signal over background ratio while section **6.6** explains the cross section extraction.

6.2 Trigger and datasets

6.2.1 Trigger design

As already described in section **3.4** a dedicated multijet trigger requiring the presence of four calorimeter jets, including tau identification was developed to record pp $\rightarrow t\bar{t} \rightarrow \tau_h$+jets events. With increasing instantaneous luminosity the L1 and HLT filters became more selective, leading to two versions of the trigger:

- HLT_QuadJet40_IsoPFTau40

- HLT_QuadJet45_IsoPFTau45.

The L1 decision of the trigger is based on the presence of four L1 jets with $p_T > 20$ GeV ($p_T > 28$ GeV for a luminosity of 2E33 $cm^{-2}s^{-1}$). At HLT, the presence of 4 jets with $p_T > 40/(45)$ GeV in $|\eta| < 2.5$, where one of the jets has to be matched with an HLT τ is required.

Due to the high cut on the tau transverse momentum, the trigger signal efficiency is expected to be small. Requiring offline the presence of 3 central jets and one hadronic tau with momenta above 45 GeV, one expects a signal efficiency of the order of $0.54 \cdot 0.027 \sim 0.015$ (where 0.54 is the efficiency to select 3 central jets with $p_T > 45$ GeV, and 0.027 is the efficiency to select one tau with $p_T > 45$ GeV). Convoluting the offline selection with the trigger efficiency measured in section **6.2.2**, ~ 0.64, the signal trigger efficiency for pp$\rightarrow t\bar{t} \rightarrow \tau_h$+jets is estimated to be of the order of 9 per mille. Figure **6.1** shows the reconstructed hadronic tau and jet momenta for simulated $t\bar{t} \rightarrow \tau_h$+jets events.

6.2.2 Trigger performance

The direct efficiency measurement of the HLT_QuadJet40(45)_IsoPFTau40(45) trigger is unfeasible due to the large QCD multijet background and to the difficulty to select a clean τ_h + 3 jets sample in data. The efficiency of the trigger can nevertheless be evaluated by measuring the efficiencies of the "tau part" and "jet part" separately.
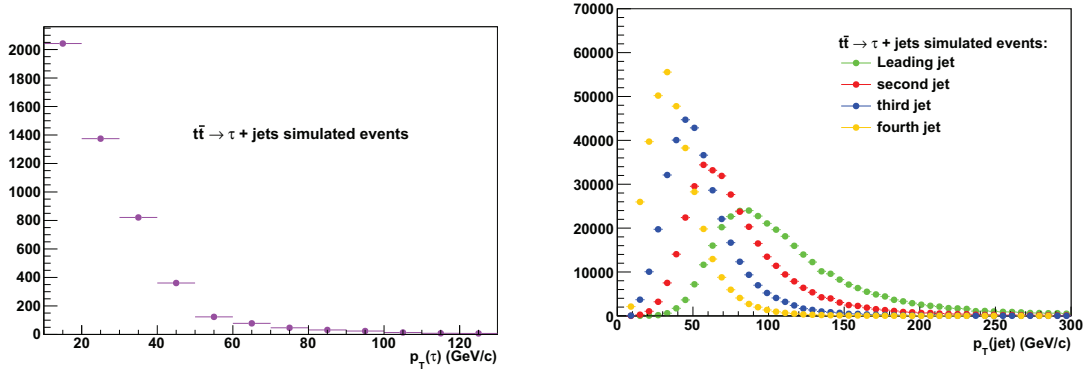
Figure 6.1: Reconstructed hadronic tau and jet transverse momenta for simulated $t\bar{t} \to \tau_h$+jets events.

The jet-leg efficiency has been measured in events fired by a single muon trigger (HLT_mu15, HLT_mu20, HLT_mu24, HLT_mu30) and containing four particle flow jets (PF jets) in the central region $|\eta| < 2.5$, three of them having $p_T > 70$ GeV and matched to trigger jets used in the HLT-jet-filter, in order to ensure that they satisfy the HLT-jet-filter requirements. The fourth jet is used as a probe jet and
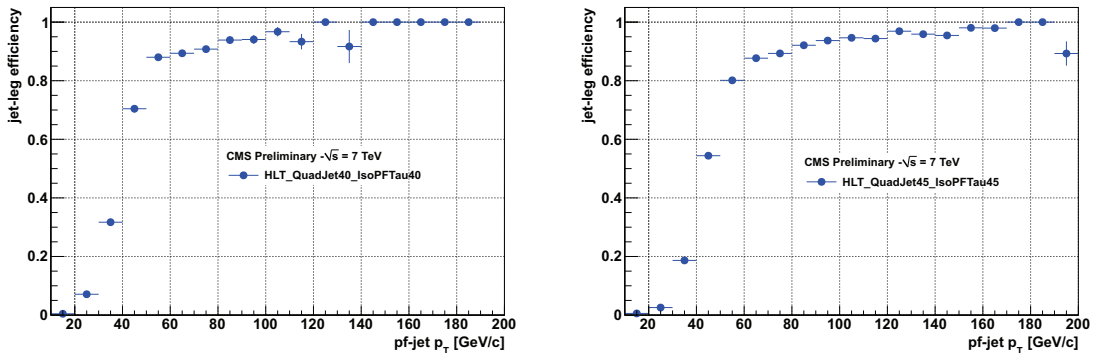
Figure 6.2: Jet-leg efficiency of the HLT_QuadJet40_IsoPFTau40 trigger (left) and HLT_QuadJet45_IsoPFTau45 trigger (right) measured in data with respect to the transverse momentum of PF jets.

its efficiency is computed with respect to the HLT jet filter response. Figure **6.2** shows the obtained efficiency turn-on per jet-leg for HLT_QuadJet40_IsoPFTau40 and HLT_QuadJet45_IsoPFTau45 triggers. The late reach of the plateau is due to the use of calorimeter jets online whereas PF jets are used offline. In figure **6.3** the jet-leg efficiency for calorimeter jets is compared to the one obtained with respect to PF jets offline.

Figure **6.4** shows the efficiency measured for the tau-leg in events of the *MultiJet* primary dataset. The events are required to contain four PF jets (not cleaned from taus) matched to the HLT-jets used in the HLT-jet-filter in order to ensure that the HLT-jet-filter has fired. In addition the events are required to contain only one HPS
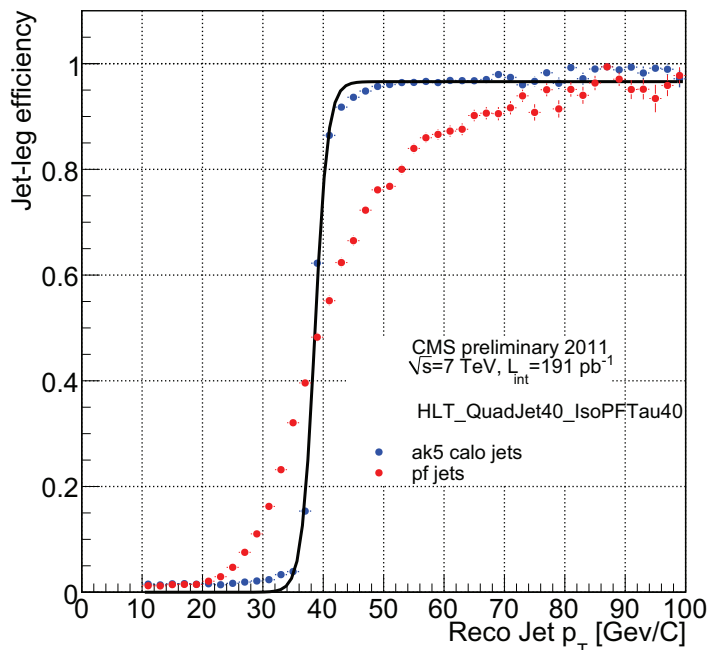
Figure 6.3: Jet-leg efficiency of the HLT_QuadJet40_IsoPFTau40 trigger for calorimeter and particle flow jets as function of their p_T.

tau identified with combined medium isolation matched to one of the four previously selected PF jets. The efficiency is computed with respect to the matching to the HLT-tau used in the HLT-tau-filter with positive decision.
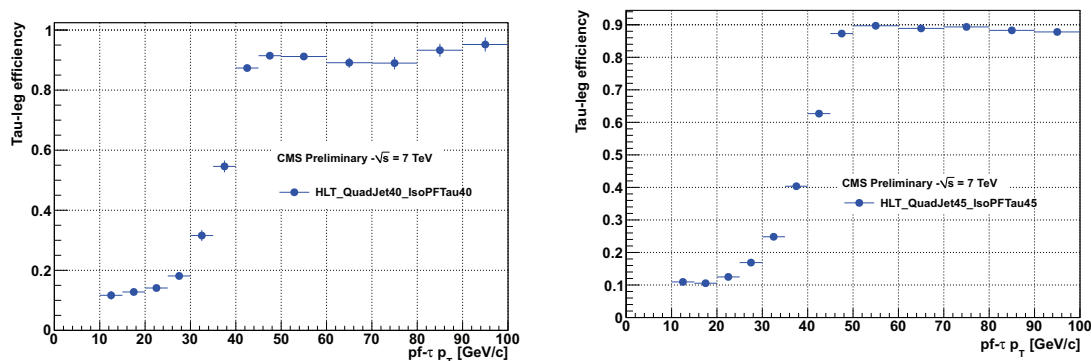

Figure 6.4: Tau-leg efficiency of the HLT_QuadJet40_IsoPFTau40 trigger (left) and HLT_QuadJet45_IsoPFTau45 trigger (right) measured in data with respect to the transverse momentum of the HPS taus selected with combined medium isolation.

For simulated events the trigger decision is emulated using the efficiencies measured previously. No direct trigger emulation is available in simulation for the HLT_QuadJet45_IsoPFTau45 trigger. Three central PF jets $|\eta| < 2.4$, ordered in p_T, (tau/jet cleaning required) are considered and the trigger weight computed in

figure **6.2** is applied on each jet depending of its p_T. In addition the presence of one PF tau with $p_T > 45$ GeV, $|\eta| < 2.3$ and combined medium isolation is required, and weighted by the efficiency measured in figure **6.4** to mimic the trigger response. All four weights are multiplied to model the trigger efficiency. Events are randomly weighted by the HLT_QuadJet40_IsoPFTau40 or HLT_QuadJet45_IsoPFTau45 trigger efficiency according to the integrated luminosity fraction of each trigger.

For recorded data the reconstructed offline tau and jets are required to match in $\Delta R < 0.4$ the tau and jet objects used in the trigger in order to model the same acceptance than the one used in simulation.

In figure **6.5** the efficiency of the tau-leg for the HLT_QuadJet45_IsoPFTau45 trigger is shown for Run2011A (left) and Run2011B (right). The mean number of pileup events increased on average from 6 to 10 between the two data taking periods. The efficiency is stable thanks to the tight criteria used online for the tau isolation.
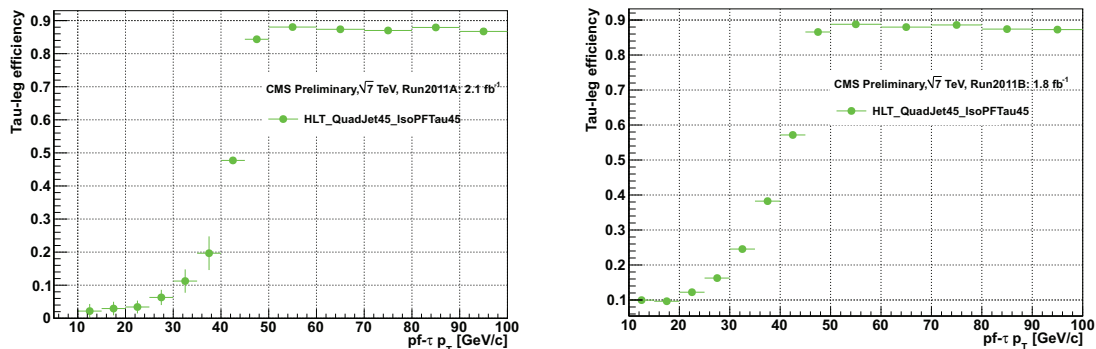

Figure 6.5: HLT_QuadJet45_IsoPFTau45 tau-leg efficiency for the two different data taking periods: Run2011A (left), Run2011B (right) as function of the τ p_T.

In figures **6.6, 6.7, 6.8** we see the registered online rates for three different runs (160873, 160956, 167041) in which the HLT_QuadJet40_IsoPFTau40 trigger was used. These three runs correspond to three different luminosities ($0.3 \cdot 10^{32}$, $0.13 \cdot 10^{33}$, $0.7 \cdot 10^{33}$ cm^{-2}.s^{-1}). The rate is not affected by the different pileup conditions since it grows up linearly with the luminosity.

6.2.3   Datasets and simulation

The HLT_QuadJet40(45)_IsoPFTau40(45) trigger is part of the so called *MultiJet* primary dataset. Table **6.1** summarizes the chosen trigger for each data taking period and corresponding integrated luminosity. The QuadJet40_IsoPFTau40 trigger was prescaled by mistake for runs 165970-166782, the HLT_QuadJet45_IsoPFTau45 trigger is used instead. The entire Run2011A (corresponding to the first half part of
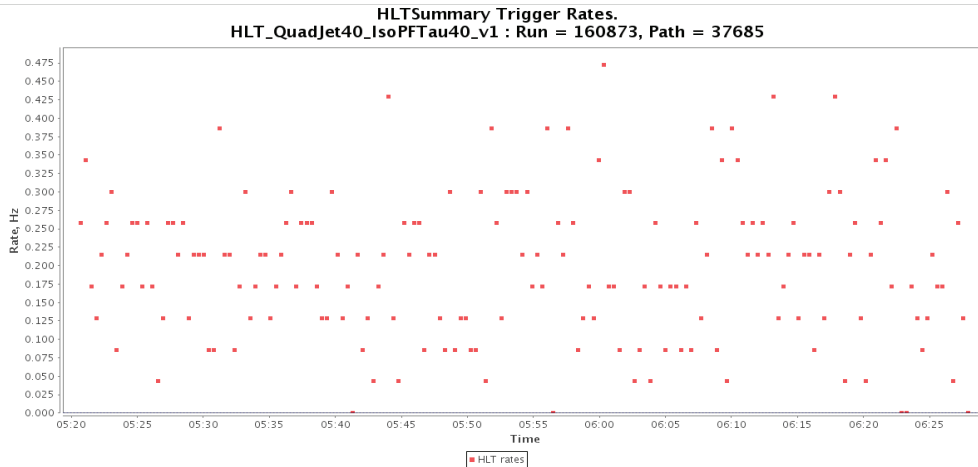
Figure 6.6: Rate for the HLT_QuadJet40_IsoPFTau40 trigger in run 160873 corresponding to a luminosity of $0.3 \cdot 10^{32}$ cm^{-2}s^{-1}. Mean recorded rate: 0.2 Hz.
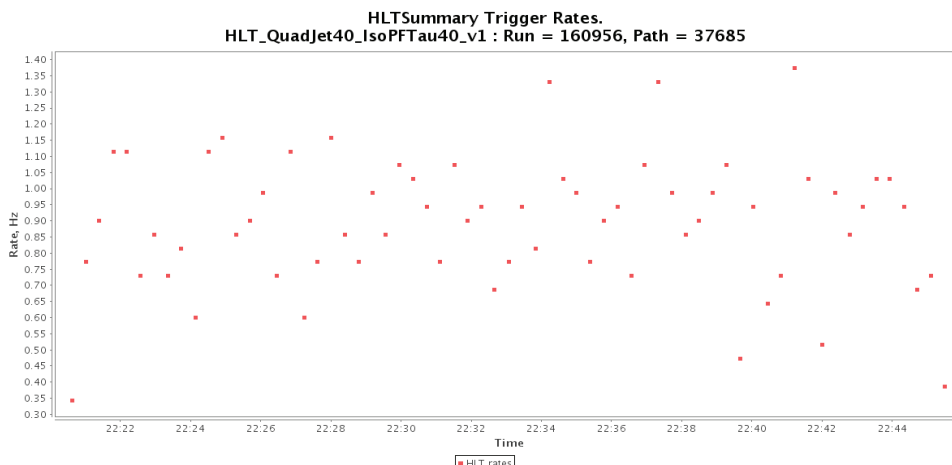

Figure 6.7: Rate for the HLT_QuadJet40_IsoPFTau40 trigger in run 160956 corresponding to a luminosity of $1.3 \cdot 10^{32}$ cm^{-2}s^{-1}. Mean recorded rate: 0.89 Hz.

the 2011 data taking with low pileup conditions) and a part of the 2011B run (corresponding to the second part of the 2011 data taking) are used for the analysis. The second part of the Run2011B was not analyzed since the HLT_QuadJet45_IsoPFTau45 trigger is prescaled from run 178421. The total integrated luminosity of the analysed dataset sums up to $\mathcal{L}$=3.9 fb^{-1}.

The simulated $t\bar{t}$ events as well as the W(Z)+jets events in CMS have been produced with MADGRAPH (73) as matrix element (ME) generator using the leading order parton distribution function set CTEQ6L1 (74), and matched with PYTHIA as parton shower (PS) which performs the parton showering, fragmentation, hadronisation and decays of short lived particles. Tau leptons are decayed using TAUOLA(24). For the $t\bar{t}$ production at parton level, all the leading order (LO) contributions to the process are simulated with the ME generator including up
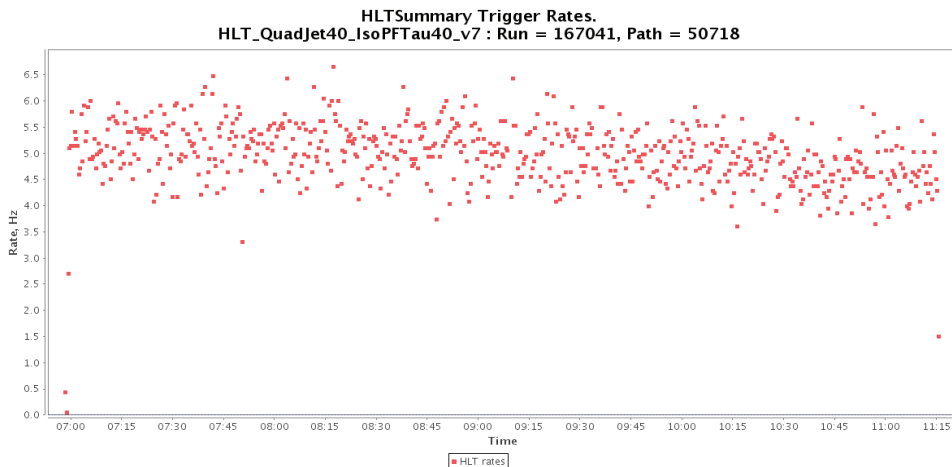
Figure 6.8: Rate for the HLT_QuadJet40_IsoPFTau40 trigger in run 167041 corresponding to a luminosity of $0.7 \cdot 10^{33}$ cm^{-2}s^{-1}. Mean recorded rate: 4.97 Hz.

Table 6.1: Chosen trigger data taking period and run range. The corresponding integrated luminosity is given in the last column of the table.

Dataset	run range	trigger	$\mathcal{L}$ (pb^{-1})
/Run2011A_May10RecReco-v1/AOD	160431-163869	HLT_QuadJet40_IsoPFTau40	194.9
/Run2011A_PromptReco-v4/AOD	165088-167913	HLT_QuadJet40_IsoPFTau40	38.8
/Run2011A_PromptReco-v4/AOD	165970-166782	HLT_QuadJet45_IsoPFTau45	282.9
/Run2011A_05Aug2011-v1/AOD	170826-170901	HLT_QuadJet40_IsoPFTau40	592.1
/Run2011A_05Aug2011-v1/AOD	171050-172619	HLT_QuadJet45_IsoPFTau45	364.2
/Run2011A_PromptReco-v6/AOD	172620-172692	HLT_QuadJet45_IsoPFTau45	655.0
/Run2011B_PromptReco-v1/AOD	175860-178420	HLT_QuadJet45_IsoPFTau45	1790

to 3 additional partons in the final state. Jets of the final state are clustered with a k_t algorithm, with a minimum p_T threshold of 20 GeV and a maximum pseudorapidity $|\eta| = 5$. A model called *Tune Z2* is used for the simulation of the underlying events. It is chosen since it gives the best data/MC agreement. Single top events are generated using POWHEG (75) interfaced to PYTHIA and TAUOLA. The top quark mass is set to 172.5 GeV and the Next-to-Next-to-Leading-Log (NNLL) $t\bar{t}$ cross section is assumed to be 164 ± 10 pb (76).

The Monte Carlo samples used for the analysis include a special simulation of the pileup events which should roughly cover the pileup conditions expected during the 2011 data-taking. The MC distribution is flat from 0 to 10 events with a poissonian tail above 10, while in data the pileup distribution has a typical poissonian distribution. Hence, in order to match both pileup distributions in data and in MC a reweighting procedure is used. A set of weights w_i is derived by normalizing the two histograms and dividing the two distributions of the number of interactions bin

by bin:

$$w_i = \frac{N_{Data}^i}{N_{MC}^i} \tag{6.1}$$

N_{Data}^i refers to the number of pileup events in the bin i in the data distribution, and N_{MC}^i to the number of pileup events in MC. These weights are then applied to the simulated events by weighting each event by the weight corresponding to the number of simulated interactions.

The Summer11 MC simulation of the CMS experiment is used to estimate the signal efficiency as well as the contribution from electroweak background processes. Table **6.2** summarizes the considered simulated datasets.

6.3 Event selection

The event reconstruction relies for all objects on the particle flow (PF) technique, described in section **4.2** of this document. The aim is to reconstruct objects stemming from $t\bar{t}$ pairs where one W boson decays hadronically and the other one to a hadronically decaying tau and a tau neutrino. The event selection is mainly based on the presence of four PF jets reconstructed with the anti-kT clustering algorithm with cone size R=0.5 and on the presence of one PF tau reconstructed with the combined medium isolation working point of the Hadron Plus Strip (HPS) algorithm described in section **4.3.2.2** of this document. Since two b-jets from the top decays are expected in the final state, one jet should be b-tagged using the medium working point of the Jet Probability algorithm described in section **5.5** of this document. Tau candidates are required to pass the tight discriminators against electrons and muons. A veto on the presence of well identified loosely isolated electrons and muons is applied to further suppress the electron-tau and muon-tau contamination. A loose cut on the transverse missing energy $E_T^{miss} > 20$ GeV is applied to get rid of the important multijet background and achieve a good separation for the input variables used in the neural network described in section **6.4**. The choice of the cut on the transverse missing energy is discussed in section **6.4.1**. The previously described criteria define what we call in the following the search sample. A more detailed description of the selected objects is given in the sub-sections **6.3.1.1 to 6.3.1.7**.

6.3.1 Object selection

6.3.1.1 Vertex selection

The primary vertex is reconstructed using the deterministic annealing algorithm (62). Selected events are required to contain at least one good quality primary vertex fulfilling the following requirements:

- the primary vertex should not be identified as *fake vertex*[1];

- the vertex fit is performed with at least 4 degrees of freedom, $ndof >= 4$;

- the primary vertex should lie inside the detector center, $z_{PV} < 24$ cm;

- the radial coordinate of the primary vertex with respect to the beam line is smaller than 2 cm.

---

[1]During the reconstruction of the PV, the collection is filled with all the vertices found and fitted. If the fit procedure doesn't find any reconstructed vertex, a vertex based on the beam spot is put into the collection. Hence the vertex made out of the beamspot is called a fake vertex, and is not made of any tracks. This kind of vertex has to be removed.

Process	σ (pb)	Dataset	Simulated events
$t\bar{t}$	164 ± 10	/TTJets_TuneZ2_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1	3.03 M
W+jets	31314 ± 1558	/WJetsToLNu_TuneZ2_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1	46.3 M
$Z/\gamma^* \to ll$ ($M_{ll} > 50GeV$)	3048 ± 132	/DYJetsToLL_TuneZ2_M-50_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1	35.8 M
single t (s-channel)	2.72 +0.11 -0.10	/T_TuneZ2_s-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	138 k
single $\bar{t}$ (s-channel)	1.49 +0.09 -0.08	/Tbar_TuneZ2_s-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	45 k
single t (tW-channel)	5.3 ± 0.6	/T_TuneZ2_tW-channel-DR_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	334k
single $\bar{t}$ (tW-channel)	5.3 ± 0.6	/Tbar_TuneZ2_tW-channel-DR_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	359 k
single t (t-channel)	42.6 +2.4 -2.3	/T_TuneZ2_t-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	1.9 M
single $\bar{t}$ (t-channel)	22.0 +0.10 -0.8	/Tbar_TuneZ2_t-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	1.9 M
$t\bar{t}$ M = 166.5 GeV	164 ± 10	/TTJets_TuneZ2_mass166_5_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v3/AODSIM	1.5 M
$t\bar{t}$ M = 178.5 GeV	164 ± 10	/TTJets_TuneZ2_mass178-5_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v3/AODSIM	1.5 M
$t\bar{t}$	164 ± 10	/TTJets_TuneZ2_scaledown_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	0.9 M
$t\bar{t}$	164 ± 10	/TTJets_TuneZ2_scaleup_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	1.0 M
$t\bar{t}$	164 ± 10	/TTjets_TuneZ2_matchingdown_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	0.5 M
$t\bar{t}$	164 ± 10	/TTjets_TuneZ2_matchingup_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	1.0 M
$Z/\gamma^* \to ll$	3048 ± 132	/ZJets_TuneZ2_scaledown_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	1.0 M
$Z/\gamma^* \to ll$	3048 ± 132	/ZJets_TuneZ2_scaleup_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	1.3 M
$Z/\gamma^* \to ll$	3048 ± 132	/ZJetsToLL_TuneZ2_matchingdown_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	1.5 M
$Z/\gamma^* \to ll$	3048 ± 132	/ZJetsToLL_TuneZ2_matchingup_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	1.6 M
W+jets	31314 ± 1558	/WJetsToLNu_TuneZ2_scaleup_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	9.2
W+jets	31314 ± 1558	/WJetsToLNu_TuneZ2_scaledown_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	10.0
W+jets	31314 ± 1558	/WJetsToLNu_TuneZ2_matchingup_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	9.8 M
W+jets	31314 ± 1558	/WJetsToLNu_TuneZ2_matchingdown_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	9.6 M

Table 6.2: List of signal and background samples used in the analysis, as well as samples used to determine the systematic uncertainties. The cross section values and related uncertainties are taken from (77).

6.3.1.2 Jet selection

The jets are reconstructed using the PF algorithm and anti-kT clustering algorithm with cone size R=0.5. Each event must have at least four reconstructed jets with pseudorapidity $|\eta| < 2.4$. In order to cope with the trigger requirements three jets are required to have $p_T >45$ GeV. The fourth jet is required to have $p_T >20$ GeV. The jet energies are corrected following the official CMS prescriptions (78). To take into account differences in the jet energy resolution of the order of 10 % between simulation and data, a last correction step is applied: *L2L3Residuals*. In addition jets are required to satisfy loose identification criteria to reduce the contamination from fake jets due to detector noise or from electrons: number of constituents >2, multiplicity of charged particles > 1, energy fractions (charged electromagnetic, charged hadronic, neutral hadronic, neutral electromagnetic) < 0.99. Jets overlapping with identified hadronic taus or electrons and muons (as defined below) are excluded ($\Delta R(jet, lepton) > 0.4$).

6.3.1.3 Tau identification

The hadronically decaying tau candidate is reconstructed with the HPS algorithm. The medium working point of the combined discriminator including $\Delta\beta$ correction is used in this analysis and requires that the sum of the transverse energies of the additional charged hadrons and photons should be less than 1 GeV. The reconstruction efficiency for this working point is estimated to be $\sim 45\%$ for real $Z \to \tau\tau$ events and 0.5% for jets faking taus (53) - (79) - (80).

Furthermore, tau candidates are required to pass the tight discriminators against muons and electrons (see section **4.3.2.2**).

To cope with the trigger design the tau candidate is required to fulfill: $p_T >45$ GeV. Additional requirements are:

- $|\eta(\tau_h)| < 2.3$ and $|\eta(\tau_h)| \notin [1.444, 1.566]$ in order to exclude the barrel-endcap transition region of the electromagnetic calorimeter;

- leading track $p_T > 10$ GeV;

- $|z_{\mathrm{vtx}}(\tau_h) - z_{\mathrm{PV}}| <1$ cm [1];

- |impact parameter of tau candidate with respect to the beam line| < 0.04 cm to ensure a prompt production from the collision vertex.

---

[1] $z_{vtx}(\tau_h)$ is the z component of the vertex associated to the tau candidate.

6.3.1.4   Muon veto

Muons are only considered to veto events containing well identified loosely isolated muons:

- $p_T > 10$ GeV and $|\eta(\mu)| < 2.4$;

- identification as a global muon;

- relative isolation RE< 0.15, where RE$=\frac{trkIso+ecalIso+hcalIso}{p_T(\mu)}$. Each component is computed from deposits in a cone of 0.3 around the muon direction, $p_T(\mu)$ represents the transverse momentum of the muon;

- $|z_{\text{vtx}}(\mu) - z_{\text{PV}}| < 1$ cm.

6.3.1.5   Electron veto

Electrons are only considered to veto events containing well identified loosely isolated electrons :

- $p_T > 15$ GeV and $|\eta(\text{electron})| < 2.5$;

- relative isolation RE< 0.15, where RE$=\frac{trkIso+ecalIso+hcalIso}{p_T(e)}$. Each component is computed from deposits in a cone of 0.3 around the electron direction, $p_T(e)$ represents the transverse momentum of the electron;

- photon conversion rejection (81): number of lost hits in tracker<2, minimal distance between the electron and its closest opposite sign track: $\Delta(\cos\theta) > 0.02$ and Δdistance(r-ϕ) > 0.02;

- $|z_{\text{vtx}}(\text{electron}) - z_{\text{PV}}| < 1$ cm.

6.3.1.6   b-jet identification

Events selected in the search sample are required to contain at least one b-tagged jet. The Jet Probability (JP) algorithm is chosen using its medium working point (JPM). In order to reproduce the b-tagging efficiencies measured in data, scale factors (SF) have to be applied on the simulated events (66) to reproduce the p_T dependence of the b-tagging algorithm and differences between simulation and data. Following the method used by the $t\bar{t}$ dilepton analysis with b-tagging (82), a probabilistic approach is used instead of selecting directly the simulated events with a cut on the b-tag discriminant. As described in (69), the probability that a jet is selected by the b-tagging discriminant is given by:

$$P_i = SF_i \cdot Eff_i^{MC} \tag{6.2}$$

where Eff_i^{MC} is the b-tagging efficiency measured in simulated $t\bar{t}$ events and SF_i the scale factor associated to the jet. SF_i depends on the p_T, η of the jet as well as

on its flavour (see section **5.7**). SF_b, SF_c, SF_l denote respectively the scale factor for b-jets, c-jets and light-jets. Eff_i^{MC} (i=b,c,l) has been measured on simulated $t\bar{t}$ ev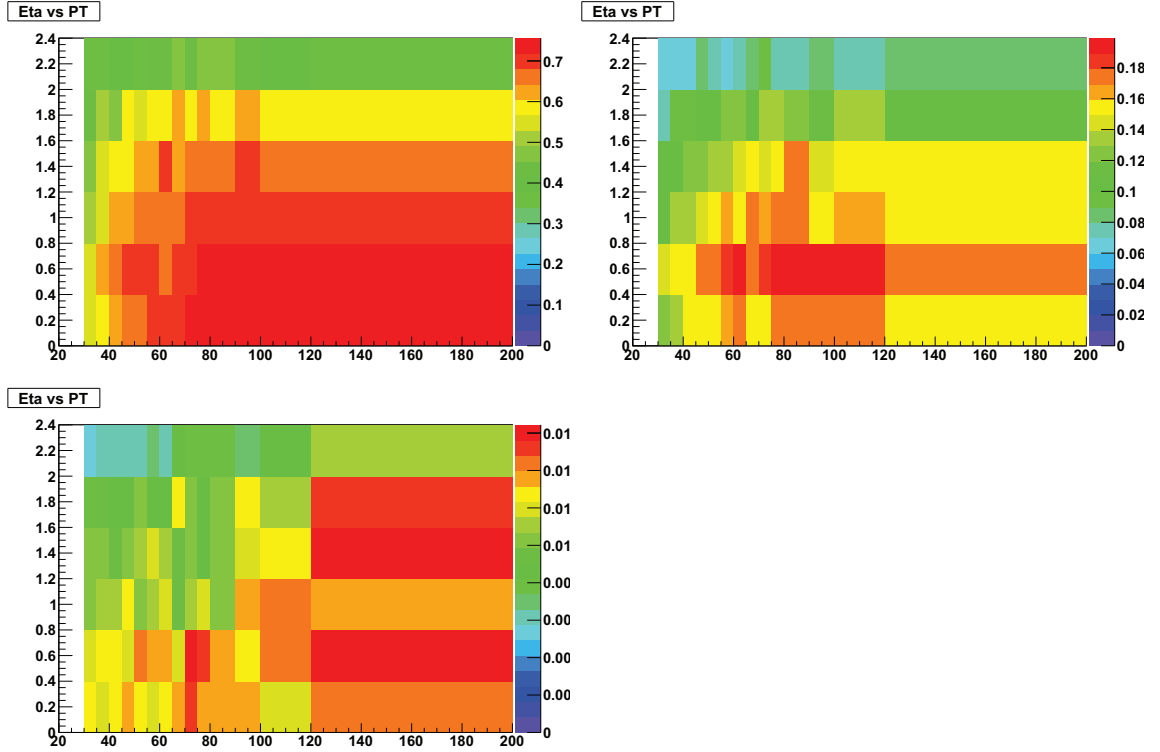ents as a function of p_T, η for the JPM working point. In figure **6.9** b-tagging efficiency for the JPM algorithm, measured on $t\bar{t}$ simulated events, as a function of η and p_T for b, c and light quarks is presented. The scale factors are taken from (66), SF_b and SF_c are assumed to be equal. The probability that the event contains

Figure 6.9: b-tagging efficiency for the JPM algorithm, measured on $t\bar{t}$ simulated events, as a function of η and p_T for b (top left), c (top right) and light (bottom left) quarks.

no b-tagged jet is defined by:

$$P(0\,tag) = \Pi_i(1 - P_i) \tag{6.3}$$

And hence the probability that the event contains one b-tagged jet and at least one b-tagged jet by:

$$P(1\,tag) = \sum_i (P_i \cdot \Pi_j(1 - P_j)) \text{ with } j \neq i \tag{6.4}$$

$$P(\geq 1\,tag) = 1 - P(0\,tag) \tag{6.5}$$

Simulated events have been weighted by $P(\geq 1\,tag)$, to take into account the JPM b-tagging requirement to select events with at least one b-tagged jet. The probabilistic approach offers the possibility to better describe the b-tagged jet multiplicity.

In section **6.9** of the appendix the validation of the probabilistic method is presented.

6.3.1.7 Transverse missing energy

The transverse missing energy is computed using the PF algorithm. The L1 jet energy scale corrections are propagated.

6.3.1.8 Pileup treatment

The pileup distribution is reproduced in simulation using the so called 3D reweighting procedure described elsewhere (83). The number of pileup events in simulation is reweighted in order to match the distribution observed in data, that is obtained by multiplying the measured luminosity per bunch-crossing times the average total inelastic cross section, $\sigma_{\text{min. bias}} = 73.5$ mb. Figure **6.10** shows the pileup distribution measured in the analysis sample for Run2011A and Run2011B.
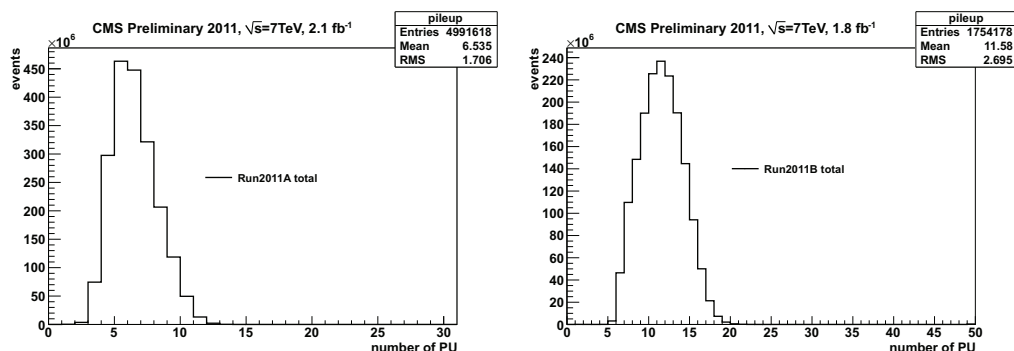
Figure 6.10: Pileup distribution observed in the analysed data samples of Run2011A ($\mathcal{L}$=2.1 fb^{-1}) and Run2011B ($\mathcal{L}$=1.8 fb^{-1}).

6.3.2 Expected event yields

Table **6.3** shows the expected number of events for the signal and the main expected backgrounds (except the multijet background) for an integrated luminosity of $\mathcal{L}$=3.9 fb^{-1} and the number of selected events in data. Given the low expected signal over background ratio obtained using a traditional cut flow, a multivariate technique is used to further discriminate signal and background, as described in section **6.4**.

6.4 Background estimation

Given the large uncertainty on the multijet cross section at the LHC and the huge number of events one would need to simulate to evaluate it, we use a data driven

search sample	→ τ_h+jets	→ll	→jets	→l+jets	→ $\tau_h\tau_h$	→ $e\tau_h$	→ $\mu\tau_h$
4 jets + τ_h	29056	577.2 ± 9.9	2.2 ± 0.4	156.3 ± 4.9	77.9 ± 3.8	24.1 ± 1.9	34.9 ±2.4	31.2 ± 2.2
e,μ veto	28951	576.7 ± 9.9	0.6 ± 0.2	156.3 ± 4.9	57.1 ± 3.3	24.1 ± 1.9	12.2 ±1.5	18.6 ± 1.8
≥ 1 btag	5047	459.0 ± 8.1	0.6 ± 0.2	118.8 ± 3.9	45.3 ± 2.7	19.7 ± 1.6	9.7 ± 1.2	14.5 ± 1.4
>20	3050	409.2 ± 7.5	0.6 ± 0.2	65.3 ± 2.9	41.1 ± 2.6	18.7 ± 1.5	9.1 ± 1.2	13.6 ± 1.3

	W+jets	Z+jets	single (tW)	single (tW)	single (s)	single (s)	single (t)	single (t)
4 jets + τ_h	547.9 ± 33.6	264.3 ± 13.6	26.7 ± 0.9	24.9 ± 0.9	0.6 ± 0.4	0.2 ± 0.1	11.5 ± 0.7	7.5 ± 0.8
e,μ veto	542.6 ± 33.5	237.6 ± 13.3	25.6 ± 0.9	23.9 ± 0.9	0.6 ± 0.4	0.2 ± 0.1	11.5 ± 0.7	7.5 ± 0.8
≥ 1 btag	67.3 ± 6.1	27.2 ± 1.8	17.8 ± 0.7	16.9 ± 0.7	0.4 ± 0.3	0.1 ± 0.1	8.9 ± 0.6	5.7 ± 0.6
>20	60.8 ± 5.9	20.8 ± 1.6	15.3 ± 0.6	14.3 ± 0.6	9.4 ± 0.7	0.1 ± 0.1	6.5 ± 0.5	3.9 ± 0.4

Table 6.3: Expected number of events for the signal and the main expected backgrounds (except the multijet background) for an integrated luminosity of $\mathcal{L} = 3.9 \ fb^{-1}$ and the number of selected events in data. The uncertainties are statistical only.

approach. Contributions from electroweak processes such as single top quark production, W(Z)+jets events and $t\bar{t}$ backgrounds are estimated by using the theoretical cross section along with the acceptance from simulation and the total integrated luminosity. Only the multijet background has been estimated using a data driven technique as will be explained in the following section.

6.4.1 Multijet background

The multijet background is inferred from data by applying the same selection cuts as for the search sample except a veto on the presence of a b-tagged jet (selected with JPM). Using simulated events, we expect the resulting sample to contain less than 0.6 % of $t\bar{t} \to \tau_h$+jets events, less than 0.3% $t\bar{t}$ background events and less than 2% W(Z)+jets events. Therefore it provides a good representation of the multijet background and is used to train the NN. Increasing the cut on E_T^{miss}, loosening the b-tagging or tau identification criteria didn't lead to a better multijet background description.

To take into account the b-tagging efficiency with respect to the momenta and the pseudorapidity of the selected jets, the multijet events in data are weighted by the mistag probability to select at least one b-jet in the event. The probability is computed in a similar way than the one described in section **6.3.1.6**, except that the mistag rate efficiency for light jets is assumed for P_i. The mistag rate efficiency has been measured in real data (see section **5.7.2**), using tracks with negative signed track impact parameter with respect to the primary vertex to describe light jets: $P(\geq 1\,mistag) = \sum_i (Pi_i \cdot \Pi_j (1 - P_j))$ with $j \neq i$. The effect of the reweighting is discussed in section **6.5.2**.

6.5 Neural network method

The neural network method is an artificial generalization of the more complex biological functioning of the neural nework of the human brain. Based on this biological model, many artifical neural network methods were developed. The simplest one, which is chosen for our analysis, is called Multi Layer Perceptron (MLP). The main goal of the method is to distinguish signal from background events in the case where a simple cut based analysis fails. The figure **6.11** represents a schematic view of the MLP method. We can summarize it in the following steps:

- the nodes of the first layer contains some input variables $(x_1...x_n)$, for example the kinematic variables of jets, taus, electrons or any other physics observables, of well known signal and background events;

- each node of the second layer (hidden layer) defines a sigmoid function [1] taking as input the NN input variables and the weights linking the input nodes to the nodes of the hidden layer. Hence if we call Z_j the content of the node j and $A(x)$ a generic sigmoid function we have: $Z_j = A(\sum_i w_{ij} x_i)$ (see figure **6.11**);

- the third layer delivers the output variable Y, which is a linear combination of the hidden nodes and their weights ($Y = \sum_j w_{jk} Z_j$), which can be used as discriminant to separate signal and background.

The goal of the algorithm is to compute the correct weights in order to obtain the best separation as possible between signal and background. Weights are computed in the so called *training* phase using an iterative technique consisting of the following steps:

- at first the algorithm starts using random weights;

- the output computed with random weights is compared with the desidered output;

- the weights are modified in order to match the computed output with the desidered one;

- the procedure to adapt the weights is iterated as many times as needed to find the best match with the desidered output.

At the end of the training phase, the MLP output distribution is tested in an independent sample of data.

From a mathematical point of view, the MLP method represents an example of regression problem from R^n →R (**?**). It relies on the following theorem:

all R^n →R functions can be approximated to any level of accurancy, with a linear combination of sigmoid functions.

Hence the weights are the coefficients of the linear combination of the sigmoid functions used to condense the information carried by the n input variables in only one scalar output.

6.5.1   Neural network design and training

We use the MLP neural network implemented inside the TMVA package (84). The following variables have been considered as input variables for the NN:

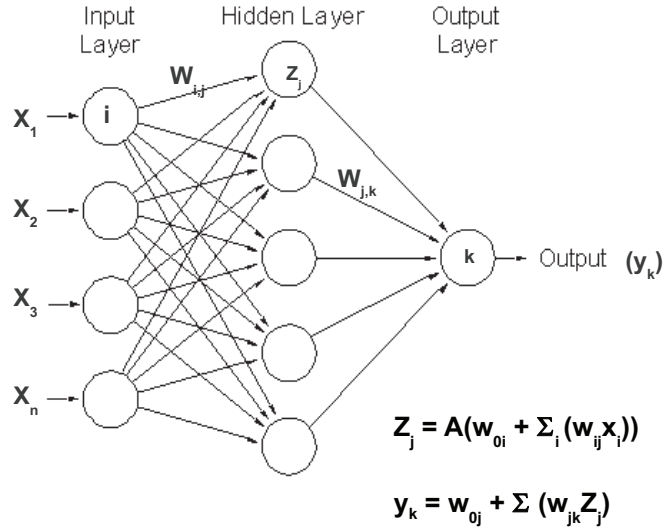[1]A sigmoid function is a function of the type: $\frac{1}{1-e^{-x}}$.

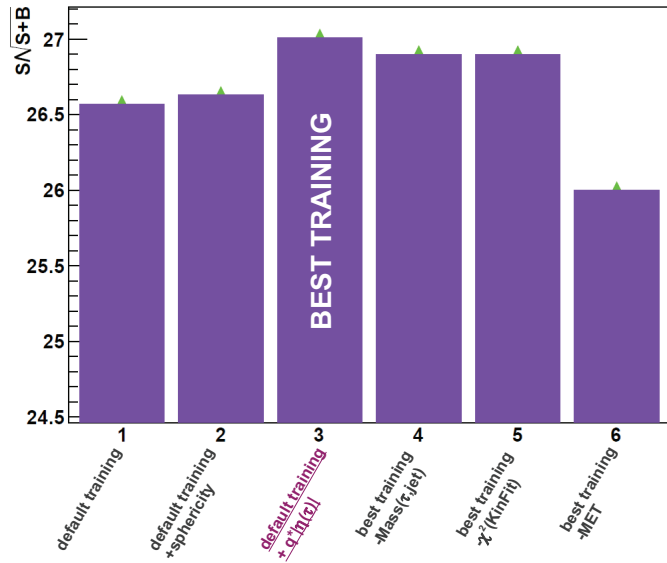**Figure 6.11:** Drawing the MLP neural network method.

Figure 6.12: Best $S/\sqrt{S + B}$ obtained for a given cut on the NN output distribution, using different sets of input variables.

- H_T: the scalar sum of the transverse momenta of all the selected jets and hadronic tau candidate;

- aplanarity: $A = \frac{3}{2}\lambda_1$, with λ_1 being the smallest eigenvalue of the momentum tensor; $M^{\alpha\beta} = \sum_i p_i^\alpha p_i^\beta / \sum_i |\vec{p}_i|^2$;

- sphericity: $A = \frac{3}{2}(\lambda_1 + \lambda_2)$, with λ_1, λ_2 being the smallest and next smallest eigenvalues of the momentum tensor; $M^{\alpha\beta} = \sum_i p_i^\alpha p_i^\beta / \sum_i |\vec{p}_i|^2$;

- $q(\tau_h) \cdot |\eta(\tau_h)|$: the hadronic tau charge multiplied by the absolute value of the pseudorapidity of the tau candidate;

- E_T^{miss}: the transverse missing energy;

- $\Delta\phi(\tau_h, E_T^{miss})$: the azimuthal angle between the hadronic tau candidate and the transverse missing energy direction;

- $M(jets, \tau_h)$: the invariant mass of the system of all the selected jets and the hadronic tau candidate;

- $M_T(\tau_h, E_T^{miss})$: the transverse mass of the hadronic tau candidate and transverse missing energy;

- χ^2: the χ^2 returned by the semileptonic top kinematic fitter constraining the hadronic W boson and top quark masses (85).

The choice of the NN input variables has been performed by optimising the best $S/\sqrt{(S+B)}$ ratio obtained for a given cut on the NN output distribution (see figure **6.12**). The study has been performed starting from a set of default variables: aplanarity, E_T^{miss}, $M(\tau_h, jets)$, H_T, $\chi^2(kin.fit)$, $\Delta\phi(\tau_h, E_T^{miss})$.

Figure **6.13** shows the discrimination power of the retained input variables: aplanarity, $q(\tau_h) \cdot |\eta(\tau_h)|$, E_T^{miss}, $M(\tau_h, jets)$, H_T, $\chi^2(kin.fit)$, $\Delta\phi(\tau_h, E_T^{miss})$. The training has been performed using 3000 simulated $t\bar{t} \to \tau_h$+jets passing the search sample selection cuts and 5000 multijet events selected in data according to the multijet selection criteria. Figure **6.14** shows the NN output distributions obtained after training and the obtained performance. We used a NN configuration including one hidden layer and 12 hidden nodes. Figure **6.15** shows the NN convergence test, no overtraining is observed.
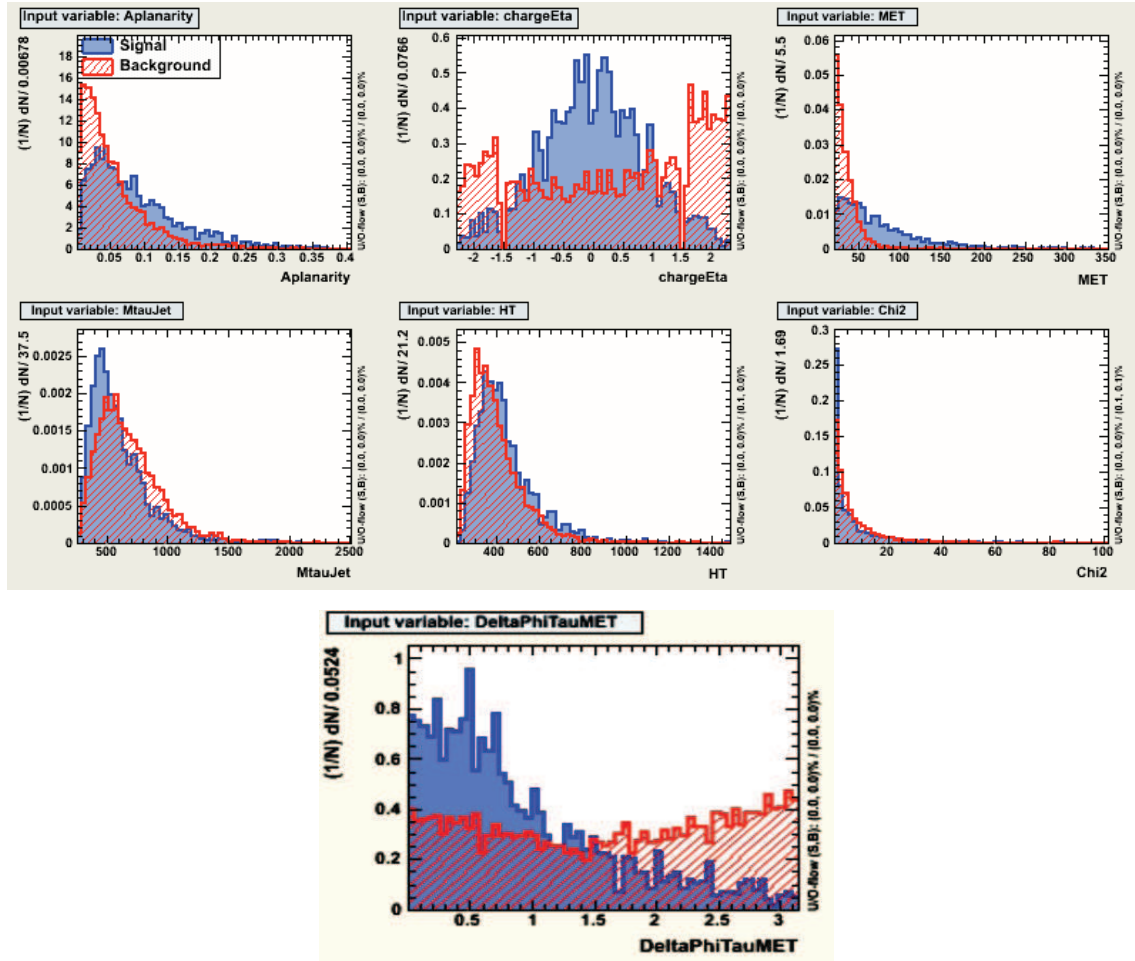
Figure 6.13: Input variables of the NN. Number of signal and multijet data events are normalized. From left to right and top to bottom: aplanarity, $q(\tau_h) \cdot |\eta(\tau_h)|$, E_T^{miss}, $M(\tau_h, jets)$, H_T, E_T^{miss}, $\chi^2(kin.fit)$, $\Delta\phi(\tau_h, E_T^{miss})$.
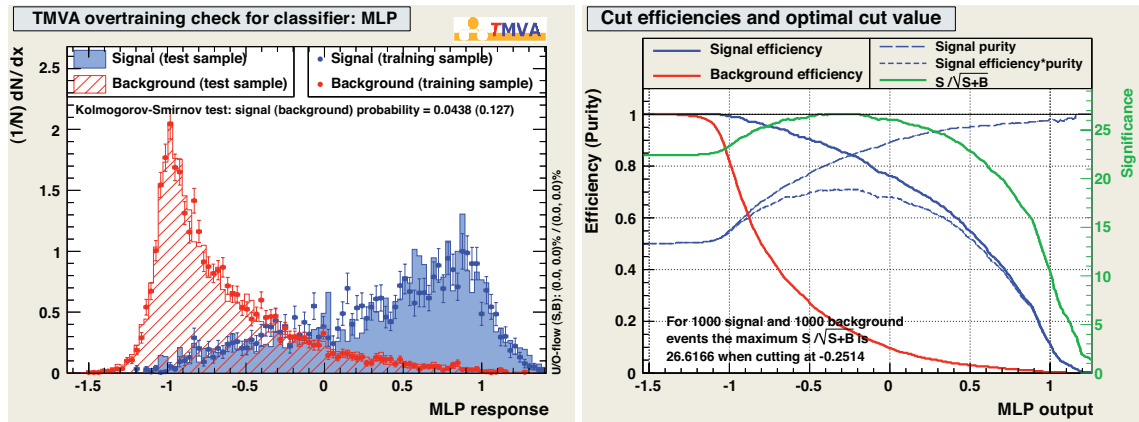

Figure 6.14: Ouput distributions of the neural network for signal and multijet data (left). NN performance (right).
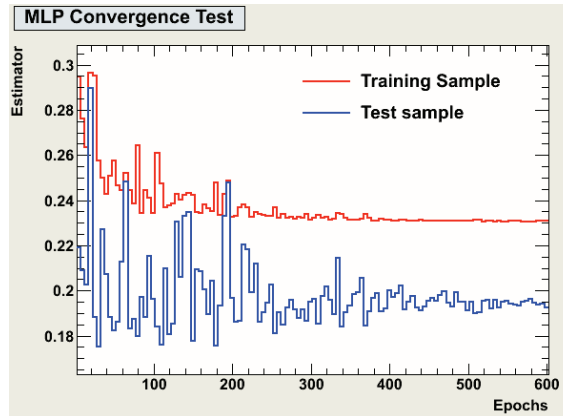
Figure 6.15: Convergence test of the NN training.

6.5.2 Fit to the data

To estimate the multijet background and minimise the statistical uncertainty of the measurement of the cross section, we fit the entire NN output distribution rather than counting events above a given value. The estimated number of signal and multijet background events is obtained from a negative log-likelihood fit to the NN output distribution (86). The number of signal events is measured to be 386.9 ± 28.8 and the number of multijet QCD events is 2393.6 ± 28.8. Table **6.4** summarizes the contribution of the various processes after a fit to the data. The expected contributions from the electroweak processes and top backgrounds are taken from table **6.3**.

Table 6.4: Estimated number of signal and multijet events after a fit to the NN output distribution and expected contributions of the electroweak processes and $t\bar{t}$ backgrounds.

Search sample	$t\bar{t} \to \tau_h + jets$	Multijet	W+jets	Z+jets	$t\bar{t}$ background	single top
3050	386.9 ± 28.8	2393.6 ± 28.8	60.8 ± 7.8	20.8 ± 4.6	147.9 ± 4.5	40.0 ± 1.3

Figure **6.16** shows the fitted NN output distribution in linear and logarithmic scale. Figures **6.17-6.20** show the fitted input variables of the neural network. Figures **6.21, 6.22** show the jet and tau momentum spectra. Figures **6.23, 6.24** show: the jet multiplicity and transverse mass of the hadronic tau candidate and transverse missing energy, $M_T(\tau_h, E_T^{miss})$ as well as the M3 variable after a cut on NN output>0.5. The M3 variable is defined as the invariant mass of the 3 jets system with highest transverse momentum. Figure **6.24** shows the agreement obtained on the number of primary vertices.
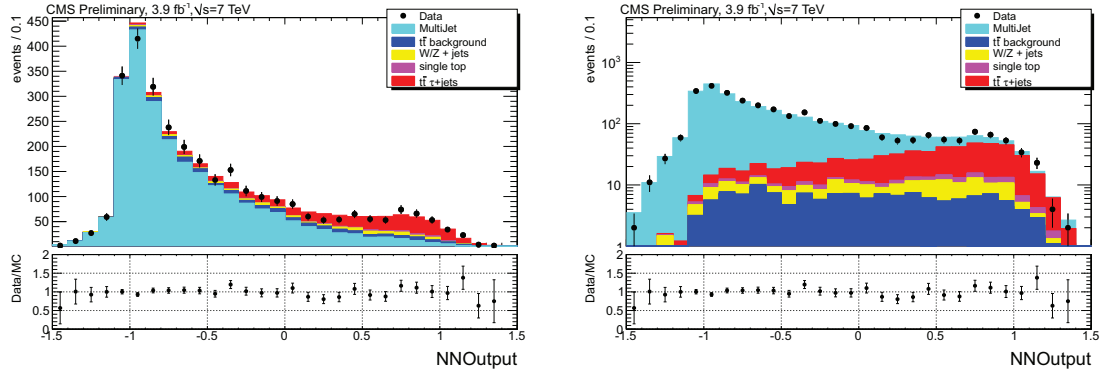
Figure 6.16: Distribution of the neural network output variable in linear (left) and logarithmic scale (right) after a fit of the signal and multijet QCD processes to the data.
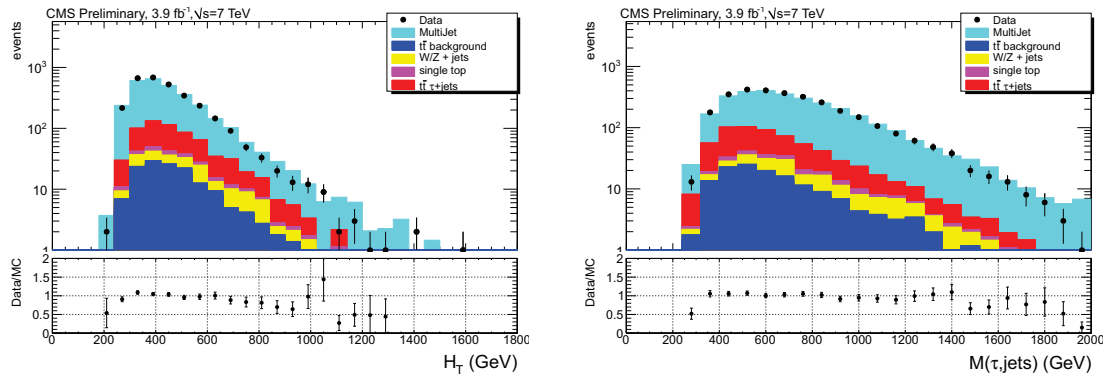

Figure 6.17: Distribution of the NN input variables H_T (left) and invariant mass of the hadronic tau candidate and selected jets $M(\tau_h, jets)$ (right) after a fit of the signal and multijet QCD processes to the data.
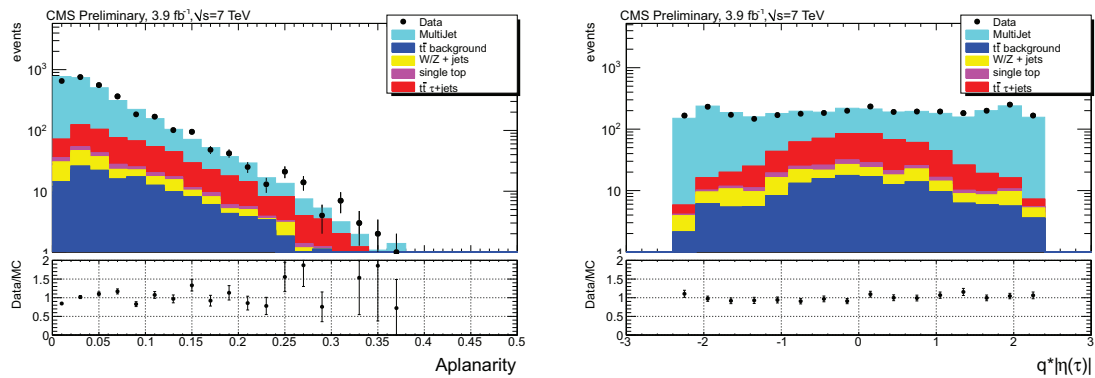

Figure 6.18: Distribution of the NN input variables aplanarity (left) and $q \cdot |\eta(\tau_h)|$ (right) after a fit of the signal and multijet QCD processes to the data.
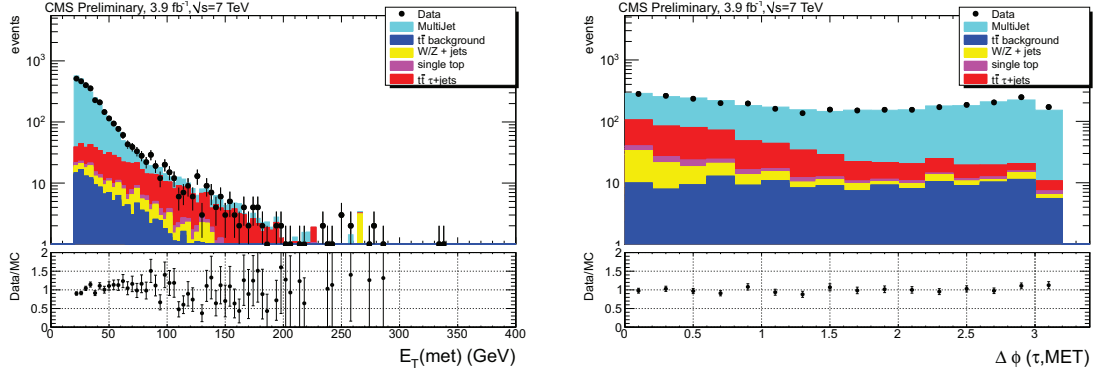
Figure 6.19: Distribution of the NN input variables E_T^{miss} (left) and $\Delta\phi(\tau_h, E_T^{miss})$ (right) after a fit of the signal and multijet QCD processes to the data.
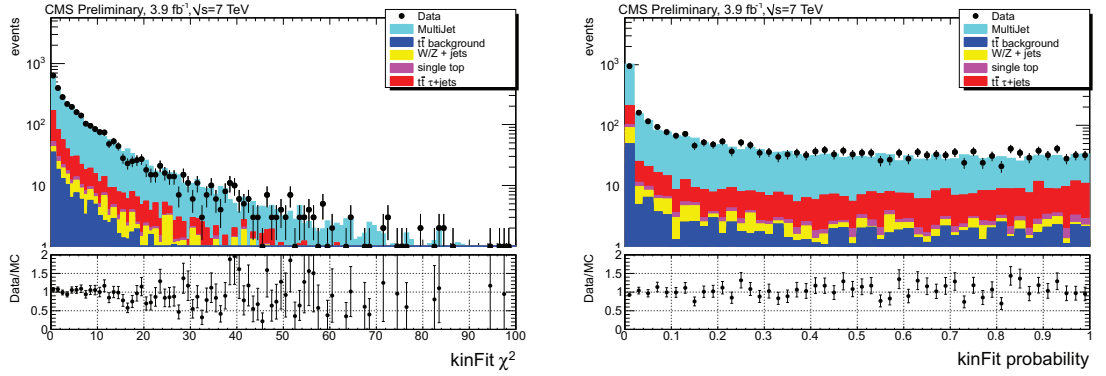

Figure 6.20: Distribution of the NN input variable $\chi^2(kin.fit)$ (left) and kin. fit probability (right) after a fit of the signal and multijet QCD processes to the data.
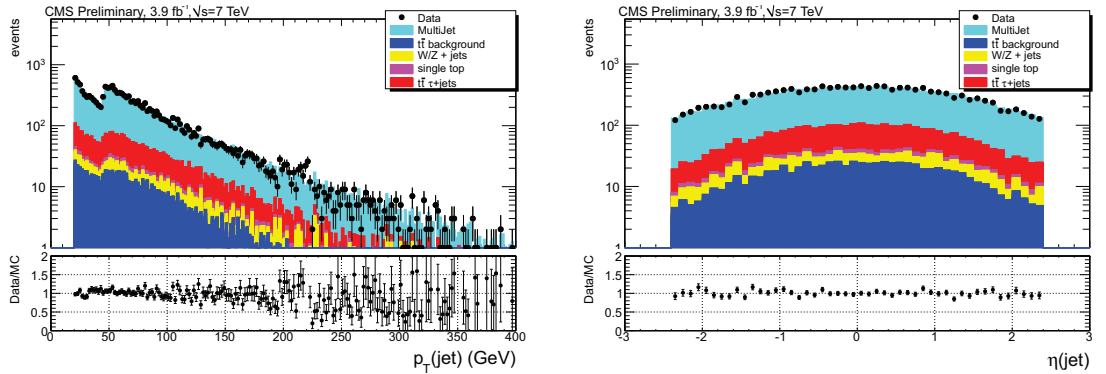

Figure 6.21: Distribution of the momentum (left) and the pseudorapidity (right) of the selected jets after a fit of the signal and multijet QCD processes to the data.
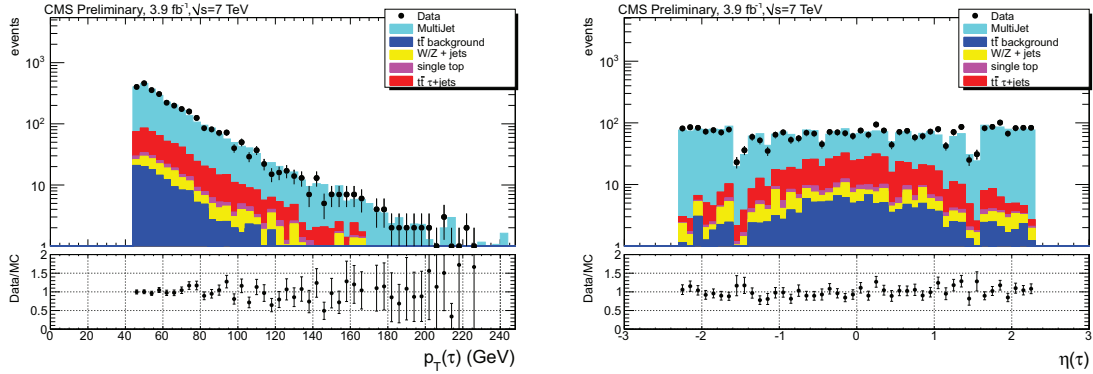
Figure 6.22: Distribution of the momentum (left) and the pseudorapidity (right) of the hadronic tau candidate after a fit of the signal and multijet QCD processes to the data.
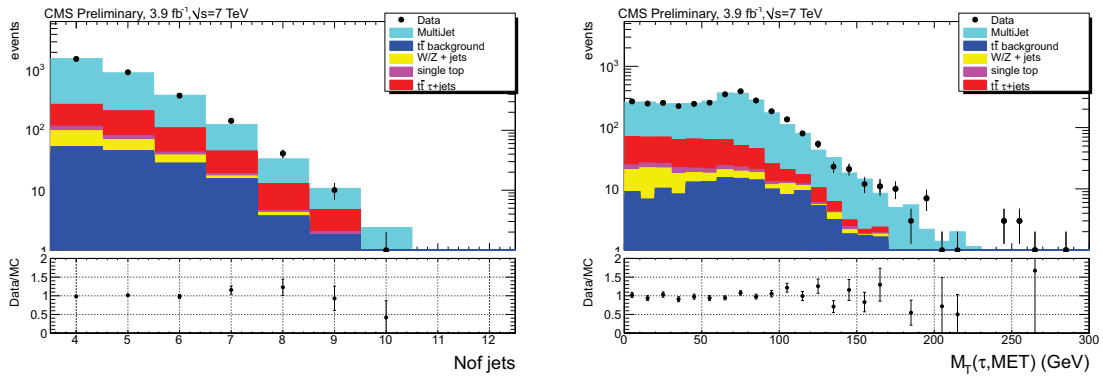

Figure 6.23: Distribution of the selected jet multiplicity (left) and the transverse mass of the hadronic tau candidate and transverse missing energy, $M_T(\tau_h, E_T^{miss})$, (right) after a fit of the signal and multijet QCD processes to the data.
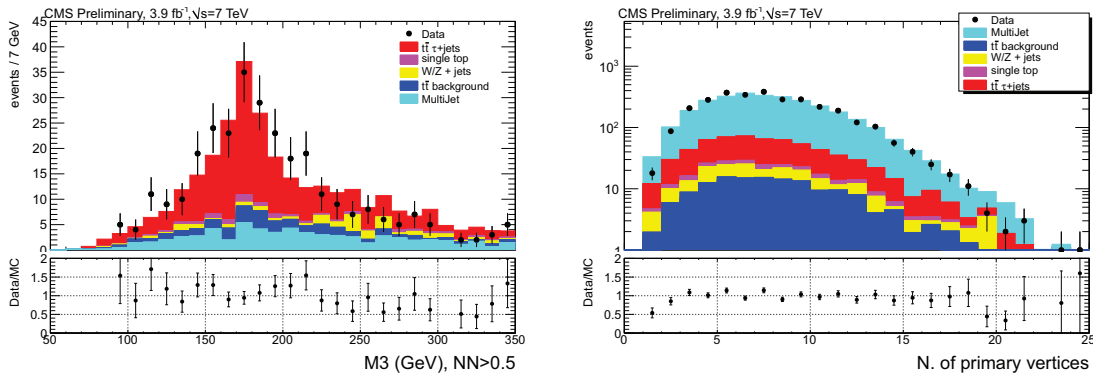

Figure 6.24: Distribution of the reconstructed M3 variable (left) after a cut on NN output¿0.5, and of the primary vertices multiplicity (right) after a fit of the signal and multijet QCD processes to the data.

The fit stability has been checked using 50000 pseudo-experiments. Events have been randomly sorted to form signal and multijet background subsamples with a number of events following a Poisson distribution with a mean value of respectively 396 and 2399 events. The fact that the reference samples consist of weighted events has been taken into account by using an *inverse transform sampling* method. Events are sorted according to a uniform distribution between 0 and the sum of the weights of the reference sample. To each event i, the sum of the previous weights is associated, $\Sigma_{j=0}^{i} w_j$. The event with closest $\Sigma_{j=0}^{i} w_j$ with respect to the sorted number is selected to enter the subsample.

The pull and the obtained fit uncertainty distributions are shown in figure **6.25**. As expected the pull distribution is gaussian with a mean value close to zero and a root mean square close to one. Furthermore the uncertainty distribution clearly shows that the observed uncertainty falls inside the core of the expected distribution of the uncertainty generated from the pseudo-experiments.
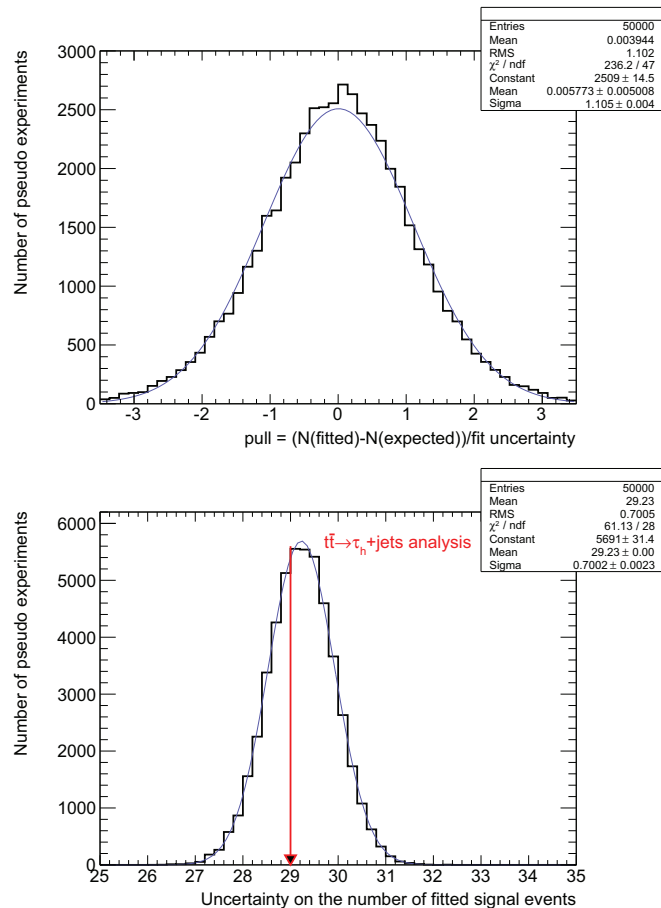

Figure 6.25: Pull distribution (top) and fit uncertainty (bottom) obtained using 50000 pseudo-experiments.

Effect of "b-mistag reweighting":

The effect of the "b-mistag reweighting" has been evaluated by iterating the NN
output fit using the non-reweighted multijet sample. The fitted number of signal
events increases by 10% as can be derived from table **6.5**.

Table 6.5: Estimated number of signal and multijet QCD events after a fit to the
NN output distribution and expected contributions of the electroweak processes and
$t\bar{t}$ backgrounds. In the first line no b-tag reweighting is applied, while the b-tag
reweighting is included in the second one.

Search sample	$t\bar{t} \to \tau_h + jets$	Multijet	W+jets	Z+jets	$t\bar{t}$ background	single top
3050	421.3 ± 29.0	2359.2 ± 29.0	60.8 ± 7.8	20.8 ± 4.6	147.9 ± 4.5	40.0 ± 1.3
3050	386.9 ± 28.8	2393.6 ± 28.8	60.8 ± 7.8	20.8 ± 4.6	147.9 ± 4.5	40.0 ± 1.3

Figures **6.30-6.37** in appendix show the corresponding NN output distributions
and input variables. A better agreement between data and expected signal and
backgrounds after reweighting is observed, especially for the H_T variable (figure
6.31) and the jet multiplicity (figure **6.37**). The χ^2/NDF improves respectively
from 2.15 to 1.49 for the H_T variable and from 6.50 to 0.70 for the jet multiplicity.
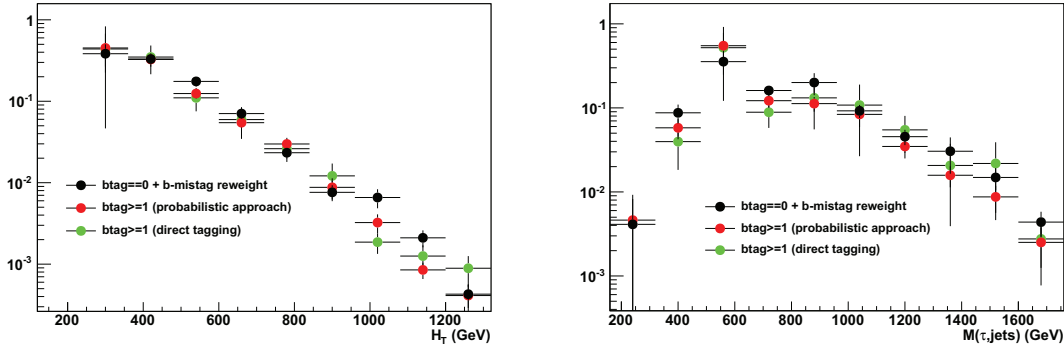

Figure 6.26: Distribution of H_T (right) and $M(\tau, jets)$ variables obtained on simulated
QCD events selecting one b-jet in the event (b-tag $>= 1$: probabilistic approach/
direct b-tagging) and applying the same selection criteria than for multijet data sample
(btag $== 0$)

In order to validate the b-mistag reweighting procedure, we tested it on simulated
QCD events. We compared the shapes of the QCD events in three different cases:

- tagging directly the b-jets using the JPM discriminator, as done on data for
  the search sample;

- weighting the events by the probability to identify at least one jet as a b-jet
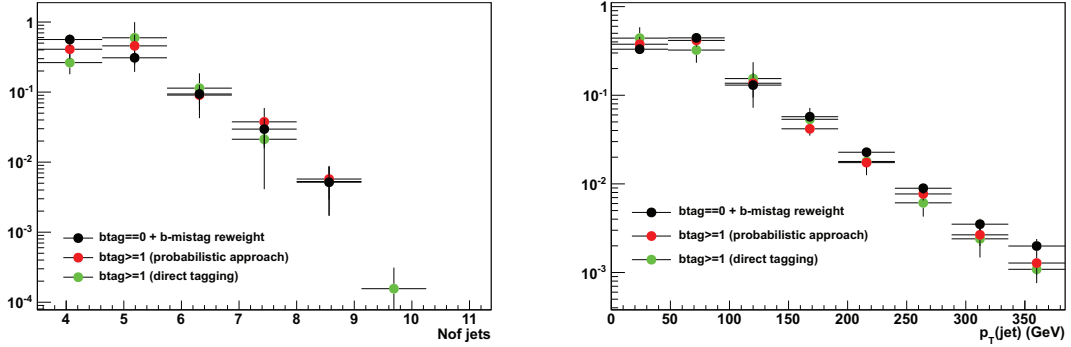 in the event (*probabilistic approach*);

Figure 6.27: The same than figure **6.26** for jet multiplicity (left) and jet p_T variables.

- applying a veto on the presence of a b-tagged jet and weighting the events by the probability to mistag at least one light jet as b-jet in the event (the same procedure is used to model the multijet background);

In figures **6.26-6.27** the obtained results on QCD simulated events are presented. The three methods are compatible within the statistical uncertainties available for the simulated events.

6.6 Cross section measurement

6.6.1 Systematic uncertainties

The main sources of systematic uncertainties are due to the jet energy scale (JES), the jet energy resolution (JER), the tau energy scale (TauES), the tau identification (TauID), the b-tagging/b-mistagging efficiencies and the pileup description. The following systematic uncertainties on the signal selection efficiency and on the background estimates are considered:

- The uncertainty on the theoretical cross section relative to the different simulated processes is taken from (77).

- The uncertainty stemming from the top quark mass is evaluated considering two simulated samples where the nominal top mass of 172.5 GeV has been shifted by ± 6 GeV. The estimated uncertainties are +9.1 % and -17.8%, scaling the uncertainty to the knowledge of the actual top mass precision of 1.1 GeV (i.e. dividing the maximum half-difference by a factor of 5), the considered uncertainty on the obtained cross section is 2.8%.

- The dependency of the selection on the hard scattering scale Q, $Q^2 = m_{top}^2 + p_T^2$, has been estimated using dedicated samples for the $t\bar{t}$ processes. The scale has been varied by a factor of 0.5 and of 2.0. The measured relative uncertainty for $t\bar{t}$ processes is estimated to be 2.2%.

6. MEASUREMENT OF THE TOP-ANTITOP PRODUCTION CROSS SECTION IN THE HADRONIC TAU+JETS FINAL STATE

- The influence of the matching thresholds used to associate the matrix elements to the parton showers has been varied from 20 GeV to respectively 10 GeV and 40 GeV. The measured relative uncertainty for $t\bar{t}$ processes is estimated to be 2.9%.

- A systematic uncertainty due to the uncertainties on the proton PDF was also estimated. The uncertainty on the CTEQ66 PDF has been taken into account using an event by event reweighting technique (87). Events have been reweighted using the 2 parameter sets leading to the maximal up and down PDF uncertainties (out of the 2·22 associated uncertainties to CTEQ66). The fit on the NN output distribution has been reiterated using the reweighted events.

- ±5% uncertainty is accounted for the tau-leg trigger efficiency measurement.

- The pileup uncertainty is estimated by varying the number of pileup interactions measured in data according to the theoretical uncertainty of the minimum bias inelastic cross section (± 8%).

- The effect of the tau energy scale is estimated by moving the tau energy by ±3% (53) and recomputing the number of expected events for each simulated process.

- The uncertainty due to the tau identification efficiency is estimated to 6% (53).

- The uncertainty due to JES and JER are estimated according to the prescription described in (50). In practice, for both JES and JER, the jet energies are rescaled up or down taking into account the JES or JER uncertainty corresponding to one standard deviation (± 1σ). The corrections are propagated to the E_T^{miss} measurement;

- The uncertainty on the unclustered E_T^{miss} is estimated by scaling up and down the unclustered E_T^{miss} by ±10%;

- The uncertainty due to applying b-tagging data/MC scale factor for b, c, light jets to the simulated events is estimated by shifting the value of the applied SF by the uncertainty corresponding to one standard deviation (± 1σ) (70).

- The uncertainty on the b-mistagging reweighting method applied to the multijet data sample is estimated to 5%.

Table **6.6** summarizes the systematic uncertainties entering the cross section measurement. The systematic uncertainties have been derived by iterating the fit on the NN output using the simulated samples with corresponding modified weights. This procedure is chosen in order to take into account possible shape variations of

Table 6.6: Sources of uncertainties on the $t\bar{t} \to \tau_h$+jets acceptance, A_{TOT} (central column), and on the final cross section measurement (right column).

Relative uncertainty [%]	A_{TOT}	cross section
Jet energy scale (JES)	±9	± 11
Jet energy resolution (JER)	±0.3	± 2
Unclustered E_T^{miss}	±4	± 7
Tau energy correction	±6	± 7
Tau identification	±6	± 9
Tau trigger leg	±5	± 7
B-tagging	±2	± 3
Multijet b-mistagging reweighting		± 5
Pileup	±1	±5
Top quark mass	± 3	±3
Q^2 scale	± 2	±2
Parton matching	± 3	±3
PDF	± 4	± 5
theoretical cross section		± 3
Systematic	± 15	± 21
Statistical from fit and MC	± 2	± 8
Statistical from trigger	± 1	± 1
Total statistical	± 2	± 8

the NN input variables. The fit procedure was not iterated for the following systematic uncertainties: tau identification, efficiency of the tau trigger leg and theoretical parameters used in the MC generation. In fact, the systematic uncertainties related to the tau identification and efficiency of the tau trigger leg only affect the number of expected events and not the NN output shape. Regarding the theoretical uncertainties, we could not iterate the fit procedure to estimate the corresponding systematic uncertainty due to the limited number of simulated events. Table **6.6** summarizes the uncertainties on the signal acceptance as well as on the final cross section measurement. The estimation of the statistical uncertainty due to the trigger efficiency is computed using the error bars of the turn-on efficiency histogram (see figure **6.2**). The uncertainties given in table **6.6** are relative ones and are expressed in terms of percentages.

6.6.2 Result

The measurement of the $t\bar{t}$ cross section in the hadronic tau+jets channel is performed using the following expression:

$$\sigma_{t\bar{t}} = \frac{N - B}{A_{TOT} \cdot \mathcal{L} \cdot BR} \tag{6.6}$$

where N is the number of observed candidate events, B is the estimation of the background, $\mathcal{L}$ is the integrated luminosity, A_{TOT} is the total acceptance which includes the trigger efficiency and the efficiency of the offline event selection, and BR is the branching ratio of the semileptonic hadronic tau $t\bar{t}$ decay ($\sim 9.8\%$). The result we get for A_{TOT} is:

$$A_{TOT} = 0.0065 \pm 0.0001(stat.) \pm 0.0010(syst.) \tag{6.7}$$

The systematic uncertainty on the total acceptance is detailed in table **6.6**. The uncertainties are propagated from Eq.(6.6) to the cross section measurement in the following way:

$$\frac{(\Delta\sigma_{t\bar{t}})^2}{(\sigma_{t\bar{t}})^2} = \frac{(\Delta(N-B))^2}{(N-B)^2} + \frac{(\Delta A_{TOT})^2}{(A_{TOT})^2} + \frac{(\Delta\mathcal{L})}{\mathcal{L}} \tag{6.8}$$

The $t\bar{t}$ cross section measured for the hadronic tau+jets channel is:

$$\sigma_{t\bar{t}} = 156 \pm 12(stat.) \pm 33(syst.) \pm 3(lumi)pb \tag{6.9}$$

which is in agreement with the theoretical NNLL value of 164 ± 10 pb and with the CMS measurements performed in other $t\bar{t}$ semileptonic final states (88). In particular the values of the $t\bar{t}$ production cross section obtained for the electron and muon + jets channels are:

$$\sigma_{t\bar{t}} = 163.0 \pm 4.4(stat.) \pm 12.7(syst.) \pm 7.3(lumi)pb \tag{6.10}$$

for the electron + jets channel,

$$\sigma_{t\bar{t}} = 163.2 \pm 3.4(stat.) \pm 12.7(syst.) \pm 7.3(lumi)pb \tag{6.11}$$

for the muon + jets channel.

In figure **6.28** the top quark pair cross section measurements in the different final states are compared. We can see the perfect agreement between all measurements. In the combined measurement (89) the hadronic tau + jets final state is not yet included, but the precision of the combination is already of the order of 8% which is compatible with the precision of the approximate NNLO theory (90), (91).

6.7   Perspectives

The measurement of the $t\bar{t}$ production cross section in hadronic tau + jets final state was also done by the ATLAS collaboration using a smaller data sample corresponding to an integrated luminosity of 1.7 fb^{-1}. The result found by the ATLAS collaboration is: $\sigma_{t\bar{t}} = 200 \pm 19$ (stat.) ± 43(syst.)pb (92), in perfect agreement with our measurement. An interesting point of the ATLAS analysis is the trigger strategy
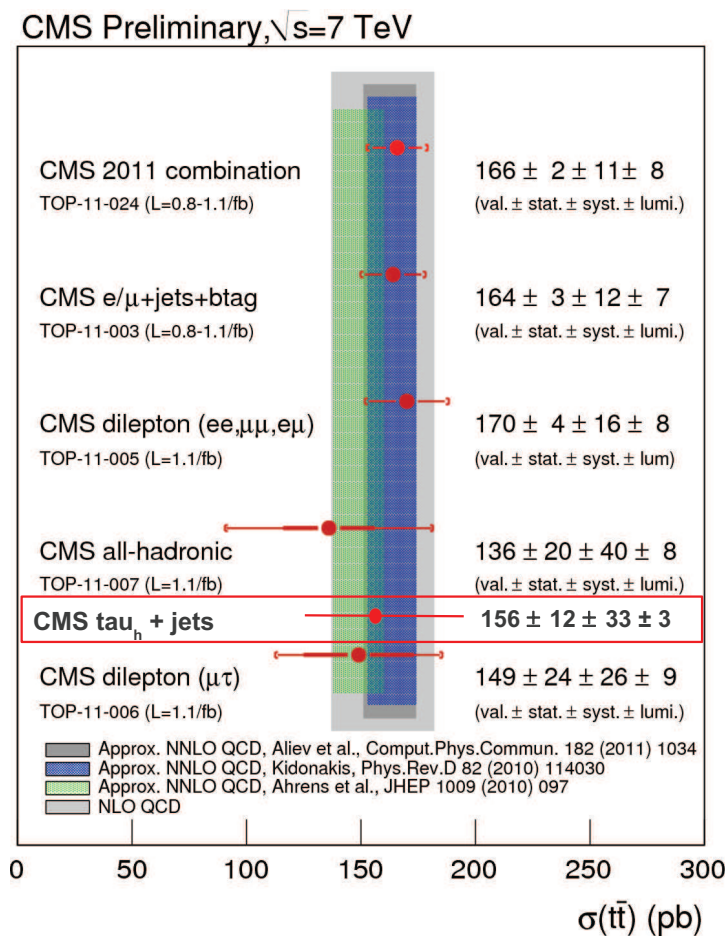
Figure 6.28: The cross section of the $t\bar{t}$ production obtained by a combination of measurements in different channels at $\sqrt{s} = 7$ TeV. The hadronic tau + jets final state is not yet included in the combination. The data are compared to the approximate NNLO calculations.

which relies on a multijet selection including two b-tagged jets without using the tau identification at online level. For the 2012 data taking in CMS, the present threshold on the tau p_T, at 45 GeV, would not be suitable, leading to a too high trigger rate. Raising the threshold would lead to a too low signal efficiency. The ATLAS trigger strategy instead could be a good alternative for two different reasons. The first one is that the presence of b-tagged jets would allow to remove the tau identification from the trigger path and hence increase the signal efficiency. The second reason is that the online jet selection could benefit online from the PF reconstruction, which was not available in 2011.

Besides that, an interesting variable exploited by the ATLAS collaboration is the number of tracks associated to a hadronic tau candidate, which in principle should be strictly correlated with the number of charged hadrons associated to the hadronic tau candidate. This variable could be considered in the analysis even if dedicated

studies on the effect of photon conversions (coming from the π^0 in the τ decay) and of nuclear interactions, would have to be done. These effects could distort the distribution especially in an high pileup environment.

On the other hand it could be interesting to re-do the analysis by excluding, from the NN training, the χ^2 variable returned by the kinematic fit. Indeed this χ^2 includes a constraint on the W boson mass. Hence, removing this variable we could keep possible charged higgs events like: $t\bar{t} \rightarrow Wb$, $H^+b \rightarrow q\bar{q}b$, $\tau_h\nu_\tau b$, and the analysis could be used to constrain the BR(t$\rightarrow H^+$b). The present upper limit imposed on the BR(t$\rightarrow H^+$b) by the CMS collaboration is of the order of 3-4% for a mass of the H^+ between 80 and 160 GeV, assuming BR($H^+ \rightarrow \tau_h\nu_\tau$)=1, as shown in figure **6.29**, (93).
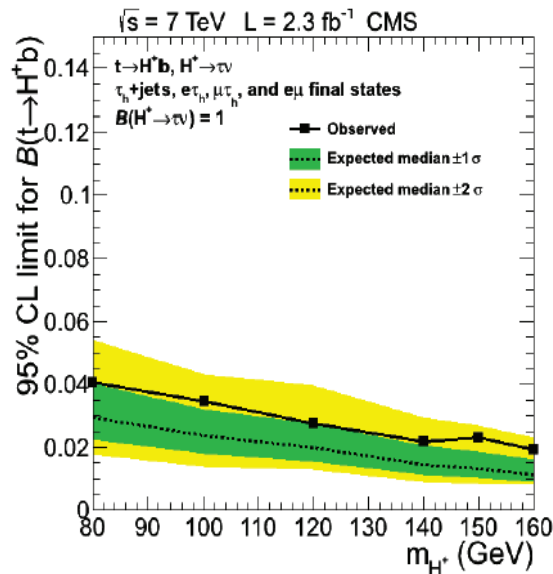

Figure 6.29: The upper limit on BR(t$\rightarrow H^+$b) as a function of m(H^+) obtained from the combination of all the considered final states.

Conclusion

The 2011 year was rich in interesting physics results for the LHC and for the CMS experiment in particular. Thanks to the good performace of the accelerator, CMS was able to collect about 5.0 fb^{-1} of data. The instantaneous luminosity provided by the LHC increased very fast, from $1 \cdot 10^{32}$ to more than $5 \cdot 10^{33}$ cm^{-2}s^{-1}, but the pileup also increased from 6 events on average in early 2011 to 10 events at the end of the 2011 data taking. The increase of the pileup conditions required a special effort from the entire collaboration both at trigger and at reconstruction level in order to rapidly adapt to the changes of the machine conditions. CMS was able to maintain many of its physics goals, regarding the Higgs and SUSY searches as well as the SM measurements.

In this thesis we presented the first measurement in the CMS experiment of the top-antitop production cross section in the tau + jets final state. This channel represents the most difficult one for the top-antitop production since both the reconstruction of the hadronic tau and the rejection of the multijet background are challenging.

To perform this measurement we started with the design of a specific trigger requiring four jets where one of these has to be identified as an hadronic tau. The performance of this trigger has been measured in this thesis. A dataset of 3.9 fb^{-1} was collected with this trigger and analyzed. At offline level we needed to apply a more sophisticated tau identification technique to identify the tau jets, based on the reconstruction of the intermediate resonances of the hadronic tau decay modes. Another crucial point was the b-jet identification, both to identify the b-jets in the final state and to modelize the background using a data driven technique. The studies done on the b-tag algorithms along the PhD period are also presented in this document with particular attention to the Jet Probability algorithm. It is the algorithm for which I performed the calibration since 2009 as well as the one used to tag the b-jets in the $t\bar{t}$ cross section measurement in the tau + jets final state. To perform the cross section measurement, we applied a Neural Network technique using as input kinematic and topological variables related to the presence of an hadronic tau candidate and jets in the final state. A fit to the Neural Network output distribution was done in order to estimate the signal contribution. A detailed estimation of the systematic uncertainties on the cross section $pp \to t\bar{t} \to \tau_h$+jets is also presented. The main contribution to these systematics are due to the tau

trigger efficiency and identification and the jet energy scale (propagated also to the E_T^{miss} measurement).

The measured cross section of the top-antitop production in hadronic tau + jets final state is:

$$\sigma_{t\bar{t}} = 156 \pm 12 \,(\text{stat.}) \pm 33, (\text{syst.}) \pm 3 \,(\text{lumi}) \,\text{pb} \qquad (6.12)$$

which is in very good agreement with the SM expectation of 164 ± 10 pb.

Appendix

6.8   Effect of the b-mistag reweighting

In this section of the appendix the final plots, for the same variables showed in figures from **6.16** to **6.24**, are presented before the application of the b-mistag reweighting procedure of the multijet background.
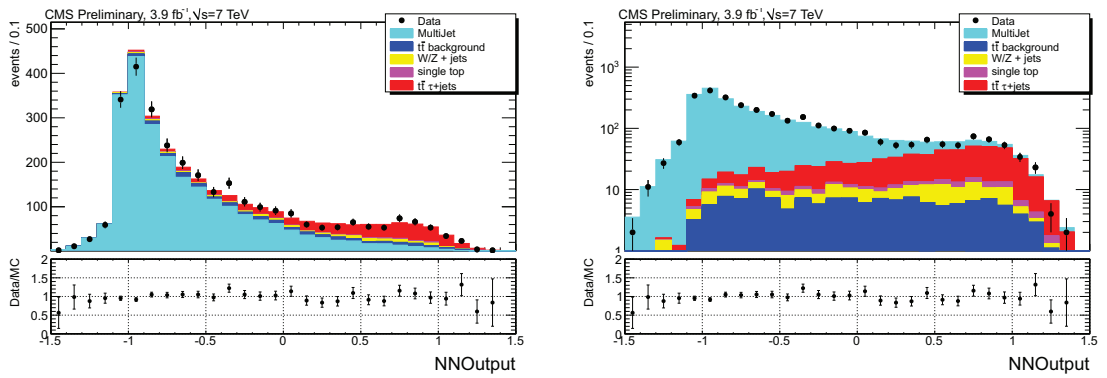
Figure 6.30: Distribution of the neural network in linear (left) and logarithmic scale (right) after a fit of the signal and multijet QCD processes to the data. No "b-mistag reweighting" is applied on the QCD multijet data sample.
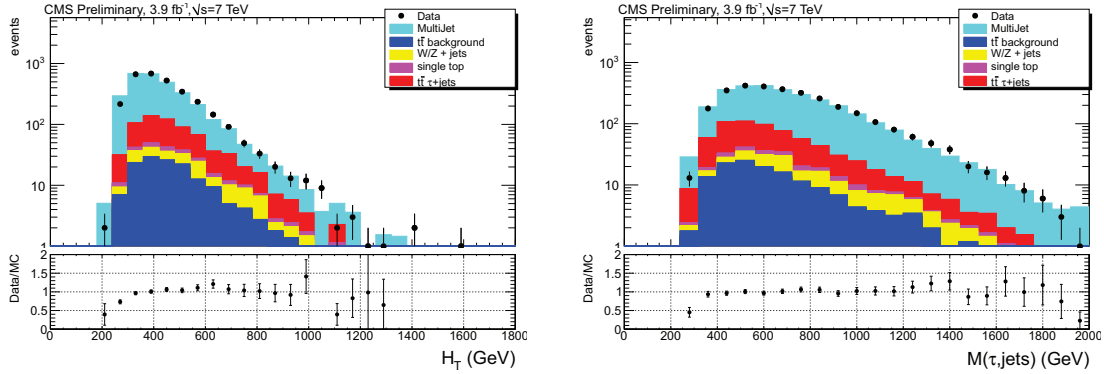
Figure 6.31: Distribution of the NN input variables (left) and invariant mass of the hadronic tau candidate and selected jets $M(\tau_h, jets)$ (right) after a fit of the signal and multijet QCD processes to the data. No "b-mistag reweighting" is applied on the QCD multijet data sample.
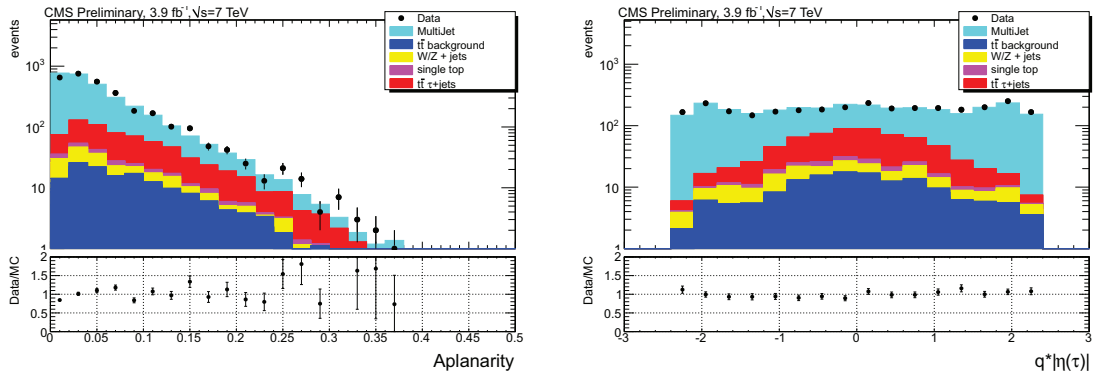

Figure 6.32: Distribution of the NN input variables aplanarity (left) and $q * |\eta(\tau_h)|$ (right) after a fit of the signal and multijet QCD processes to the data. No "b-mistag reweighting" is applied on the QCD multijet data sample.
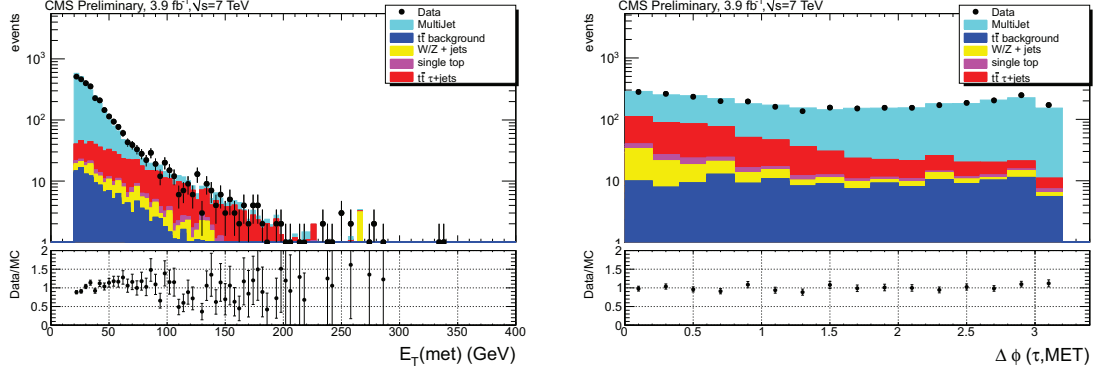
Figure 6.33: Distribution of the NN input variables (left) and $\Delta\phi(\tau_h,)$ (right) after a fit of the signal and multijet QCD processes to the data. No "b-mistag reweighting" is applied on the QCD multijet data sample.

Figure 6.34: Distribution of the NN input variable $\chi^2(kin.fit)$ (left) and kin. fit probability (right) after a fit of the signal and multijet QCD processes to the data. No "b-mistag reweighting" is applied on the QCD multijet data sample.
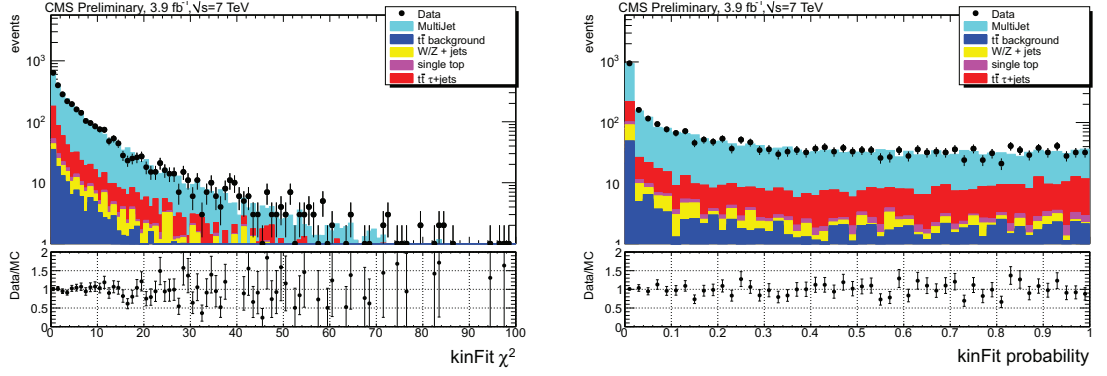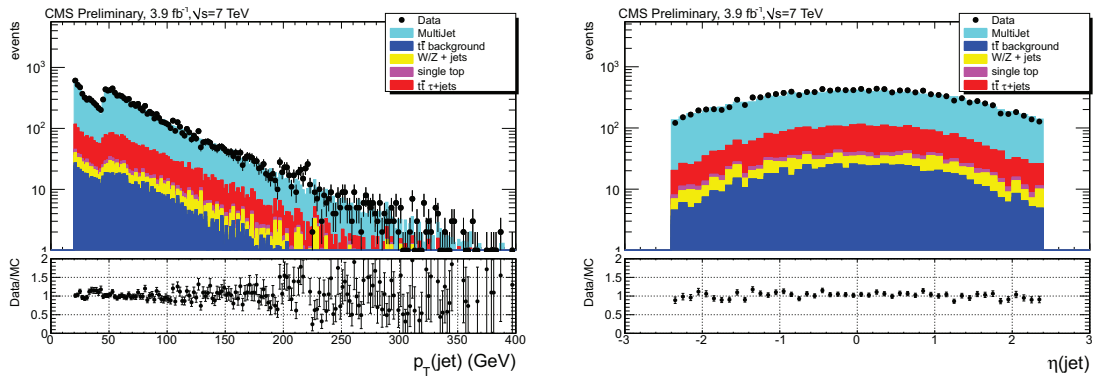

Figure 6.35: Distribution of the momentum (left) and the pseudorapidity (right) of the selected jets after a fit of the signal and multijet QCD processes to the data. No "b-mistag reweighting" is applied on the QCD multijet data sample.

Figure 6.36: Distribution of the momentum (left) and the pseudorapidity (right) of the hadronic tau candidate after a fit of the signal and multijet QCD processes to the data. No "b-mistag reweighting" is applied on the QCD multijet data sample.
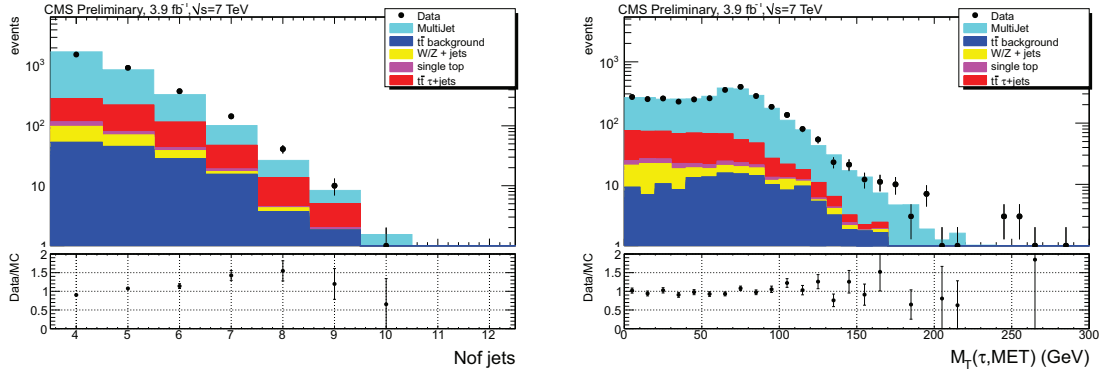
Figure 6.37: Distribution of the selected jet multiplicity (left) and the transverse mass of the hadronic tau candidate and transverse missing energy, $M_T(\tau_h,)$, (right) after a fit of the signal and multijet QCD processes to the data. No "b-mistag reweighting" is applied on the QCD multijet data sample.
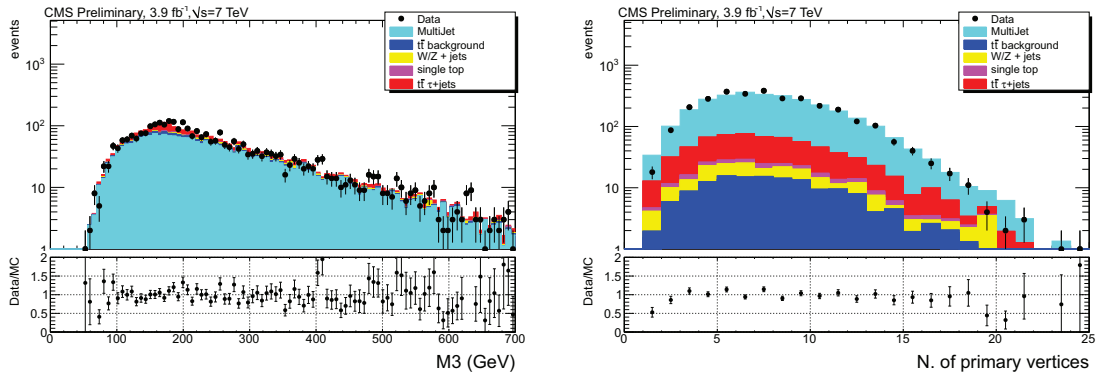

Figure 6.38: Distribution of the reconstructed M3 variable (left) and of the primary vertices multiplicity (right) after a fit of the signal and multijet QCD processes to the data. No "b-mistag reweighting" is applied on the QCD multijet data sample.

6.9 Validation of the b-tagging approach

In order to validate the probabilistic approach (see section **6.3.1.6**) we compared the shapes obtained for the NN input variables using simulated $t\bar{t}$ signal events (figures **6.39**-**6.41**). As we can see the shapes obtained using the probabilistic approach and the direct b-tagging are compatible for all NN input variables.
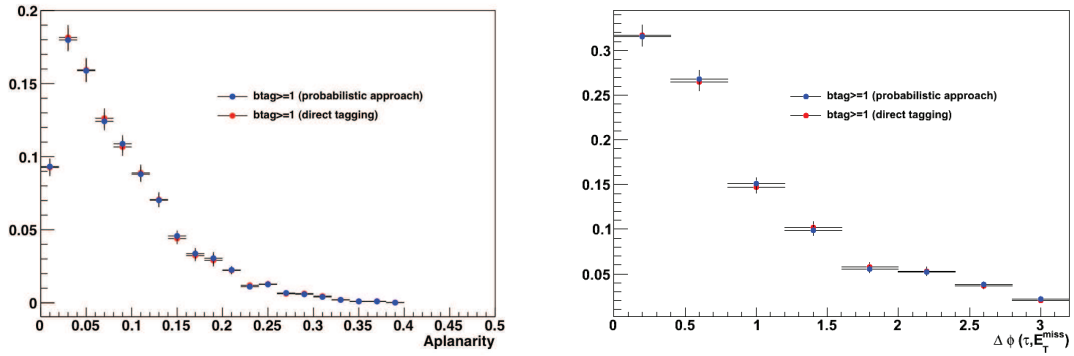

Figure 6.39: Comparison between the distribution obtained applying the direct b-tag method and the probabilistic approach to the simulated $t\bar{t}$ events for aplanarity and $\Delta\phi(\tau, E_T^{miss})$ variables.
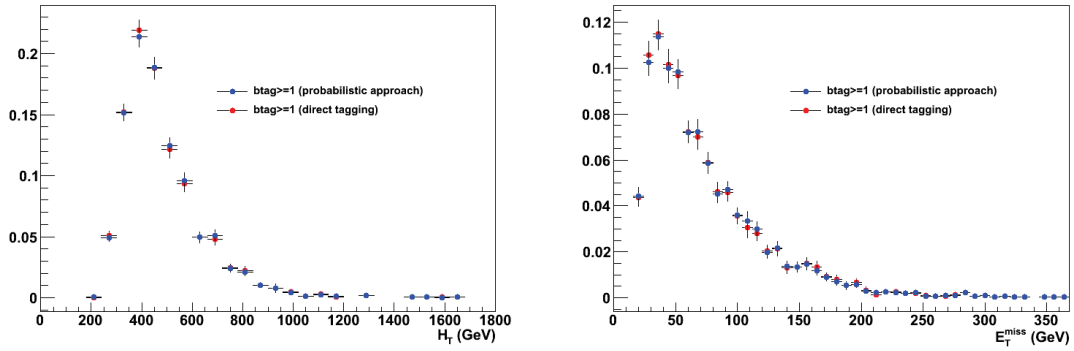

Figure 6.40: The same than figure **6.39** for H_T and E_T^{miss}.
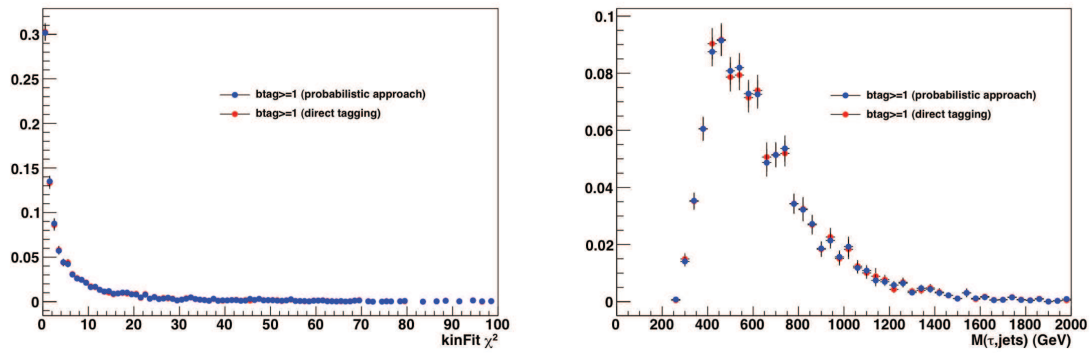
Figure 6.41: The same than figure **6.39** for χ^2 (kin. fit) and M(τ,jets).

References

[1] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967.

[2] Richard L. Garwin, Leon M. Lederman, and Marcel Weinrich. Observations of the failure of conservation of parity and charge conjugation in meson decays: the magnetic moment of the free muon. *Phys. Rev.*, 105:1415–1417, Feb 1957.

[3] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. EXPERIMENTAL TEST OF PARITY CONSERVATION IN BETA DECAY. *Phys. Rev.*, 105:1413–1414, 1957.

[4] M. Goldhaber, L. Grodzins, and A. W. Sunyar. Helicity of neutrinos. *Phys. Rev.*, 109:1015–1017, Feb 1958.

[5] O. W. Greenberg. Spin and Unitary Spin Independence in a Paraquark Model of Baryons and Mesons. *Phys. Rev. Lett.*, 13:598–602, 1964.

[6] R.Brock *et all* CTEQ collaboration. Handbook of perturbative QCD: Version 1.0. *Rev. Mod. Phys.*, 67:157–248, 1995.

[7] CMS collaboration. Constraints on the top-quark charge from top-pair events. *CMS PAS TOP-11-031*, 2012.

[8] CDF collaboration. Observation of Top Quark Production in $\overline{p}p$ Collisions with the Collider Detector at Fermilab. *Phys. Rev. Lett.*, 74:2626–2631, Apr 1995.

[9] Tevatron Electroweak Working Group, CDF, and D0 Collaborations. Combination of CDF and DO results on the mass of the top quark using up to 5.8 fb-1 of data. arXiv:1107.5255v3, 2011.

[10] CMS collaboration. Combination of atlas and cms results on the mass of the top quark using up to 700 pb^{-1} of data. *CMS PAS*, TOP-12-001, 2012.

[11] J. Alcaraz et al. A Combination of preliminary electroweak measurements and constraints on the standard model. arXiv:hep-ex/0612034v2, 2006.

[12] https://twiki.cern.ch/twiki/bin/viewauth/CMS/StandardModelCrossSection.

REFERENCES

[13] Wolfgang Wagner. Top quark physics in hadron collisions. *Rept. Prog. Phys.*, 68:2409–2494, 2005.

[14] Chia-Hung V. Chang and Hsiang-nan Li. Three-scale factorization theorem and effective field theory: Analysis of nonleptonic heavy meson decays. *Phys. Rev. D*, 55:5577–5580, May 1997.

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

[16] CMS collaboration. Measurement of btagging efficiency using ttbar events. *CMS PAS*, BTV-11-003, 2012.

[17] D. N. Spergel et al. Wilkinson Microwave Anisotropy Probe (WMAP) three year results: Implications for cosmology. *Astrophys. J. Suppl.*, 170:377, 2007.

[18] G.C. Branco, P.M. Ferreira, L. Lavoura, M.N. Rebelo, Marc Sher, et al. Theory and phenomenology of two-Higgs-doublet models. *Phys.Rept.*, 516:1–102, 2012.

[19] B. Fuks and M. Rausch de Traubenberg. *Supersymétrie - exercices avec solutions*. Editions Ellipses Marketing, 2011.

[20] Wim de Boer and Christian Sander. Global electroweak fits and gauge coupling unification. *Phys.Lett.*, B585:276–286, 2004.

[21] F. Krauss, R. Kuhn, and G. Soff. AMEGIC++ 1.0: A Matrix element generator in C++. *JHEP*, 0202:044, 2002.

[22] John Campbell and R. K. Ellis. Next-to-leading order corrections to $w + 2$ jet and $z + 2$ jet production at hadron colliders. *Phys. Rev. D*, 65:113007, Jun 2002.

[23] F. Krauss, A. Schalicke, and G. Soff. APACIC++ 2.0: A Parton cascade in C++. *Comput.Phys.Commun.*, 174:876–902, 2006.

[24] Z. Was. TAUOLA the library for tau lepton decay, and KKMC / KORALB / KORALZ /... status report. *Nucl.Phys.Proc.Suppl.*, 98:96–102, 2001.

[25] Yves Baconnier, Giorgio Brianti, P Lebrun, A G Mathewson, R Perin, and Yves Baconnier. LHC: the Large Hadron Collider accelerator project. 1993.

[26] R. Adolphi et al. The cms experiment at the cern lhc. *JINST*, 0803:S08004, 2008.

[27] Atlas collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *J. Instrum.*, 3:S08003. 437 p, 2008.

[28] LHCb collaboration. *LHCb : Technical Proposal.* Tech. Proposal. CERN, Geneva, 1998.

[29] ALICE collaboration. *A large ion collider experiment technical proposal design.* lhcc 95-71, lhcc/p3,1995. Cern, Geneva.

[30] CMS collaboration. The CMS magnet project: Technical Design Report. 1997.

[31] CMS collaboration. Tracking and vertexing results from first collisions. *CMS PAS*, TRK-10-001, 2010.

[32] CMS collaboration. Studies for tracker material. *CMS PAS*, TRK-10-003, 2010.

[33] CMS collaboration. The CMS tracker system project: Technical Design Report. 1997.

[34] CMS collaboration. The CMS tracker: addendum to the Technical Design Report. 2000.

[35] CMS collaboration. The Electromagnetic Calorimeter Project: Technical Design Report. 1997.

[36] CMS collaboration. Electromagnetic calorimeter calibration with 7 TeV data. 2010.

[37] CMS collaboration. The Hadron Calorimeter Project: Technical Design Report. 1997.

[38] The CMS collaboration. The Muon Project: Technical Design Report. 1997.

[39] CMS collaboration. Cms tridas project: Technical design report; 1, the trigger systems.

[40] Daniel de Florian and Massimiliano Grazzini. Higgs production through gluon fusion: Updated cross sections at the Tevatron and the LHC. *Phys.Lett.*, B674:291–294, 2009.

[41] Paolo Bolzoni, Fabio Maltoni, Sven-Olaf Moch, and Marco Zaro. Vector boson fusion at next-to-next-to-leading order in qcd: Standard model higgs boson and beyond. *Phys. Rev. D*, 85:035002, Feb 2012.

[42] The Higgs PAG. Trigger strategies for higgs searches. CMS Analysis Note 2011/065, 2011.

[43] CMS collaboration. Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV. *CMS PAS PFT-10-002*, 2010.

REFERENCES

[44] CMS collaboration. Tracking and Primary Vertex Results in First 7 TeV Collisions. *CMS PAS TRK-10-005*, 2010.

[45] Pioppi Michele. Iterative traking. CMS Internal Note 2011/334, 2011.

[46] W Adam, R Frhwirth, A Strandlie, and T Todorov. Reconstruction of electrons with the gaussian-sum filter in the cms tracker at the lhc. *Journal of Physics G: Nuclear and Particle Physics*, 31(9):N9, 2005.

[47] CMS collaboration. Jet Performance in pp Collisions at 7 TeV. *CMS PAS JME-10-003*, 2010.

[48] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

[49] CMS Collaboration. Determination of jet energy calibration and transverse momentum resolution in CMS. *Journal of Instrumentation*, 6:11002, November 2011.

[50] The CMS collaboration. Jet energy corrections determination at 7 tev. *CMS PAS JME-10-010*, 2011.

[51] K. Nakamura et al. Review of particle physics. *J. Phys.*, G37:075021, 2010.

[52] M. Bachtis at all. Prospects for measurement of $\sigma(\text{pp} \rightarrow \text{z})*\text{br}(\text{z} \rightarrow \tau^+ \tau^-)$ with cms in pp collisions at $\sqrt{s} = 7$ tev. CMS Analysis Note 2010/082, 2010.

[53] The CMS collaboration. Performance of tau reconstruction algorithms in 2010 data collected with cms. Technical report, 2011.

[54] The TAU POG. Measurement of hadronic tau identification efficiency in 2011 run *b* data. CMS Analysis Note 2011/514, 2011.

[55] The TAU POG. Measurement of tau identification efficiency with 2011 cms data. CMS Analysis Note 2011/200, 2011.

[56] M. Bachtis at all. Search for higgs bosons in di-tau final states with cms run2011 data. CMS Analysis Note 2011/160, 2011.

[57] J.Conway et all. The tau neural classifier algorithm: tau identification and decay mode reconstruction using neural networks. CMS Analysis Note 2010/099, 2010.

[58] CMS Collaboration. Performance of -lepton reconstruction and identification in cms. *Journal of Instrumentation*, 7(01):P01001, 2012.

[59] M. Bachtis et all. Study of tau reconstruction algorithms using pp collisions data collected at $\sqrt{s} = 7$ tev. CMS Analysis Note 2011/001, 2011.

[60] CMS Collaboration. The cms physics technical design report, volume 1. *CERN/LHCC*, 2006-001, 2006.

[61] Wolfgang Waltenberger. Adaptive vertex reconstruction. CMS Analysis Note 2008/33, 2008.

[62] W. Erdmann. Offline primary vertex reconstruction with deterministic annealing clustering. CMS Internal Note 2011/014, 2011.

[63] R.Castello and others. Estimation of the b-tagging efficiency scale factor with a profile likelihood ration method on top di-lepton events with 2011 data. CMS Analysis Note 2011/180, 2011.

[64] CMS collaboration. Status of b-tagging tools for 2011 data analysis. *CMS PAS*, BTV-11-002, 2011.

[65] https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideBTagJetProbabilityCalibration.

[66] CMS collaboration. b-jet identification in the cms experiment. *CMS PAS*, BTV-11-004, 2012.

[67] D.Bloch and others. Measurement of the b-tagging efficiency at high p_t using the jet probability. CMS Analysis Note 2011/448, 2011.

[68] http://root.cern.ch/root/html/TFractionFitter.html.

[69] J.Andrea and others. Estimation of the b-tagging efficiency scale factor with a profile likelihood ration method on top di-lepton events with 2011 data. CMS Analysis Note 2011/438, 2011.

[70] The CMS collaboration. Measurement of btagging efficiency using ttbar events. Technical report, 2011.

[71] The Collaboration. Measurement of the production in the τ+jets topology using $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ tev. *Physical Review D*, 071102(R), 2010.

[72] The CDF Collaboration. Measurements of top quark properties in the τ + jets decay channel at cdf. *CDF note 10562*, 2011.

[73] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5: Going Beyond. *JHEP*, 1106:128, 2011.

[74] J. Pumplin et al. New generation of parton distributions with uncertainties from global qcd analysis. *JHEP 0207 (2002) 012, hep-ph/0201195*, 2002.

[75] S. Alioli et al. A general framework for implementing nlo calculations in shower monte carlo programs: the powheg box. *JHEP 1006 (2010) 043, hep-ph/10022581*, 2010.

REFERENCES

[76] Nikolaos Kidonakis. Higher-order corrections to top-antitoppair and single top quark production. In *Proceedings of the DPF-2009 Conference, Detroit, MI, July 27-31, 2009*. hep-ph, 2009. arXiv:0909.0037v1.

[77] https://twiki.cern.ch/twiki/bin/viewauth/CMS/StandardModelCrossSections.

[78] https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookJetEnergyCorrections.

[79] M. Chiochia et al. Measurement of tau identification efficiency with 2011 cms data. CMS Analysis Note 2011/200, 2011.

[80] M. Chen et al. Measurement of hadronic tau identification efficiency in 2011 run b data. CMS Analysis Note 2011/514, 2011.

[81] D. Barge et al. Study of photon conversion rejection at cms. CMS Analysis Note 2009/159, 2009.

[82] J. Andrea et al. Measurement of the top dilepton cross section using b-tagging at $\sqrt{s} = 7$ tev with 1.14 fb^{-1} of pp collisions. CMS Analysis Note 2011/334, 2011.

[83] https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupInformation.

[84] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss. TMVA: Toolkit for Multivariate Data Analysis. *PoS*, ACAT:040, 2007.

[85] J. D'Hondt et al. Fitting with event topologies with external kinematic constraints in cms. 2005/025, 2005.

[86] W. Verkerke and D. Kirkby. The roofit toolkit for data modeling. *physics.data-an/0306116v1*, 2003.

[87] P Biallass et al. Parton distribution uncertainty determination within cmssw. CMS Analysis Note 2009/048, 2009.

[88] The CMS collaboration. Measurement of the $t\bar{t}$ pair production cross section at $\sqrt{s} = 7$ using b-quark jet identification techniques in lepton + jet events. *CMS PAS TOP-11-003*, 2011.

[89] The CMS collaboration. Combination of top quark pair production cross section measurements. *CMS PAS TOP-11-024*, 2011.

[90] Nikolaos Kidonakis. Next-to-next-to-leading soft-gluon corrections for the top quark cross section and transverse momentum distribution. *Phys.Rev.*, D82:114030, 2010.

[91] Valentin Ahrens, Andrea Ferroglia, Matthias Neubert, Ben D. Pecjak, and Li Lin Yang. Renormalization-Group Improved Predictions for Top-Quark Pair Production at Hadron Colliders. *JHEP*, 1009:097, 2010.

[92] The ATLAS collaboration. Measurement of the $t\bar{t}$ production cross section in the final state with a hadronically decaying tau lepton and jets using the atlas detector. Technical Report ATLAS-CONF-2012-032, CERN, Geneva, Mar 2012.

[93] The CMS collaboration. Search for the light charged higgs boson in top quark decays in pp collisions at $\sqrt{s} = 7$. *CMS PAS HIG-11-019*, 2011.